Supplementary Information

Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans

Keinan A, Mullikin JC, Patterson N, and Reich D

Supplementary Methods

SNP filtering and balancing of SNPs from different sources: We eliminated SNPs that did not pass HapMap's quality control filters, were monomorphic across all samples or were less than 80% complete in any one sample. Following the application of these criteria, we balanced the genotyping failure rate of HapMap Phase 1[1] with that of HapMap Phase 2[2]: Since Phase 1 involved several different genotyping centers, using different genotyping technologies, and since for the most part each chromosome was genotyped by one center[1], we estimated the per-chromosome failure rate of each phase and removed random SNPs in that chromosome from the phase with the lower failure rate. We removed the ten ENCODE regions from the data set since multiple attempts were made to genotype them.[1] Similarly, we randomly removed SNPs with allele frequency information from the Hinds et al. data set[3] to balance the success rate with that of HapMap Phase 2 (using information on failure rates estimated from chr2p genotyping).

Chromosome 2p validation: We validated our data processing steps by studying chromosome 2p. For these data, all SNPs in dbSNP at the time of marker picking were attempted, regardless of whether they had previously been tried in Phase 1, or their status in the Hinds et al. data set[3]. We were thus able to test whether the characteristics of the allele frequency spectrum extracted from the smaller but perfectly collected chromosome 2p data set are in agreement with the larger whole-genome data set (Supp. Note 2).

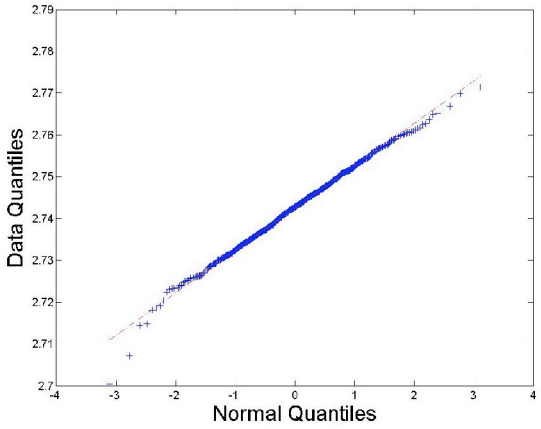
Sequence diversity and divergence estimates: All sequence traces for each individual (Supp. Table 4) were aligned to the reference genome using ssahaSNP, with strict NQS settings to reduce false positives (Qsnp>=40, Qneighbor>=15, Nneighbor=5, maxNeighborhoodDiffs=1,

maxSNPs/kb=15). A subset of non-overlapping sequence traces for each individual was selected, ensuring that we have a single mosaic, haploid genome for each individual. Within-population sequence diversity was computed by examining if there are two or more individual haploid genomes aligned at each consecutive base position in the autosomal genome, other than bases in CpG dinucleotides. If so, the count of bases, L, was incremented. To count differences, two haploid genomes were selected at random and if the alleles between these two individuals differed, the SNP count, K, was incremented. A value for diversity, K/L, was reported each time L reached 100,000, at which point L and K were reset to zero. We report the autosomal mean diversity as the mean of the values reported every 100kb; we use the jackknife method[4] to obtain a standard deviation. To compute between-population sequence divergence, the individual haploid genomes from one population were combined so that at any base, only one individual was represented (we selected one individual at random if more than one individual haploid genome was available at that position). These population-merged mosaic haploid representations were then pairwise-compared (as in the within-population method) to obtain between-population sequence divergence estimates.

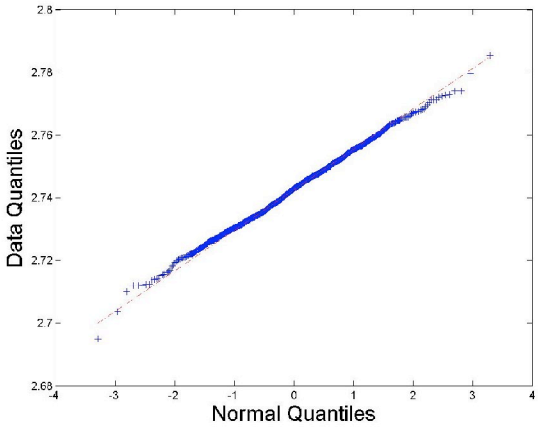
Test for difference in frequency spectrum of two samples: We tested whether the frequency spectrum obtained for two different samples is significantly different by a t-test for a difference in the mean derived allele frequency. This test is conservative as it reflects only a significant deviation of the mean, rather than other differences in the shape of the distribution, but it was still powerful enough to detect highly significant differences among the samples. We report $P << 10^{-12}$ throughout the paper for any value lower than 10^{-12}.

Tajima's D test: To test for a significant rise or fall in Tajima's D statistic compared with the expectation for a constant population size, we used Hudson's ms coalescent simulation software[5]. To mimic the ascertainment scheme in our data, we simulated N+2 chromosomes, only reporting the allele frequency of a SNP (in N chromosomes) if it was polymorphic in the 2 chromosomes used for ascertainment. We matched the number of individuals and the number of SNPs between the data and simulations. For this test we only studied SNPs successfully genotyped in all samples (except for the CHB+JPT analysis, as described below). The null Tajima's D distribution is assumed to be normal with the mean and standard deviation estimated from the simulations (the below figure presents quantile-quantile plots from simulation that justify the normality assumption). Since in the CHB+JPT sample there is one individual who is missing from Phase 1 but not Phase 2 of HapMap[1], ignoring SNPs with missing genotypes results in a data set of only Phase 2 SNPs, which as described in the main text has biased frequencies relative to Phase 1 SNPs and the entirety of HapMap. To deal with this bias, we included in the CHB+JPT Tajima's D analysis SNPs with one missing individual, substituting the genotype of this individual to be homozygous for the major allele. This is conservative for the purpose of a test for deviation from size constancy when Tajima's D is above expectation, as is the case for the CHB+JPT sample.

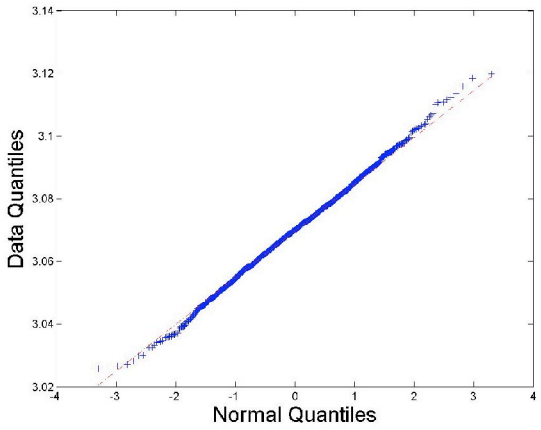
YRI



CEU



CHB+JPT



Quantile-quantile plots between simulated Tajima's D values and a normal distribution.

Allele frequency modeling and inference (expansion of Methods): To estimate model parameters, we employed a likelihood approach that captures the probability of the data conditioned on the ascertainment scheme and the demographic history. Specifically, let s denote the number of SNPs and n denote the number of chromosomal samples (120 for CEU and YRI, 180 for CHB+JPT and 90 for CHB and JPT when considered separately). Additionally, let x_i and m_i denote the number of derived alleles and the total number of successful genotypes at SNP i ($x_i \leq m_i \leq n$). Then, the likelihood of the data is given by

$L = \prod_{i=1}^{s} P(x_i \text{ of } m_i)$, where $P(x_i \text{ of } m_i)$ is the probability of observing x_i derived alleles out of m_i, conditioned on ascertainment and demography. In order to account for the sampling of the m_i successful genotypes out of the possible n, we derived L by the law of total probability to be

$$L = \prod_{i=1}^{s} \sum_{j=x_i}^{n-m_i+x_i} P(x_i \text{ of } m_i \mid j \text{ of } n) P(j \text{ of n}) = \prod_{i=1}^{s} \sum_{j=x_i}^{n-m_i+x_i} \frac{\binom{m_i}{x_i}\binom{n-m_i}{j-x_i}}{\binom{n}{j}} P(j \text{ of } n).$$

Let $P_n(j)$ denote the demography-dependent probability of observing j derived alleles at a polymorphic site out of n ($j = 1,..,n-1$). Then, considering the ascertainment scheme in two chromosomes independent of the samples, it follows from the above, by Bayes' theorem, that

$$L = \prod_{i=1}^{s} \sum_{j=x_i}^{n-m_i+x_i} \frac{\binom{m_i}{x_i}\binom{n-m_i}{j-x_i}}{\binom{n}{j}} \frac{P_{n+2}(j+1)\cdot(j+1)\cdot(n-j+1)}{\sum_{l=0}^{n} P_{n+2}(l+1)\cdot(l+1)\cdot(n-l+1)},$$

Where $(j+1)\cdot(n-j+1)$ is proportional to the probability of one random chromosomal sample out of the $n+2$ having the derived allele and another random chromosomal sample having the ancestral allele, when $j+1$ samples have the derived allele. Last, we note that the normalization factor, which captures the overall probability of ascertaining a SNP from the demography captured by P_{n+2}, depends on the demography alone, not the data, which results in

$$L = \prod_{i=1}^{s} \frac{\sum_{j=x_i}^{n-m_i+x_i} \frac{\binom{m_i}{x_i}\binom{n-m_i}{j-x_i}}{\binom{n}{j}} P_{n+2}(j+1)(j+1)(n-j+1)}{\sum_{l=0}^{n} P_{n+2}(l+1)(l+1)(n-l+1)}.$$

To profile the likelihood and determine the maximum likelihood estimates, we evaluated $\log L$ for different P_{n+2} functions. This likelihood formulation assumes that the SNPs are independent, which is not the case for SNPs in linkage disequilibrium. To avoid underestimation of the errors due to SNPs dependency, we used a Moving Block Bootstrap (MBB) for dependent data for estimating the standard deviation of the maximum likelihood estimates (Methods).

To evaluate the log-likelihood under the modeled demographic histories, we first evaluated the frequency spectrum of the multi-epoch model[6], given by

$$P_n(i) \propto \frac{N_1}{i} + \sum_{m=1}^{M-1} \left\{ \frac{N_{m+1} - N_m}{i} \binom{n-1}{i}^{-1} \sum_{k=2}^{n} \left[\binom{n-k}{i-1} \sum_{j=k}^{n} \left(e^{-\binom{j}{2} \sum_{l=1}^{m} T_l / 2N_l} \prod_{\substack{l:l\neq j; \\ k\leq l\leq n}} \frac{l(l-1)}{l(l-1) - j(j-1)} \right) \right] \right\}$$

for $i = 1,..,n-1$. M denotes the number of epochs, N_m the effective population size during epoch m and T_m the length of epoch m, in generations. We evaluated this equation using software originally described by Marth et al.[6].

The theoretical frequency spectrum, $P_n(i)$, remains unchanged under simultaneous scaling of all effective population sizes and epoch durations[6]. Hence, an *M*-epoch model entails $2M-1$ degrees of freedom. We evaluated $2M-1$ parameters by keeping the ancestral effective population size fixed at an arbitrary value $(10{,}000)$. Then, we normalized the maximum

likelihood estimates of all parameters, including the ancestral effective population size, to fit the observed sequence heterozygosity which is expected to be equal to $4\mu\left\{N_1 + \sum_{m=1}^{M-1}\left[\left(N_{m+1} - N_m\right)e^{-\sum_{l=1}^{m} T_l / 2N_l}\right]\right\}$, where μ is the per-generation mutation rate. We separately applied the normalization to each bootstrap run.

The simplest model we used is a model that assumes a constant population size throughout history. Then, we considered a model with one change of effective population size, capturing an expansion or contraction, with two parameters that capture the time of change and relative population size before and after the change[6]. Bottlenecks were modeled as a crash in population size for a fixed number of generations followed by re-expansion to the same effective population size as before the bottleneck, with two parameters capturing the time and inbreeding coefficient of the bottleneck. The inbreeding coefficient, defined here as $F = T/2N$, is approximately the probability that two alleles randomly picked from the population after the bottleneck derive from the same ancestral allele just before the bottleneck.[7] The effect of a bottleneck on the frequency spectrum depends primarily on this ratio and is practically independent of the predefined value of T, the number of generations used to model the bottleneck, since a simultaneous scaling of both *T* and *N* does not change the results (Supp. Note 6). Supp. Note 1 provides further details about each of the models.

Joint-frequency modeling and inference: We employed a similar maximum likelihood formulation for modeling joint-frequencies in two populations, capturing the probability of the observed derived and ancestral counts of each SNP in both analyzed samples, while

conditioning on ascertainment in either of the corresponding populations (the analysis is repeated for ascertainment in each population). For each possible history, we obtained the expected proportion of SNPs of any allele joint-frequency by simulating ten million different genealogies by coalescent simulations (using the ms program[5]). Standard deviation estimates were obtained by bootstrapping one thousand random data sets using MBB.

Let n_1 and n_2 denote the number of chromosomal samples in the two populations. Let x_i^1 and m_i^1 denote the number of derived alleles and the number of successful genotypes at SNP i in population 1 ($x_i^1 \le m_i^1 \le n_1$), the population in which SNPs are ascertained, and similarly x_i^2 and m_i^2 for population 2. Let $P_{n_1,n_2}(j_1, j_2)$ denote the demography-dependent probability of observing j_1 derived alleles out of n_1 in population 1 and j_2 derived alleles out of n_2 in population 2 ($j_1 = 1,..,n_1 - 1; j_2 = 0,..,n_2$). Then, the likelihood of the data conditioned on demography and on ascertainment in population 1 is given by

$$L = \prod_{i=1}^{s} \frac{\sum_{j_1=x_i^1}^{n_1-m_i^1+x_i^1} \sum_{j_2=x_i^2}^{n_2-m_i^2+x_i^2} \frac{\binom{m_i^1}{x_i^1}\binom{n_1-m_i^1}{j_1-x_i^1}}{\binom{n_1}{j_1}} \frac{\binom{m_i^2}{x_i^2}\binom{n_2-m_i^2}{j_2-x_i^2}}{\binom{n_2}{j_2}} P_{n_1+2,n_2}(j_1+1, j_2)(j_1+1)(n_1-j_1+1)}{\sum_{l_1=0}^{n_1}\sum_{l_2=0}^{n_2} P_{n_1+2,n_2}(l_1+1, l_2)(l_1+1)(n_1-l_1+1)}.$$

We used this inference framework to estimate divergence time by considering the joint allele frequencies of the CEU and CHB+JPT samples. Other parameters of demographic history were inferred based on the maximum likelihood estimates of the two-bottleneck model in the separate populations. We evaluated the likelihood over a grid of a hundred values of the parameter.

These values allow the divergence to have occurred before both bottlenecks, between them, after both of them or during either of the two bottlenecks, with different coalescent simulations capturing each of these possibilities. Whenever a bottleneck is shared, we averaged its parameters across the maximum likelihood estimates of the two samples. If it is partially shared, we averaged the time of the bottleneck, while keeping the inbreeding coefficient of both samples equal to the maximum likelihood estimates. We similarly estimated F_{ST} for the cases of divergence after the ancient bottleneck and of divergence after the recent bottleneck by simulating data sets under each of these two scenarios.

Ancestry determination: We determined the ancestry of the HuAA and HuFF libraries[8] by examining the alleles these individuals carried at SNPs that we knew were highly informative about biogeographic ancestry, and for which frequencies in West Africans, Europeans, East Asians, and Amerindians have been previously measured[9]. We used Maximum likelihood to identify the best-matching populations[10], using 367 SNPs for HuAA and 235 SNPs for HuFF. This inference matched reported ancestry for those samples for which we could obtain both types of information.

Supplementary Table 1: Tajima's D test

	YRI	CEU	CHB+JPT
Tajima's D estimate	2.279	3.043	3.437
Expectation	2.743	2.743	3.07
Standard deviation	0.01	0.01	0.016
p-value	$<< 10^{-12}$	$<< 10^{-12}$	$<< 10^{-12}$

While theory suggests that the expectation of Tajima's D is zero[11], it largely depends on the ascertainment. The expectation and standard deviation reported in the table are for a constant population size under the same ascertainment scheme as the data. P-value is for a two-tailed test of deviation from this expectation.

Supplementary Table 2: Libraries for sequence diversity and divergence estimates

Library	Sequencing center	Population, Coriell ID	Traces*	Aligned bases**
ABC7	Agencourt	YRI, NA18517	1,239,791	401,628,732
ABC8	Agencourt	YRI, NA18507	2,759,301	776,422,446
ABC9	Agencourt	JPT, NA18956	1,597,706	448,184,613
ABC10	Agencourt	YRI, NA19240	1,608,045	441,885,522
ABC12	Agencourt	CEU, NA12878	574,271	178,168,556
Cor7340	Sanger Institute	CEU, NA07340	2,721,720	424,328,147
Cor10470	Sanger Institute	Biaka Pygmy, NA10470	358,490	31,209,978
Cor11321	Sanger Institute	East Asian, NA11321	1,828,769	550,965,952
HuAA	Celera	European American, Individual A	2,408,092	566,649,249
HuBB	Celera	European American, Individual B	5,851,971	985,652,063
HuFF	Celera	East Asian, Individual F	1,272,561	356,161,178

* total number of traces uniquely aligned to the reference genome sequence.
** total number of aligned bases used in analysis.

Supplementary Table 3: F_{ST} estimates between West Africans and non-Africans

Ascertainment library	F_{ST}(YRI, CEU)	F_{ST}(YRI,CHB+JPT)	F_{ST}(YRI,CHB+JPT) – F_{ST}(YRI, CEU)	p-value
Cor10470 (Biaka Pygmy)	**0.143 ± 0.0054**	**0.182 ± 0.0068**	**0.0380 ± 0.0061**	**$3.1 \cdot 10^{-10}$**
Cor17119 (African American)	**0.158 ± 0.0015**	**0.187 ± 0.0017**	**0.0289 ± 0.0017**	**$<< 10^{-12}$**
Cor17109 (African American)	**0.146 ± 0.0026**	**0.179 ± 0.0033**	**0.0332 ± 0.0030**	**$<< 10^{-12}$**
Cor11321 (East Asian)	0.152 ± 0.0021	0.158 ± 0.0019	0.0062 ± 0.0021	0.0015
HuFF (East Asian)	0.158 ± 0.0036	0.162 ± 0.0034	0.0042 ± 0.0037	0.13
HuAA (European American)	0.140 ± 0.0018	0.179 ± 0.0025	0.0388 ± 0.0024	$<< 10^{-12}$
Cor7340 (European American)	0.142 ± 0.0019	0.182 ± 0.0024	0.0404 ± 0.0023	$<< 10^{-12}$

F_{ST} calculation follows Supp. Note 10, with each row considering all SNPs ascertained in the indicated library. All three columns present mean and standard deviation over 1000 bootstraps using MBB. P-value is for testing whether F_{ST}(YRI,CHB+JPT) minus F_{ST}(YRI,CEU) is larger than zero, based on the bootstrap estimates for this difference (fourth column; one-tailed z-test). The estimation does not account for the ascertainment scheme, resulting in our estimates being larger than previously reported[1,12]. Note that F_{ST}(YRI,CEU) is lower for ascertainment in European libraries and similarly, F_{ST}(YRI, CHB+JPT) is lower for ascertainment in East Asian libraries, due to the frequency spectra these ascertainment schemes entail. However, the comparison of F_{ST}(YRI,CHB+JPT) and F_{ST}(YRI,CEU) for ascertainment in libraries from populations that diverged from the ancestral population of CEU and CHB+JPT (indicated in bold) is not biased by ascertainment (Supp. Note 10).

Supplementary Table 4: F_{ST} estimates between Europeans and East Asians

Ascertainment library	Observed F_{ST}	Expected F_{ST} for divergence after the recent bottleneck; p-value	Expected F_{ST} for divergence after the ancient bottleneck; p-value
Cor10470 (Biaka Pygmy)	0.105 ± 0.0057		
Cor17119 (African American)	0.112 ± 0.0015		
Cor17109 (African American)	0.107 ± 0.0027		
Cor11321 (East Asian)	0.099 ± 0.0017	0.0327 ± 0.0002; $<<10^{-12}$	0.2166 ± 0.0011; $<<10^{-12}$
HuFF (East Asian)	0.098 ± 0.0025	0.0328 ± 0.0007; $<<10^{-12}$	0.2166 ± 0.0031; $<<10^{-12}$
HuAA (European American)	0.109 ± 0.0017	0.0332 ± 0.0005; $<<10^{-12}$	0.2353 ± 0.0020; $<<10^{-12}$
Cor7340 (European American)	0.106 ± 0.0019	0.0332 ± 0.0002; $<<10^{-12}$	0.2353 ± 0.0014; $<<10^{-12}$

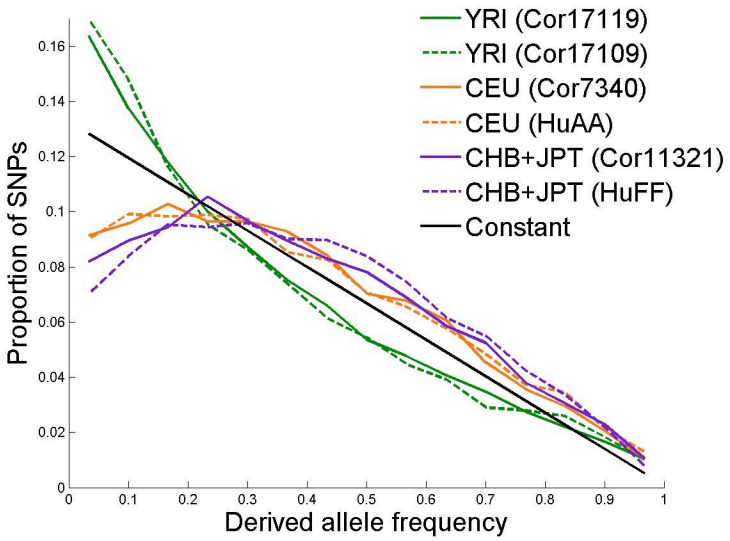
F_{ST} estimates between the CEU and the CHB+JPT samples (mean and standard deviation over 1000 bootstraps using MBB), together with the expectation if the populations diverged *immediately* after the recent bottleneck or the ancient bottleneck (mean and standard deviation across 100 randomly generated data sets). The estimates are affected by the ascertainment scheme, resulting in larger estimates than previously reported, but we emphasize that the theoretical expectations were derived under the same ascertainment scheme (based on the ascertainment population and the number of SNPs as dictated by the ascertainment library), allowing for a meaningful comparison. P-values are for the deviation of the reported F_{ST}(CEU,CHB+JPT) from these theoretical expectations (two-tailed z-test). Since the theoretical expectations account for the ascertainment, they are reported only for ascertainment in either an East Asian library or a European library.

Supplementary Table 5: Libraries and SNP Density

Individual	Ancestry	Chromosomes	# SNPs	median spacing (bp)	median block size for bootstrapping (kb)
Cor17119	African American	2-14; 20-22	114,198	6,224	654
Cor17109	African American	1; 16-19	17,176	10,953	576
Cor7340	European American	1; 6; 9-13; 20; 22	52,184	5,985	594
HuAA	European American	1-22	18,006	55,490	3,197
Cor11321	East Asian	1; 6; 9-13; 20; 22	45,721	6,332	637
HuFF	East Asian	1-22	5,250	254,900	7,552
Cor10470	Biaka Pygmy	20; 22	3,997	8,759	262

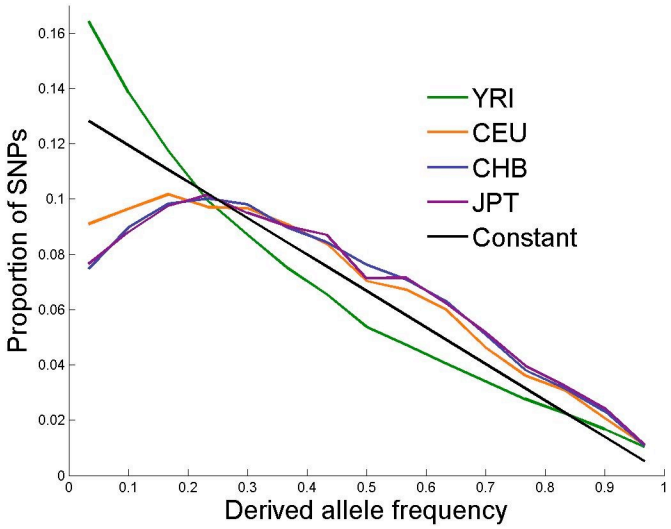
The table lists, for each of the seven ascertainment libraries described in Table 1, the chromosomes used for ascertainment, the number of SNPs following all data correction, the median spacing between a pair of consecutive SNPs and the median block size of the Moving Block Bootstrap (the block size was chosen to be $\sqrt[3]{n}$, where *n* is the number of SNPs; see ref. 13).

Supplementary Figure 1: Derived allele frequency spectra: Comparing ascertainment libraries



Supplementary Figure 1: Derived allele frequency spectrum for each of the HapMap samples for ascertainment in libraries of similar ancestry. Each spectrum is plotted twice, once for each of the corresponding ascertainment libraries. For each pair of libraries, the one with a larger number of SNPs is represented by a solid line and the other by a dashed line, which hence tends to be noisier (particularly the one based on HuFF that has only 5,250 SNPs). Each pair of libraries of similar ancestry generates very similar spectra.

Supplementary Figure 2: Derived allele frequency spectra: CHB versus JPT



Supplementary Figure 2: Derived allele frequency spectrum for each of the HapMap samples for ascertainment in libraries of similar ancestry. The YRI spectrum is based on SNPs from both the Cor17109 and the Cor17119 libraries; the CEU spectrum is based on the Cor7340 and the HuAA libraries; and the CHB and JPT spectra are based on the Cor11321 and the HuFF libraries. For comparison, the black line depicts the expected derived frequency spectrum of a population of effective constant size throughout history under the same ascertainment scheme. CHB and JPT exhibit very similar spectra, motivating our decision to pool them for evaluating the history of East Asian populations.

Supplementary Note 1
Detailed modeling results

This note details the different models used in the paper, along with the maximum likelihood estimates obtained for the different populations. The different models considered, with their parameters, are:

1. Constant population size, with one parameter: the effective population size (Ne).

2. 2-epoch model, with three parameters: the effective population size before population size change; the time of population size change; the effective population size after population size change.

3. Bottleneck model, with three parameters: the intensity of the bottleneck; the time of the bottleneck; the effective population size which we model to be the same before and after the bottleneck.

4. Two-bottleneck model, with five parameters: the intensity of each bottleneck separately; the time of each bottleneck separately; the effective population size which we model to be the same before, between and after the bottlenecks.

Since the proportions of allele frequencies are independent of a simultaneous scaling of all effective population sizes and time estimates[6], we normalized the maximum likelihood estimates of all parameters to fit the observed sequence heterozygosity in the population analyzed (Supp. Methods). As a consequence, the number of degrees of freedom is one less than the number of parameters indicated above. Hence, in the main text of the paper, as well as in the following, the ancestral effective population size, a common parameter of all models, is estimated by the normalization and not considered as a parameter.

Two-epoch model

The 2-epoch model[6] allows one population size change in the history of a population, with two parameters capturing the time of change and relative population size before and after this demographic event. The maximum likelihood estimates and standard deviations are summarized in Table 1. This is a very simplified model that uses one simple demographic event to capture different features of the population history. This averaged signal points to about 1.8-fold population expansion of the West African (YRI) sample and about 0.4-fold contraction of both non-African populations. Fig. 1 depicts the maximum log-likelihood versus the factor of population size change. We note that the likelihood peaks for the two epoch model and the other analyses below are too narrow due to linkage disequilibrium among SNPs (we estimate they are too narrow by 1.5- to 2-fold based on the Moving Block Bootstrap results). However, we expect that the likelihood peaks are unbiased. A caveat to interpreting these results is that for the African analysis, SNPs were discovered in African Americans, who have some European

ancestry. However, repeating the analyses only using SNPs discovered in sections of the genome with all African ancestry, the modeling results are similar (Supp. Note 3).

The results, as well as the results of the next models, are normalized by sequence heterozygosity, based on a value of 0.0008359 for YRI, 0.0006044 for CEU and 0.0005741 differences per base pair for CHB+JPT (Methods). Under a constant population size model and assuming 25 years per generation, these correspond to an effective population size of 10,573, 7,612 and 7,231, correspondingly, which are concordant with the commonly estimated value of about 10,000 for the ancestral population[14-16].

	YRI	CEU	CHB+JPT
Ne before change	9069 ± 54	23,344 ± 18,583	11,556 ± 1926
Ne after change	16,196 ± 289	5,634 ± 127	4,465 ± 168
change factor	**1.79 ± 0.04**	**0.35 ± 0.14**	**0.39 ± 0.03**
time of change (kya)	186 ± 7	506 ± 254	208 ± 61
p-value	$<< 10^{-12}$	$4.9 \cdot 10^{-6}$	$<< 10^{-12}$

Table 1: **2-epoch modeling results**. Ne estimates are based on 25 years per generation and on normalization by sequence heterozygosity. The time of change is also normalized by sequence heterozygosity, but is independent of the number of years per generation. The change factor, which is the ratio of Ne after the change to Ne before the change, is independent of both the normalization and the number of years per generation, and hence it is the most robust model statistic from these analyses. A value greater than 1 denotes a population expansion, while a value smaller than 1 denotes a contraction. P-value is for testing whether the change factor is different from 1 (two-tailed z-test).

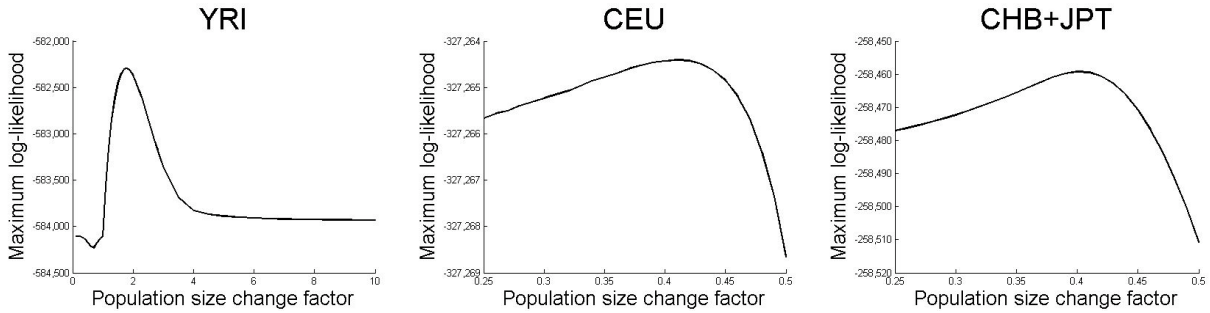


Figure 1: **2-epoch likelihood profiling.** Maximum log-likelihood (maximized over the time of event) versus the factor of population size change, which is independent of both the normalization and the number of years per generation, exhibiting a distinct peak. The observed maximum differs slightly from the estimates detailed in Table 1, as the estimates in Table 1 are based on bootstrapping, while this figure profiles the likelihood function from the full data set. (Unlike the bootstrap analysis, which we use to calculate the P-values in the paper, the likelihood analysis based on the full data set does not take into account correlation among SNPs in LD, and so the likelihood surfaces are also more peaked than is in fact appropriate.)

Bottleneck model

The bottleneck model allows one bottleneck event in the history of the population, with two parameters capturing the time of the demographic event and its inbreeding coefficient. Table 2 presents the maximum likelihood estimates and standard deviations for the European (CEU) and

East Asian (CHB+JPT) samples. The significance of a test for difference between the inbreeding coefficient of CEU and CHB+JPT bottlenecks is $P = 4.5 \cdot 10^{-5}$ (one-tailed two sample z-test), pointing to a significantly less intense bottleneck in the history of Europeans, compared to that in the history of East Asians. Fig. 2 depicts the maximum log-likelihood against the inbreeding coefficient.

	CEU	CHB+JPT
Ne before and after bottleneck	8712 ± 65	8695 ± 74
inbreeding coefficient *F*	**0.151 ± 0.009**	**0.201 ± 0.009**
time (kya)	32 ± 3	23 ± 2
p-value	$<< 10^{-12}$	$<< 10^{-12}$

Table 2: **Bottleneck modeling results**. Ne is based on 25 years per generation and on normalization by sequence heterozygosity. The time of the bottleneck is also normalized by sequence heterozygosity, but is independent of the number of years per generation. The inbreeding coefficient is independent of both the normalization and the number of years per generation, hence being the most robust model statistic. P-value is of testing the null hypothesis of no population bottleneck in the history of the population (one-tailed z-test of no change in effective population size).

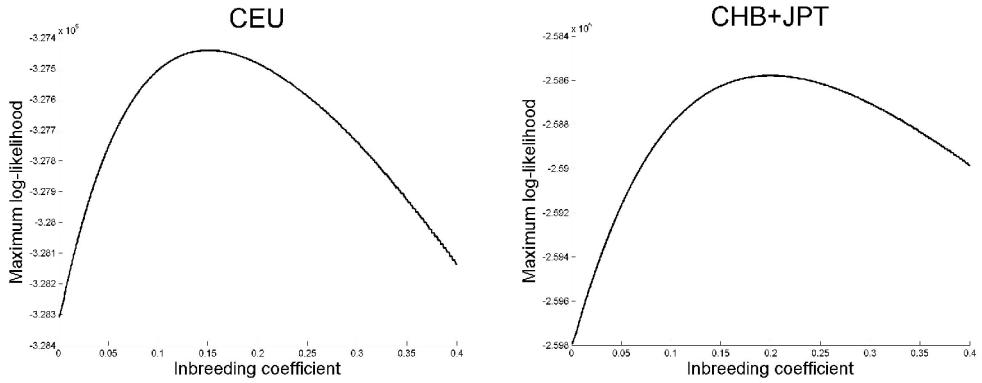


Figure 2: **Bottleneck likelihood profiling.** Maximum log-likelihood (maximized over the time of the bottleneck) versus the inbreeding coefficient, which is independent of both the normalization and the number of years per generation. The European and East Asian peaks are non-overlapping, with East Asians showing evidence of a more severe bottleneck (higher inbreeding coefficient).

Two-bottleneck model

The two-bottleneck model extends the above bottleneck model by allowing an additional bottleneck in the history of the modeled population. Table 3 presents the maximum likelihood estimates and standard deviations when this analysis is carried out independently in the European and East Asian samples.

There is no significant evidence for a difference in the inbreeding coefficient (P=0.81; two-tailed two sample z-test) or time (P=0.43; two-tailed two sample z-test) of the ancient bottleneck,

pointing to no significant difference in the characteristics of this bottleneck between the two populations (P=0.71; χ^2 with 2 degrees of freedom for any difference in the two parameters). The significance of a test for difference between the inbreeding coefficient of the recent bottleneck is P = 0.07 (one-tailed two sample z-test), somewhat supporting the conclusion that the East Asian recent bottleneck was significantly more intense than the European recent bottleneck, and is responsible for the greater genetic drift in East Asian than European populations since migrating from Africa. Last, the time of the recent bottleneck is estimated to be the same (P=0.41; two-tailed two sample z-test). Figs. 3 and 4 depict the maximum log-likelihood against the inbreeding coefficient of each of the bottlenecks.

While the CHB+JPT sample points to a more intense bottleneck, with smaller effective population size during the bottleneck, the model predicts the same effective population size before the bottlenecks for CHB+JPT and CEU, as expected by their shared history. This result further corroborates the modeling and the normalization by the different heterozygosity estimates.

	CEU	CHB+JPT
Ne before, after and between bottlenecks	10,085 ± 364	10,063 ± 310
time of ancient bottleneck (kya)	118 ± 21	98 ± 16
inbreeding coefficient 1	**0.264 ± 0.043**	**0.279 ± 0.039**
p-value	$3.6 \cdot 10^{-7}$	$3.9 \cdot 10^{-9}$
time of more recent bottleneck (kya)	18 ± 3	16 ± 2
inbreeding coefficient 2	**0.091 ± 0.016**	**0.123 ± 0.015**
p-value	$4.2 \cdot 10^{-7}$	$1.3 \cdot 10^{-12}$

Table 3: **Two-bottleneck modeling results**. Ne is based on 25 years per generation and on normalization by sequence heterozygosity. The time of each of the bottlenecks is also normalized by sequence heterozygosity, but is independent of the number of years per generation. The inbreeding coefficient of each of the bottlenecks is independent of both the normalization and the number of years per generation, hence they are the most robust model statistics; the only ones that are unaffected by normalization by heterozygosity and the fossil-based estimates of human-chimpanzee genetic divergence. P-values are obtained by testing, for each of the two modeled bottlenecks, whether it is consistent with the null hypothesis of no change in effective population size (one-tailed z-test).

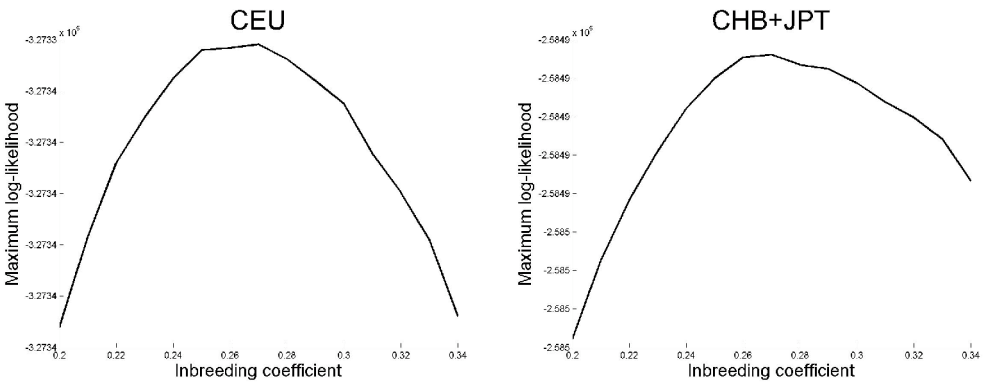


Figure 3: **Ancient bottleneck likelihood profiling.** Maximum log-likelihood versus the inbreeding coefficient of the ancient bottleneck (maximized over all other three parameters), which is independent of both the normalization and the number of years per generation. Inbreeding coefficients for the ancient bottleneck, estimated independently for the European and East Asian data sets, are consistent.

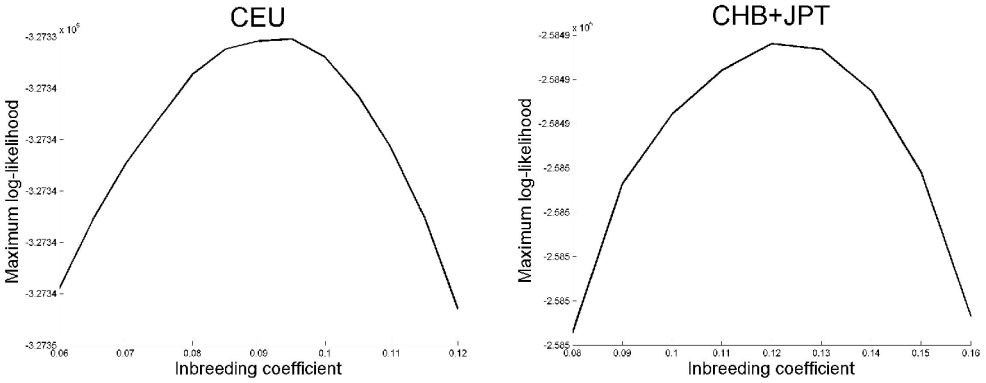


Figure 4: **Recent bottleneck likelihood profiling.** Maximum log-likelihood versus the inbreeding coefficient of the recent bottleneck (maximized over all other three parameters), which is independent of both the normalization and the number of years per generation. There is weak evidence of a stronger inbreeding coefficient in East Asians.

Additional models

We considered additional models that allow for a recent population expansion in Europe and East Asia, either expansion alone or following one or two bottlenecks. The aim of this modeling was to capture the profound population expansion that (based on the large current human population sizes and the archaeological record) we know must have occurred in the last ten thousand years of human history. The modeling estimates no recent expansion, that is, models of expansion do not fit the data significantly better than models of no expansion (data not shown).

How do these results reconcile with the obvious recent explosive increase in population size? A recent expansion, due to the small number of mutations since it took place, leaves a weak signal that affects only the very low end of the frequency spectrum. Since the number of chromosomes we study is small, 120 to 180 chromosomes depending on the population, this signal is not evident in the data. A larger sample size should detect the more recent expansions.

Divergence time modeling

The divergence time estimated by the joint-frequency spectrum modeling of CEU and CHB+JPT (Methods) is 17,484 ± 131 years ago when considering SNPs ascertained in Europeans and 17,498 ± 190 years ago when considering SNPs ascertained in East Asians, with the two estimates in good agreement. The standard deviations are underestimated since, when assuming the two-bottlenecks modeling results, we considered only the maximum likelihood estimates, without accounting for variation of the estimated bottlenecks. We emphasize, however, that in all one thousand different bootstraps in both data sets, the divergence time consistently matches the beginning of the recent bottleneck. Since the recent bottleneck time is the main factor in determining the divergence time, we concluded that the divergence time is 17 ± 3 kya, based on the range of the recent bottleneck in both populations (Table 3). We note that this standard deviation might still be an underestimate and that further modeling and simulations are necessary in order to accurately understand the divergence time.

Estimate of the total number of SNPs in human populations

Our modeling also allows us to estimate the total number of common SNPs (e.g., >1% or >5% minor allele frequency) in West Africans, European Americans and East Asian populations. To do this, we use the model that best fit each population (and that provides a significantly better fit to the data than models of lower dimension). Table 4 summarizes the results.

Population	Model	Number of SNPs (millions)	
		> 5%	> 1%
West African	two-epoch (expansion)	6.80	12.17
European	two-bottleneck	6.14	8.89
East Asian	two-bottleneck	5.86	8.20

Table 4: **Predicted number of common SNPs.** The number of autosomal SNPs of minor allele frequency greater than 5% and the number of autosomal SNPs of minor allele frequency greater than 1% are predicted based on the model that best fits the data for each population. The modeling is based on SNPs outside of CpG dinucleotides, and we multiplied by ~1.3 to include CpG SNPs.

Supplementary Note 2
Chr2p pilot data set

All SNPs on the p-arm of chromosome 2 (0-91.3Mb; NCBI human build 34) that were available in dbSNP at the time of marker picking (build 121) were attempted as part of a pilot of HapMap Phase 2, using the same genotyping technology later used for all of Phase 2 genotyping.[1,2] These SNPs, as opposed to other Phase 2 SNPs, were attempted regardless of whether they had been tried previously in Phase 1 and regardless of any information from the Hinds et al. data set[3]. Hence, this set of SNPs is relieved of several data collection complexities that affected the rest of the data: (1) replacing SNPs in complete LD with another SNP attempted in HapMap. (2) Replacing SNPs that were less than 5% minor allele frequency in all three Hinds et al. samples. (3) Balancing the different failure rates of Phase 1 and Phase 2 of HapMap (Methods).

We used the chr2p pilot data as a "gold standard" to validate the corrections we applied to our whole-genome data sets. Based on the available sequence traces, we could ascertain chr2p SNPs for three of our ascertainment libraries, one for each ancestry of interest (Table 1 in main text). The chr2p data for each ascertainment library consisted of SNPs successfully genotyped using HapMap Phase 2 genotyping technology. We applied exactly the same set of filters to the chr2p data that we applied to the whole-genome data set: (1) the standard HapMap quality control filters, (2) requiring that all SNPs be polymorphic in at least one sample, (3) requiring that all SNPs have $\geq 80\%$ genotyping completeness in all populations, (4) requiring that all SNPs not be in CpG dinucleotides, and (5) requiring that the ancestral allele was successfully determined and in agreement between orangutan and chimpanzees. Data from Phase 1 genotyping of chr2p was left out of the analysis to ensure that our chr2p data sets were truly homogeneous, with all SNPs discovered and genotyped in a uniform way.

Fig. 1 compares the derived allele frequency spectra between our whole-genome data sets and the chr2p data sets. While the chr2p plots are quite noisy due to the small number of SNPs, the allele frequencies match well. We tested for deviation of mean derived allele frequency of our whole-genome data from chr2p and found no significant deviation: P=0.250 for YRI, P=0.073 for CEU, P=0.815 for CHB and P=0.949 for JPT, each for ascertainment in a library of the same ancestry. Tables 1-5 reproduce the main results based on the chr2p data sets only. While some of the qualitative results still hold, we note that due to the small number of SNPs available for chr2p, these data alone cannot be used for robust inference about history.

We used information from chr2p also for a validation of the corrections we applied for SNPs not considered in HapMap Phase 2 based on information from the Hinds et al. data set[3] (Supp. Note 5), as well as for balancing the success rate of the Hinds et al. data set with that of HapMap Phase 2 (Methods).

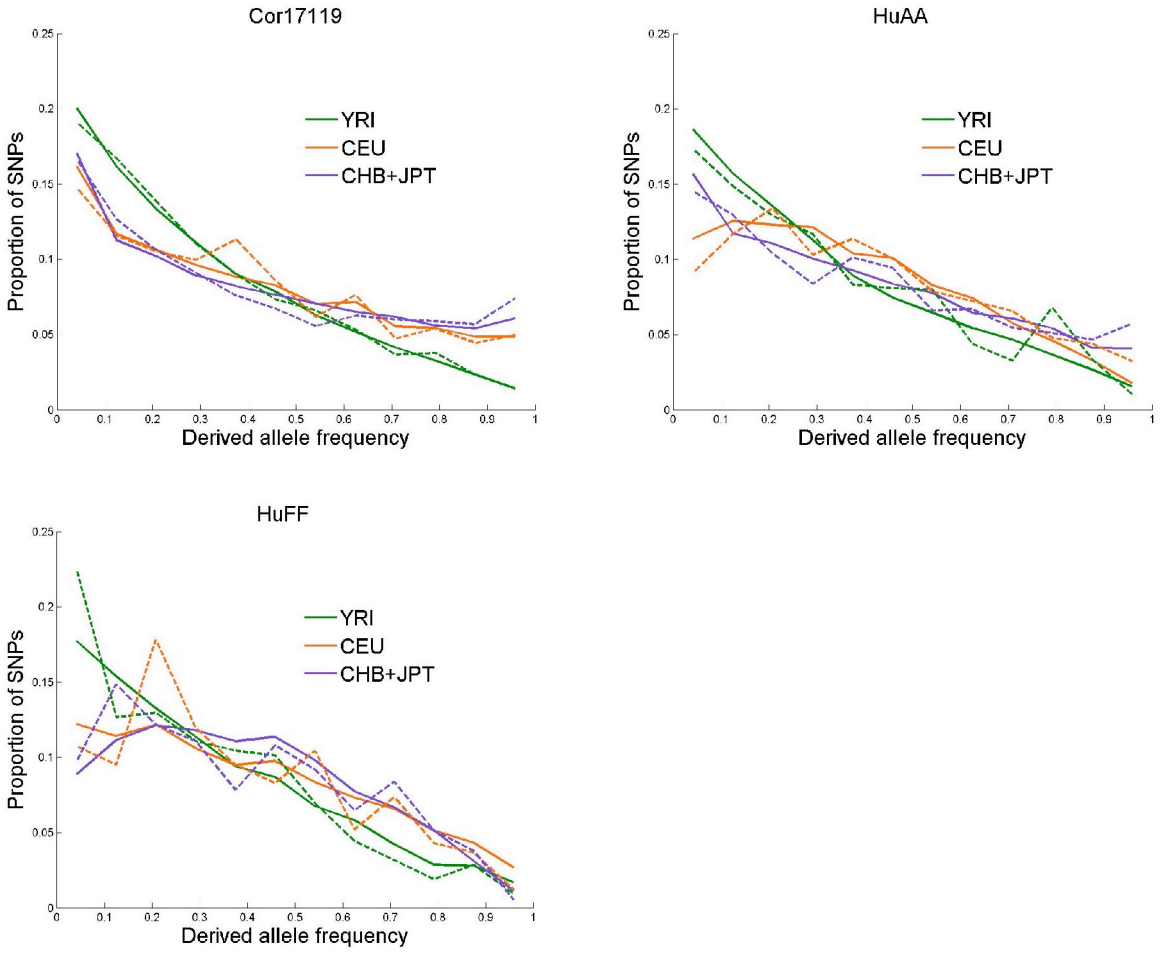


Figure 1: **Derived allele frequency spectra for whole-genome data sets and chr2p data sets.** The derived allele frequency spectrum for each of the HapMap samples based on the whole-genome data set (solid lines) and the chr2p data sets (dashed lines) for the three ascertainment libraries for which chr2p data were available. Chr2p frequency spectra are noisy since they are based on a small number of SNPs: 6,219 for Cor17119, 643 for HuAA and 186 for HuFF. The whole-genome data sets are based on 114,198, 18,006 and 5,250 SNPs, respectively.

	YRI	CEU	CHB+JPT
Tajima's D estimate	2.062	3.883	2.857
Expectation	2.740	2.703	2.954
Standard deviation	0.044	0.157	0.252
p-value	$<< 10^{-12}$	0.25	0.70

Table 1: **Tajima's D test based on chr2p data sets.**

Ascertainment library	F_{ST}(YRI, CEU)	F_{ST}(YRI,CHB+JPT)	p-value
Cor17119 (African American)	**0.158 ± 0.0026**	**0.208 ± 0.0033**	**<< 10^{-12}**
HuFF (East Asian)	0.151 ± 0.0137	0.178 ± 0.0145	0.092
HuAA (European American)	0.163 ± 0.0086	0.197 ± 0.0111	0.007

Table 2: **F_{ST} estimates between West Africans and non-Africans based on chr2p data sets.**

Ascertainment library	Observed F_{ST}	Expected F_{ST} for divergence after the recent bottleneck; p-value	Expected F_{ST} for divergence after the ancient bottleneck; p-value
Cor17119 (African American)	0.125 ± 0.0024		
HuFF (East Asian)	0.108 ± 0.0106	0.0328 ± 0.0040; $<10^{-10}$	0.2166 ± 0.0173; $<10^{-7}$
HuAA (European American)	0.123 ± 0.0071	0.0332 ± 0.0018; $<<10^{-12}$	0.2353 ± 0.0096; $<<10^{-12}$

Table 3: **F_{ST} estimates between Europeans and East Asians based on chr2p data sets.**

	YRI	CEU	CHB+JPT
Ne before change	9044 ± 208	132,190 ± 127,900	97,691 ± 124,270
Ne after change	14,748 ± 654	5,288 ± 59,362	4616 ± 7785
Change factor	**1.63 ± 0.09**	**0.04 ± 8.38**	**0.24 ± 1.10**
time of change (kya)	224 ± 29	1057 ± 475	626 ± 555
p-value	$<10^{-11}$	0.67	0.49

Table 4: **2-epoch modeling results based on chr2p data sets**.

	CEU	CHB+JPT
Ne before and after bottleneck	10,424 ± 1,460	8445 ± 874
inbreeding coefficient *F*	**0.344 ± 0.143**	**0.174 ± 0.124**
time (kya)	41 ± 21	28 ± 38
p-value	0.06	0.13

Table 5: **Single bottleneck modeling based on chr2p data sets**.

Supplementary Note 3
African American ascertainment

The two West African ascertainment libraries used in this study, Cor17109 and Cor17119, are in fact African Americans with some European ancestry. This poses a challenge to some of our analyses. While ideally the SNPs in these libraries would all be discovered in two chromosomes of African origin, in these libraries some will be discovered between one African chromosome and one European origin chromosome, or even two European origin chromosomes. The relative rates of each scenario will depend on the European ancestry proportions in these individuals.

We used the software ANCESTRYMAP[17] to obtain probabilities in Cor17109 for the number of chromosomes of European ancestry at each point of the genome, in one centimorgan resolution (data were not available to carry out the same analysis on Cor17119). ANCESTRYMAP implements a Markov Chain Monte Carlo algorithm to make these probabilistic assessments of ancestry. The assessments have been extensively validated in simulation, so that there is, for example, every reason to think that over regions assigned in a sample as having no European chromosomes (both are of African ancestry) with probability .95, we will make an incorrect assignment about 5% of the time.[17]

Based on the output of ANCESTRYMAP, we repeated all analyses that involved Cor17109 twice, restricting them to SNPs discovered in regions in which this individual is determined to have no European ancestry with either probability >.9 or probability >.95. Out of the 17,176 SNPs ascertained in this library, 15,031 and 13,951 are discovered in regions that with probability >.9 and >.95, respectively, have two African chromosomes. Based on ANCESTRYMAP, we also estimated that Cor17109 has about 4% European ancestry overall.

Table 1 revisits the West African Tajima's D results for all SNPs discovered in Cor17109 and for SNPs discovered after applying the above thresholds. In all cases, Tajima's D is significantly below expectation ($P<<10^{-12}$). Similarly, Table 2 presents the maximum likelihood estimates of the 2-epoch model for all SNPs discovered in Cor17109 and for SNPs discovered after applying the above thresholds. In all cases, the results point to about 1.8-fold population expansion of the West African sample, with the signal of expansion highly significant ($P<<10^{-12}$). Last, Table 3 revisits the F_{ST} results that are based on SNPs ascertained in Cor17109. The F_{ST} estimates do not vary substantially. The observation that East Asian allele frequencies are more differentiated from West Africans than Europeans is highly significant in all cases ($P<<10^{-12}$).

	Cor17109	Cor17109 > 0.9	Cor17109 > 0.95
Tajima's D estimate	2.273	2.272	2.300
Expectation	2.742	2.741	2.740
Standard deviation	0.020	0.020	0.021
p-value	$<<10^{-12}$	$<<10^{-12}$	$<<10^{-12}$

Table 1: **Tajima's D test**. The results of the Tajima's D test are presented for YRI using all SNPs in the Cor17109 library, using SNPs in regions with >.9 probability of two African chromosomes and using all SNPs in regions with >.95 probability of two African chromosomes. The expectation and standard deviation reported in the table are for a constant population size under the same ascertainment scheme as the data. P-value is for a two-tailed test of deviation from this expectation. The results for the full Cor17109 data set slightly differ from the ones presented in Supp. Table 1 since the results there are for pooling SNPs across both Cor17109 and Cor17119.

	Cor17109	Cor17109 > 0.9	Cor17109 > 0.95
Ne before change	8879 ± 112	8808 ± 122	8887 ± 133
Ne after change	16,008 ± 452	16,071 ± 494	16,601 ± 490
change factor	**1.80 ± 0.06**	**1.83 ± 0.07**	**1.76 ± 0.07**
time of change (kya)	211 ± 13	218 ± 14	219 ± 14
p-value	$<<10^{-12}$	$<<10^{-12}$	$<<10^{-12}$

Table 2: **2-epoch modeling results**. The 2-epoch modeling results are presented for the YRI sample using all SNPs in the Cor17109 library, using all SNPs in regions with >.9 probability of two African chromosomes and using all SNPs in regions with >.95 probability of two African chromosomes. Ne estimates are based on 25 years per generation and on normalization by sequence heterozygosity. The time of change is also normalized by sequence heterozygosity, but is independent of the number of years per generation. The change factor, which is the ratio of Ne after the change to Ne before the change, is independent of both the normalization and the number of years per generation. P-values are obtained by testing whether the change factor is different from 1 (two-tailed z-test). The results for the full Cor17109 data set slightly differ from the ones presented in Supp. Note 1 since the results there are for pooling SNPs across both Cor17109 and Cor17119.

Ascertainment library	F_{ST}(YRI, CEU)	F_{ST}(YRI,CHB+JPT)	F_{ST}(YRI,CHB+JPT) – F_{ST}(YRI, CEU)	p-value
Cor17109	0.146 ± 0.0026	0.179 ± 0.0033	0.0332 ± 0.0030	$<< 10^{-12}$
Cor17109 > 0.9	0.144 ± 0.0027	0.180 ± 0.0034	0.0355 ± 0.0033	$<< 10^{-12}$
Cor17109 > 0.95	0.143 ± 0.0029	0.179 ± 0.0037	0.0363 ± 0.0033	$<< 10^{-12}$

Table 3: **F_{ST} estimates between West Africans and non-Africans** using all SNPs ascertained in the Cor17109 library, using all SNPs ascertained in regions with >.9 probability of two African chromosomes and using all SNPs ascertained in regions with >.95 probability of two African chromosomes**.** P-values are obtained by testing whether F_{ST}(YRI,CHB+JPT) minus F_{ST}(YRI,CEU) is larger than zero (Supp. Note 10). The results for the full Cor17109 data set are reproduced from Supp. Table 2.

Supplementary Note 4
The potential confounding factor of migration

We were concerned that our inference of more East Asian genetic drift might reflect migration between African and European, after divergence of Europeans and East Asians. This would decrease F_{ST}(YRI, CEU), creating an artifactual impression of more East Asian drift.

To test whether the evidence for more drift on the East Asian than European lineage might be due to more migration between Europe and West Africa than between East Asia and West Africa since divergence, we assembled an independent data set based on comparison of millions of DNA sequence reads from unrelated West Africans, Europeans, East Asians and Biaka Pygmy chromosomes (Supp. Table 4). Aligning these reads, we counted the number of differences per base pair between individuals across populations (Methods). While migration would affect sequence divergence as well as the allele frequency spectrum, a population bottleneck occurring after the divergence would not be expected to affect sequence divergence. Table 1 presents the sequence divergence between each pair of populations, showing that the sequence divergence between Africa and Europe is similar to that between Africa and East Asia. In fact, European divergence is slightly higher, though not significantly ($P = 0.74$), providing no evidence at all of preferential European-African migration.

	Yoruba	Biaka Pygmy	European	East Asian
Yoruba	0.8359 ± 0.0048	0.8535 ± 0.0195	0.8345 ± 0.0033	0.8312 ± 0.0038
Biaka Pygmy		n/a	0.8259 ± 0.0175	0.8247 ± 0.0187
European			0.6044 ± 0.0038	0.6579 ± 0.0033
East Asian				0.5741 ± 0.0051

Table 1: **Sequence divergence and diversity estimates**. For each pair of populations, the sequence divergence is provided as the mean and standard deviation of the mean of the number of differences per 1,000 base pairs. The diagonal indicates the within-population sequence diversity (except for the Biaka Pygmy, where only one library is available so that no within-population heterozygosity estimate is possible).

To conservatively estimate the amount of uniquely European back migration to West Africa (above and beyond any East Asian back migration to West Africa) that is consistent with the sequence divergence data (Table 1), we allowed both the West Africa-Europe divergence and the West Africa-East Asia divergence to vary up to one standard deviation away from the mean. Our calculation considers *b* as the fraction of West African sites that are due to European migration, and thus of European ancestry. The observed West Africa-Europe sequence divergence is then a weighted mean, weighted by *b* and 1-*b*, of the European-European diversity and the "real" West Africa-Europe divergence. Based on the divergence estimates and on Europe-Europe diversity (Table 1), *b* is estimated to be < 0.0165, with the bound of 0.0165 corresponding to the case of West Africa-Europe divergence being a standard deviation below its estimated mean and the West Africa-East Asia divergence being a standard deviation above its estimated mean. A caveat is that we could not place a similar bound on the reverse migration pattern, that is, the amount of West African to European migration that is consistent with the data. The reason for this is that the West Africa-West Africa diversity is approximately the same as the West Africa-Europe divergence (Table 1). Thus, any West African to European migration would not be detectable from the sequence divergence data alone.

To test the robustness of our results on a migration rate of up to 1.65% between Europeans and West Africans (occurring after East Asian-European divergence), we reanalyzed our data sets while adding migration from East Asia to West Africa, to balance the possible effect of European to West African migration. We implemented this by replacing each genotype in the YRI sample, with probability 1.65%, with a random genotype from the CHB+JPT sample. The only result that might be sensitive to this type of migration is the F_{ST} with the YRI sample. Table 2 shows that this result did not change substantially with the introduction of migration, with the evidence of increased genetic drift among East Asians still highly significant.

Ascertainment library	F_{ST}(YRI, CEU)	F_{ST}(YRI,CHB+JPT)	F_{ST}(YRI,CHB+JPT) – F_{ST}(YRI, CEU)	p-value
Cor10470 (Biaka Pygmy)	**0.140 ± 0.0051**	**0.177 ± 0.0065**	**0.0364 ± 0.0060**	**$8.4 \cdot 10^{-10}$**
Cor17119 (African American)	**0.155 ± 0.0015**	**0.182 ± 0.0017**	**0.0273 ± 0.0017**	**$<< 10^{-12}$**
Cor17109 (African American)	**0.143 ± 0.0026**	**0.174 ± 0.0032**	**0.0316 ± 0.0029**	**$<< 10^{-12}$**
Cor11321 (East Asian)	0.148 ± 0.0021	0.153 ± 0.0018	0.0051 ± 0.0021	0.0072
HuFF (East Asian)	0.153 ± 0.0036	0.157 ± 0.0033	0.0031 ± 0.0037	0.2
HuAA (European American)	0.137 ± 0.0018	0.174 ± 0.0024	0.0367 ± 0.0024	$<< 10^{-12}$
Cor7340 (European American)	0.138 ± 0.0019	0.177 ± 0.0023	0.0383 ± 0.0023	$<< 10^{-12}$

Table 2: **F_{ST} estimates between West Africans and non-Africans, modeling in the confounding effect of migration.** The table is identical to Supp. Table 2, except that we randomly choose 1.65% of the "West African" chromosomes in our analysis to be sampled from the East Asians.

We similarly considered the effect of West Africa to Europe migration by reanalyzing the data while adding the same amount of migration from West Africa to East Asia. We note, however, that the amount of migration we used, b=0.0165, was estimated as an upper bound for migration in the other direction, from Europe to West Africa. The difference in the frequency spectrum between European and East Asian is still evident (Fig. 1; $P << 10^{-12}$). Tajima's D of the East Asian sample still significantly points to a contraction ($D = 3.442$; $P << 10^{-12}$).

Tables 3-5 show that the East Asian modeling results do not change considerably due to the introduction of migration. The difference between the European inbreeding coefficient and the East Asian inbreeding coefficient in the one bottleneck model also remains highly significant ($3.7 \cdot 10^{-4}$; one-tailed two sample z-test; Table 4). The results still significantly point to two bottlenecks in the history of East Asians (Table 5). There is still no significant evidence for a difference between East Asians and Europeans in the inbreeding coefficient (P=0.67; two-tailed two sample z-test; Table 5) or time (P=0.34; two-tailed two sample z-test; Table 5) of the ancient bottleneck, pointing to no significant difference in the characteristics of this bottleneck among the two populations (P=0.58; χ^2 with 2 degrees of freedom for any difference in the two parameters; Table 5). F_{ST} results did not change considerably, with the difference between Europeans and East Asians significant even after the introduction of migration (Table 6). Last, the European-East Asian divergence time is estimated to be 17,196 years ago, similar to the estimate without any migration (Supp. Note 1).

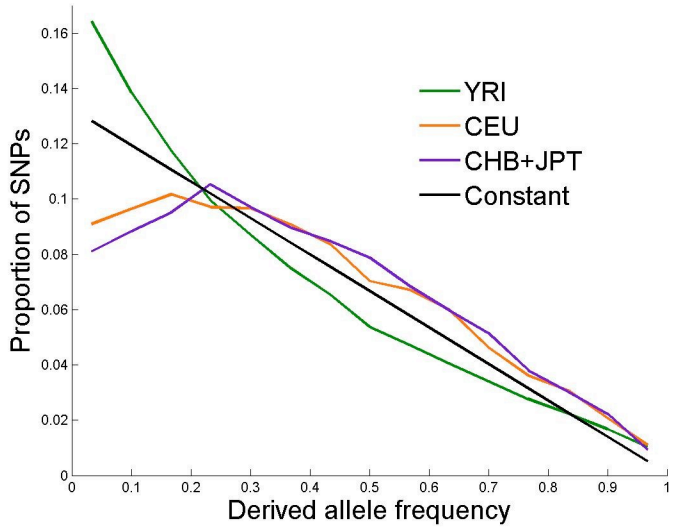


Figure 1: **Derived allele frequency spectra with West African migration to East Asia.** Derived allele frequency spectrum for each of the HapMap samples for ascertainment in libraries of similar ancestry, with 1.65% of the East Asian data representing African ancestry.

	Without migration	With migration
Ne before change	11,556 ± 1926	10,474 ± 824
Ne after change	4,465 ± 168	4,369 ± 154
change factor	**0.39 ± 0.03**	**0.42 ± 0.02**
time of change (kya)	208 ± 61	166 ± 34
p-value	$<< 10^{-12}$	$<< 10^{-12}$

Table 3: **2-epoch modeling results with West African migration to East Asia**. Maximum likelihood estimates for CHB+JPT are presented without migration (reproduced from Table 1 in Supp. Note 1) and with migration.

	Without migration	With migration
Ne before and after bottleneck	8695 ± 74	8641 ± 73
inbreeding coefficient *F*	**0.201 ± 0.009**	**0.193 ± 0.009**
time (kya)	23 ± 2	22 ± 2
p-value	$<< 10^{-12}$	$<< 10^{-12}$

Table 4: **Bottleneck modeling results with West African migration to East Asia**. Maximum likelihood estimates for CHB+JPT are presented without migration (reproduced from Table 2 in Supp. Note 1) and with migration.

	Without migration	With migration
Ne before, after and between bottlenecks	10,063 ± 310	9,728 ± 312
time of ancient bottleneck (kya)	98 ± 16	90 ± 20
inbreeding coefficient 1	**0.279 ± 0.039**	**0.239 ± 0.039**
p-value	$3.9 \cdot 10^{-9}$	$2.1 \cdot 10^{-7}$
time of more recent bottleneck (kya)	16 ± 2	16 ± 2
inbreeding coefficient 2	**0.123 ± 0.015**	**0.116 ± 0.019**
p-value	$1.3 \cdot 10^{-12}$	$4.8 \cdot 10^{-8}$

Table 5: **Two-bottleneck modeling results with West African migration to East Asia**. Maximum likelihood estimates for CHB+JPT are presented without migration (reproduced from Table 3 in Supp. Note 1) and with migration.

Ascertainment library	F_{ST}(YRI, CEU)	F_{ST}(YRI,CHB+JPT)	F_{ST}(YRI,CHB+JPT) – F_{ST}(YRI, CEU)	p-value
Cor10470 (Biaka Pygmy)	**0.143 ± 0.0054**	**0.176 ± 0.0063**	**0.0320 ± 0.0061**	**$8.3 \cdot 10^{-8}$**
Cor17119 (African American)	**0.158 ± 0.0015**	**0.181 ± 0.0017**	**0.0227 ± 0.0016**	**$<< 10^{-12}$**
Cor17109 (African American)	**0.146 ± 0.0026**	**0.173 ± 0.0032**	**0.0270 ± 0.0029**	**$<< 10^{-12}$**
Cor11321 (East Asian)	0.152 ± 0.0021	0.153 ± 0.0018	0.0016 ± 0.0020	0.22
HuFF (East Asian)	0.158 ± 0.0036	0.157 ± 0.0033	0.0001 ± 0.0037	0.51
HuAA (European American)	0.140 ± 0.0018	0.174 ± 0.0023	0.0332 ± 0.0023	$<< 10^{-12}$
Cor7340 (European American)	0.142 ± 0.0019	0.176 ± 0.0023	0.0348 ± 0.0023	$<< 10^{-12}$

Table 6: **F_{ST} estimates between West Africans and non-Africans with West African migration to East Asia.** The table is identical to Supp. Table 2, with the difference of incorporating 1.65% migration to the data.

Supplementary Note 5

Corrections and validations based on Hinds et al.

Hinds et al.[3] genotyped 1,586,383 SNPs in 71 unrelated individuals from three populations: 24 European Americans, 23 African Americans, and 24 Han Chinese. HapMap Phase 2 used these data for choosing which SNPs to genotype. First, they avoided genotyping SNPs of less than 5% minor allele frequency in all three samples of the Hinds et al. data set. 118,927 out of all SNPs considered for genotyping in HapMap Phase 2 were in this category. Second, they avoided SNPs for which another SNP in complete LD ($r^2 = 1$), according to the Hinds et al. data set, was considered for genotyping by HapMap. Specifically, a SNP was ignored if a SNP in complete LD was already considered in HapMap Phase 1 (122,174 SNPs). If no SNP in complete LD was considered in Phase 1, only one out of each $r^2 = 1$ equivalence class was considered in HapMap Phase 2, not attempting to genotype the rest (61,930 SNPs). In either case, whether a SNP in complete LD was considered in Phase 1 or Phase 2, it might have not been successfully genotyped.

The criteria applied by HapMap Phase 2 based on the Hinds et al. data set will introduce biases to our cleanly ascertained data sets, if not handled correctly, since the missing SNPs are of characteristically different allele frequencies. The SNPs that are less than 5% minor allele frequency in all three samples of the Hinds et al. data set would tend to be of either very low or very high derived allele frequency also in the HapMap samples. SNPs that have another SNP in complete LD according to the Hinds et al. data set would tend to have, on average, higher minor allele frequencies (a derived allele frequency closer to 0.5) since they tend to be due to mutations deep in the genealogical tree.

To correct the biases due to preferential missingness of SNPs of characteristic allele frequencies, we set out to estimate the derived allele frequency of each missing SNP as accurately as possible. For SNPs missing due to the consideration of a SNP in complete LD, we mined the HapMap data set for another SNP in complete LD (LD calculation based on the Hinds et al. data set). If such a SNP exists, and since it is in complete LD with the missing SNP, we assumed the allele frequencies of the missing SNP to be the same as the ones of that SNP. We noted that the $r^2 = 1$ criterion is not sensitive to which allele is the ancestral and which is the derived. Hence, derived allele frequencies of the missing SNP can either be the derived allele frequencies of the one in complete LD or the *ancestral* allele frequencies. We determined which case it is by the alleles that match across the two SNPs in the Hinds et al. data set. We replaced the allele frequencies of 5% of all SNPs in our data set by the allele frequencies of such a proxy, out of a total of 6.1% ignored due to being in complete LD with another SNP.

If a SNP in complete LD to the missing SNP has not been successfully genotyped in HapMap, we substituted the allele frequencies with the allele frequencies of that SNP in the Hinds et al. data set. We mapped the allele frequencies of the European American Hinds et al. sample to the CEU HapMap sample, of the African American sample to the YRI sample and of the Han Chinese sample to the CHB, JPT and the combined CHB+JPT samples. To avoid dependence between the ascertainment library and the sample, we ignored the Cor17109 sample, which is one of our ascertainment libraries, when considering the allele frequency of the Hinds et al.

African American sample. Using this mapping, we replaced the allele frequencies of 1.1% of all SNPs in our data sets by the allele frequencies of the SNPs in the Hinds et al. data set. We similarly substituted the allele frequencies with the allele frequencies in the Hinds et al. data set for SNPs missing due to being less than 5% minor allele frequency in all Hinds et al. samples. This class of SNPs contributed 0.4% of all SNPs in our data sets.

Replacing a SNP allele frequency by that of a HapMap SNP in complete LD might not be an accurate correction since LD is estimated to be complete by a different set of samples used in the Hinds et al. data set. As a consequence, the LD might be less than complete across the HapMap samples, resulting in different true allele frequencies of the two SNPs. In order to estimate the accuracy of the applied correction, we considered sets of SNPs in complete LD in the Hinds et al. data set that were all genotyped in HapMap (the elimination criterion was overridden by several other considerations[2]). For each equivalence class of such SNPs, we randomly chose one pair. In order to obtain a large sample of SNP pairs to accurately validate our correction, we considered the entire Hinds et al. and HapMap data sets, not restricting to SNPs ascertained in our libraries, resulting in 24,134 such pairs. We tested whether the pairs serve as good proxies for one another by estimating the correlation coefficient between the derived allele frequencies of the pairs. The correlation coefficient is 0.998 for the allele frequency of the CEU, CHB and JPT samples and 0.995 for the allele frequency of the YRI sample ($P<10^{-12}$), supporting the accuracy of the correction we applied.

A more important concern regarding these corrections has to do with the SNPs for which we substituted the allele frequencies with the allele frequencies in the Hinds et al. data set. These are SNPs omitted due to being in complete LD, but without a HapMap proxy available (1.1% of SNPs in our data sets) and due to being less than 5% minor allele frequency in all Hinds et al. samples (0.4% of SNPs in our data sets). To examine how good a proxy the Hinds et al. samples are to the HapMap samples, we considered all SNPs on chr2p that were successfully genotyped by HapMap Phase 2 and that would not have been genotyped due to these criteria if they had not been on chr2p (Supp. Note 2). There are 2,299 such SNPs that would have been omitted due to complete LD with a HapMap Phase 1 SNP; 1,152 such SNPs that would have been omitted due to complete LD with another SNP considered for Phase 2; and 413 such SNPs that would have been omitted due to the <5% criterion. These SNPs allow us to compare the allele frequencies mapped from the Hinds et al. allele frequencies with the allele frequencies of the HapMap genotyping.

Table 1 details the differences between the HapMap derived allele frequencies and the ones mapped from the Hinds et al. data set. The correlation coefficient of both is very high, always above 0.93. We further studied whether there are consistent biases of either of the two derived allele frequencies being higher. Such significant deviation is observed for the YRI sample, but while the deviation is significant, it is very small: The derived allele frequency of a SNP that would have been omitted due the complete LD criterion in the YRI sample is expected to be, on average, 0.012 lower than in the Hinds et al. African American sample. On the other hand, the derived allele frequency of a SNP that would be omitted due to the <5% criterion is expected to be, on average, 0.018 higher. We also observed a significant deviation in the other samples for SNPs omitted due to the <5% criterion (Table 1). For these, the difference is even smaller, with the HapMap sample derived allele frequency expected to be, on average, 0.006, 0.003, and 0.004

higher than the corresponding Hinds et al. sample, for CEU, CHB, and JPT, respectively. We further note that this deviation affects only about 0.4% of SNPs in our data set, and hence would not have a large effect on our results. We conclude that although the corrections described in this note are not exact, they overcome the obstacles to using HapMap for studies of demographic history.

Elimination criterion	CEU	CHB	JPT	YRI
Complete LD with a HapMap Phase 1 SNP	r = 0.95 0.002 ± 0.0735 P=0.34	r = 0.96 -0.004 ± 0.0738 P=0.08	r = 0.94 -0.003 ± 0.0908 P=0.22	r = 0.93 -0.012 ± 0.0914 $P<10^{-5}$
Complete LD with a HapMap Phase 2 SNP	r = 0.95 -0.0009 ± 0.0707 P=0.52	r = 0.96 0.001 ± 0.0726 P=0.49	r = 0.94 0.0001 ±0.0901 P=0.95	r = 0.93 -0.013 ± 0.0895 $P<10^{-12}$
Less than 5% minor allele frequency	r = 0.993 0.006 ± 0.0319 $P<10^{-4}$	r = 0.998 0.003 ± 0.0182 $P<10^{-3}$	r = 0.996 0.004 ± 0.0259 P=0.002	r = 0.98 0.018 ± 0.0532 $P<10^{-10}$

Table1: **Comparison of Hinds et al. mapped derived allele frequencies with HapMap derived allele frequencies.** For chr2p SNPs that otherwise would not have been genotyped, the table details a comparison of the Hinds et al. mapped allele frequencies with the HapMap allele frequencies. For each elimination criterion and for each HapMap sample, the correlation coefficient r of the two allele frequencies is indicated, followed by the mean and standard deviation of the HapMap allele frequency minus the Hinds et al. one and by a p-value of a t-test for the difference in the mean allele frequency of the two.

Supplementary Note 6
Impact of length of bottleneck on modeling

The effect of a bottleneck on the frequency spectrum depends primarily on the inbreeding coefficient, the ratio of its length to (twice) the effective population size during the bottleneck. Hence, we modeled a bottleneck using two parameters, the time of the bottleneck and its inbreeding coefficient. We implemented this by keeping the number of generations of the bottleneck fixed on a predefined value of T=100 and varying the effective population size to match the desired inbreeding coefficient values.

This note tests to what extent the length of the bottleneck and the effective population size individually, rather than their ratio, affect the frequency spectrum and as a consequence our results. We reanalyzed our data with different values for the number of generations of the bottleneck, scaling the effective population size as the inverse of the number of generations to keep the inbreeding coefficient constant. Table 1 presents the maximum likelihood estimates for the bottleneck model for T=10 and T=50. For comparison, the results with T=100 are reproduced. As evident from the table, the results are not sensitive to the choice of T (very small differences are due to rounding errors). Furthermore, the significance of the difference between CEU and CHB+JPT inbreeding coefficients, which is P = $4.5 \cdot 10^{-5}$ for T=100 (Supp. Note 1), is P = $6.6 \cdot 10^{-5}$ and P = $9.2 \cdot 10^{-5}$ for T=50 and T=10, correspondingly. These results show that both the estimated bottleneck intensity and the conclusion of a more severe East Asian bottleneck are robust to the exact details of bottleneck modeling.

	CEU			CHB+JPT		
	T=100	T=50	T=10	T=100	T=50	T=10
Ne before and after bottleneck	8712 ± 65	8711 ± 64	8710 ± 61	8695 ± 74	8688 ± 78	8686 ± 82
inbreeding coefficient *F*	**0.151 ± 0.009**	**0.148 ± 0.008**	**0.147 ± 0.008**	**0.201 ± 0.009**	**0.197 ± 0.010**	**0.198 ± 0.011**
time (kya)	32 ± 3	32 ± 3	32 ± 3	23 ± 2	23 ± 2	23 ± 2
p-value	$<< 10^{-12}$	$<< 10^{-12}$	$<< 10^{-12}$	$<< 10^{-12}$	$<< 10^{-12}$	$<< 10^{-12}$

Table 1: **Bottleneck modeling results with different bottleneck lengths**. The results for T=100 are reproduced from Table 2 in Supp. Note 1.

Supplementary Note 7

Derived allele state accounting for recurrent mutation

The power of a study of demographic history is substantially increased by considering the unfolded allele frequency spectrum—that is, by knowing which is the ancestral allele, and which is the new mutation. However, it can be difficult to accurately determine the ancestral allele, due to the possibility of a mutation recurring in the same nucleotide. We determined the ancestral state as reliably as possible by removing all SNPs in hypermutable CpG dinucleotides and by considering both the chimpanzee and orangutan alleles. We aligned chimpanzee and orangutan traces across human SNPs and found that the chimpanzee and orangutan alleles mismatch for over 7% of human SNPs, stressing the importance of accounting for recurrent mutations. We substantially decreased the probability of a SNP in our data sets being due to recurrent mutations by considering only SNPs for which both chimpanzee and orangutan sequences were available and in agreement. Using two species, closely related to humans, for determination of the ancestral allele constitutes a great advantage over using only one as the latter strategy is vulnerable to errors due to the same mutation recurring twice, once within human population history and once on the lineage of the species used for determination of the ancestral allele.

The determination of which allele is the new mutation, by comparison to chimpanzee and orangutan and excluding CpG dinucleotides, is still vulnerable to errors due to recurrent mutations. The first type of error is due to a mutation that recurs three times, once on the orangutan lineage, once on the chimpanzee lineage, and once within human history. For such a mutation the chimpanzee and orangutan alleles agree, but match the human derived allele. The probability of the same mutation occurring three times is small, however, and we ignore this possibility. The second type is a mutation that recurs twice along the human lineage, with the second mutation occurring within human population history and reverting the base back to its ancestral state. For such a mutation, the allele that matches the chimpanzee and orangutan alleles is the derived allele of the second, more recent, mutation. In the remainder of this note, we estimate the probability of a SNP in our data sets to be an outcome of a recurrent mutation of this second type and we show that these errors do not considerably change our results.

To estimate the probability that a SNP in our data sets results from recurrent mutations, we took into account our SNP ascertainment scheme. Since we ascertained in two chromosomes, a recurrent mutation is one that has occurred twice: once, after the divergence of the two human chromosomes that are compared, and once before the divergence of the two human chromosomes (but after divergence from chimpanzee). Since the SNP considered is also polymorphic in a larger sample (we ignored SNPs that were monomorphic across samples), and only two alleles are observed, we assumed the divergence time of the human samples to be one million years ago (1/7 of human-chimpanzee divergence assuming human-chimpanzee genetic divergence is about 7 million years ago). This assumption is an overestimate given that the average divergence of two chosen chromosomes is typically estimated to be 500,000 years. By allowing more historical opportunity for mutation, we obtain a conservative overestimate of the recurrent mutation rate.

Using this framework, and assuming a constant mutation rate across sites, we estimated a probability of a recurrent mutation of 0.16% to 0.22% (depending on the exact transition:transversion ratio assumed). We note that the assumption of a constant mutation rate across nucleotides is certainly inaccurate, even after removing hypoermutable CpG dinucleotides, which is anti-conservative for our analysis as it results in an underestimate of the recurrent mutation rate. We therefore also carried out an analysis using the recurrent mutation rates estimated empirically from data by Patterson et al[18]. Two sequence alignment strategies were considered in that paper, with five and six primate species, and the recurrent mutation rate (excluding CpG dinucleotides) was separately estimated for each using an expectation maximization algorithm. We normalized the sum of branch lengths in each of these trees to match the one in the three species tree considered here (human, chimpanzee and orangutan). The resulting estimates for the probability of a recurrent mutation are 0.26% based on the five species tree and 0.15% based on the six species tree. Combining these independent estimates with the above estimate based on a constant mutation rate, we concluded that the probability of a recurrent mutation affecting any given SNP in our data set is smaller than 0.26%.

Recurrent mutations can be accounted for in our analyses by considering the frequency spectrum as partially folded – considering that for some SNPs the derived allele is actually the ancestral one and vice versa. Let l be the probability of misclassifying the alleles due to a recurrent mutation and let $P_n(j)$ denote the expected demography-dependent probability of observing j derived alleles out of n. Then, the expected probability of observing j alleles classified as derived out of n is actually $(1-l)P_n(j)+lP_n(n-j)$. This change to the theoretical frequency spectrum amounts to a simple modification of the likelihood formulation, without any additional changes to the modeling and inference: Writing the likelihood function as $L=\prod_{i=1}^{s}L_i$ (Supp. Methods), we accounted for the probability of recurrent mutations by changing the likelihood function to be $\prod_{i=1}^{s}\left[(1-l)L_i+lL_i^{'}\right]$, where $L_i^{'}$ is identical to L_i but replacing the number of derived alleles with the number of ancestral alleles.

We repeated the main analyses of this paper while considering that the derived allele is misclassified for 0.26% of our SNPs (l=0.0026), the worst of the aforementioned probability estimates. Tables 1-3 present the maximum likelihood estimates for l=0.0026 and compare them with the results without any folding (l=0), reproduced from Supp. Note 1. For the former, only maximum likelihood estimates are reported, without bootstrapping-based standard deviations. For all the different models and parameters, accounting for recurrent mutations does not change the results significantly, with the maximum likelihood estimate always being within one standard deviation of the results reported in the paper. Importantly, the differences between Europeans and East Asians are still evident.

	CEU		CHB+JPT	
	l=0	*l*=0.0026	*l*=0	*l*=0.0026
Ne before change	23,344 ± 18,583	12,547	11,556 ± 1926	10,884
Ne after change	5,634 ± 127	5,646	4,465 ± 168	4,462
change factor	**0.35 ± 0.14**	**0.45**	**0.39 ± 0.03**	**0.41**
time of change (kya)	506 ± 254	354	208 ± 61	188

Table 1: **2-epoch modeling results with partial folding**.

	CEU		CHB+JPT	
	l=0	*l*=0.0026	*l*=0	*l*=0.0026
Ne before and after bottleneck	8712 ± 65	8681	8695 ± 74	8666
inbreeding coefficient *F*	**0.151 ± 0.009**	**0.147**	**0.201 ± 0.009**	**0.197**
time (kya)	32 ± 3	31	23 ± 2	23

Table 2: **Bottleneck modeling results with partial folding**.

	CEU		CHB+JPT	
	l=0	*l*=0.0026	*l*=0	*l*=0.0026
Ne before, after and between bottlenecks	10,085 ± 364	9,961	10,063 ± 310	9,916
time of ancient bottleneck (kya)	118 ± 21	118	98 ± 16	95
inbreeding coefficient 1	**0.264 ± 0.043**	**0.25**	**0.279 ± 0.039**	**0.26**
time of more recent bottleneck (kya)	18 ± 3	19	16 ± 2	16
inbreeding coefficient 2	**0.091 ± 0.016**	**0.09**	**0.123 ± 0.015**	**0.12**

Table 3: **Two-bottleneck modeling results with partial folding**.

Supplementary Note 8
Testing inference procedure by simulation

To test the maximum likelihood inference framework that we applied (Methods and Supp. Methods) and its accuracy, as well as the bootstrapping method and hypothesis testing procedure, we generated data sets by coalescent simulations under different demographic histories[5] and estimated the demographic parameters using our inference framework.

For each demographic history considered in this note, we simulated one thousand chromosomes. For each SNP we then randomly chose two chromosomes to mimic the ascertainment scheme used in our data sets: If the two samples carry the same allele, the SNP is ignored. Otherwise, 120 samples are randomly chosen out of the remaining samples to constitute the data for this SNP. Finally, each chromosomal sample is randomly removed with probability 2%, to simulate genotyping failure. The procedure is repeated until 25,000 SNPs are sampled. We note that this is the same order of magnitude as the number of SNPs studied in our paper (Table 1 in main text).

Each such data set was generated twice: (1) SNPs are in linkage equilibrium and hence independent. This was achieved by considering at most one SNP per simulated genealogy. (2) SNPs are highly dependent locally, which is achieved by considering each simulated SNP as a block of several SNPs in complete LD. The number of SNPs in a block is governed by a Poisson distribution with mean 5 (that is, the data set consists of 25,000 SNPs, but on average, there are 5 perfect copies of each SNP). In practice, the data sets analyzed in the paper had weak linkage disequilibrium, which put them in between these two scenarios (see Supp. Table 5 for the median distance between each pair of SNPs in our data sets).

We tested the accuracy of inference for the 2-epoch model by simulating an expanding population. The simulations assumed that 400kya the population expanded 2-fold. The maximum likelihood estimates based on a sample generated by these simulations, by bootstrapping over one thousand data sets, point to a 1.97 ± 0.04-fold expansion that occurred 417 ± 28 kya. For the correlated data set, the estimate is of a 2.15 ± 0.12-fold expansion that occurred 514 ± 96 kya. Both analyses are thus consistent with the simulation parameters, and correctly infer an expansion ($P << 10^{-12}$ for both data sets by a two-tailed z-test).

To similarly test the inference of the bottleneck model, we simulated a population that experienced a bottleneck 40kya, with an inbreeding coefficient of 0.2. The maximum likelihood estimates are of a bottleneck with an inbreeding coefficient of 0.19 ± 0.008, estimated to have occurred 41.5 ± 1.9 kya. The corresponding estimates for the correlated data set are of an inbreeding coefficient of 0.18 ± 0.019 and a bottleneck time of 44.7 ± 4.3 kya. Both analyses are consistent with the simulation parameters, and correctly reject the null hypothesis of no population bottleneck: $P << 10^{-12}$ by a one-tailed z-test of no change in effective population size.

To test whether our inference procedure is robust to overfitting, we applied the two-bottleneck model to the one-bottleneck data sets. This model estimates the first bottleneck to be significant

(P=0.02). However, the second bottleneck is estimated to have an inbreeding coefficient 0.06 ± 0.066, consistent with no change in population size (P=0.38; an inbreeding coefficient of 0.005 denotes no change in population size). For the correlated data set, the two-bottleneck model similarly estimates the first bottleneck to be significant ($P = 4.3 \cdot 10^{-6}$), and again the second bottleneck is consistent with no change in population size (inbreeding coefficient of 0.02 ± 0.030; P=0.60).

To test the more detailed two-bottleneck model, we simulated a population that experienced two bottlenecks in its recent history. The first bottleneck was simulated to be of inbreeding coefficient 0.2 and to take place 60kya, and the second bottleneck was simulated to be of inbreeding coefficient 0.1 and to have taken place 20kya. The maximum likelihood estimates are of an ancient bottleneck of intensity 0.21 ± 0.017 and a recent bottleneck of intensity 0.07 ± 0.015, consistent with the simulation parameters. The bottlenecks are estimated to have occurred 52 ± 6 and 21 ± 2 kya, respectively, with both being significant ($P << 10^{-12}$ and $P = 6.9 \cdot 10^{-6}$). For the correlated data set, we estimate inbreeding coefficients of 0.19 ± 0.056 and 0.11 ± 0.044, and dates of 64 ± 23 and 23 ± 5 kya, with both being significant ($P = 5.9 \cdot 10^{-4}$ and P=0.01).

We next applied the one-bottleneck model to these two-bottleneck data sets. The model estimates a significant bottleneck ($P << 10^{-12}$) with an inbreeding coefficient of 0.22 ± 0.009 occurring 32 ± 2 kya. For the correlated date, the inbreeding coefficient is estimated to be 0.22 ± 0.020 and the time is estimated to be 31 ± 3 kya ($P << 10^{-12}$). We note that this inferred bottleneck combines the two simulated bottlenecks into one event that occurred between the two events, as we hypothesized might be the case for the modeling of the European and East Asian data sets.

For all three models, the maximum likelihood estimates match the true parameters well. Out of 16 different parameters across the different models and the two types of data sets, the most significant deviation from the true value is 2 standard deviations. Although this deviation is borderline significant (P=0.0455; two-tailed z-test), correcting for the 16 different hypotheses tested results in a non-significant deviation from the true value (P=0.73). We conclude that the estimates resulting from our maximum likelihood framework are unbiased.

As a final test for the possibility of overfitting, we considered a data set generated by simulations of a population that was of constant size throughout history. Fitting the bottleneck model, the estimated inbreeding coefficient is 0.005 ± 0.003, implying no change in effective population size during the "bottleneck" event (P=0.9) (0.005 ± 0.007; P=1 for the data set with LD). Similarly, fitting the two-bottleneck model, both bottlenecks are consistent with no change in effective population size (P=0.84 and P=0.96 for the data set in linkage equilibrium; P=0.58 and P=0.66 for the data set with LD).

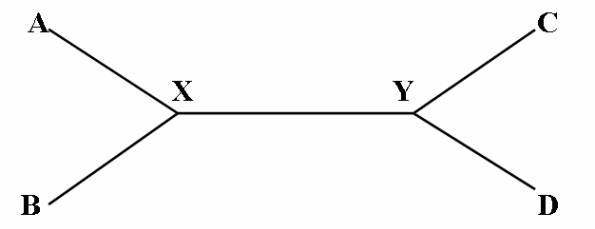
Supplementary Note 9
Quartet test for migration

This note describes a test that is an ingredient in new methods for modeling demographic history within a species. These methods are also described in ref. [19].

Our central idea is to consider 4 subpopulations, each supposed to be (approximately) homogeneous. We attempt to build a 'phylogenetic tree' with these 4 populations as leaves. With 4 populations {A,B,C,D} there are 3 labeled unrooted trees with clades

{(A,B),(C,D)}, {(A,C),(B,D)}, {(A,D),(B,C)}

respectively. Consider the tree:



where we have introduced internal nodes X and Y. Let f_A, f_B, f_C and f_D be population frequencies for a certain allele. Then, f_A-f_B has mean 0 and variance that depends on the demographic history of A and B since the split from X. Furthermore, under our assumptions

$$E[(f_A-f_B)(f_C-f_D)] = 0$$

We implemented this idea by computing sample frequencies $\widehat{f}_A$ etc. for a set of alleles and then evaluating $\rho(A,B;C,D)$ as the correlation coefficient between f_A-f_B and f_C-f_D. We tested the significance of ρ based on a block jackknife[20] (two-tailed z-test).

We repeat the test for the 3 possible trees. When using a substantial number of SNPs, usually at most one of the trees does not result in a highly significant correlation. If the correlation for all 3 trees is significant, this is evidence that our basic demographic model is overly simplistic, and that more complex models with migration or admixture are needed to explain the data.

We applied this quartet test on data from 250K Affymetrix SNP array using the YRI, CEU and CHB samples from HapMap (Methods) and 12 samples from a population of Mbuti Pygmies (to restrict our analysis to SNPs with almost no missing genotypes, we applied the test to a subset of 111,604 SNPs out of the 250K). Our main interest here is to test for migration since the Out-of-Africa event between African and non-African populations. The results show that only one tree is statistically acceptable and that we have failed to detect any migration between either CEU or CHB on the one hand and either Mbuti Pygmy or YRI on the other (Table 1).

Clade	ρ	p-value
(CEU, YRI) , (CHB, Pygmy)	0.524	$<< 10^{-12}$
(CEU, Pygmy) , (CHB, YRI)	0.512	$<< 10^{-12}$
(CEU, CHB) , (YRI, Pygmy)	-0.004	0.21

Table 1: **Quartet test results for the possible 3 clades**. The p-value is for a two-sided z-test.

This test strongly suggests that migration between YRI and CEU or CHB has been at a low level since the population split, and has not seriously distorted the joint allele frequency spectrum. We note that the HapMap CEU population is Northern European in origin. Hence, this result does not exclude some level of migration between Africa and Southern Europe, which has been suggested from analysis of Y-chromosome data[21].

Our quartet test seems to be new, but there is certainly a relationship with 'quartet puzzling'[22] and other quartet methods[23] that have been applied to determining phylogenetic trees at a species level. The issues here are somewhat different, as we can expect that species will be sufficiently divergent that differences will fix.

Supplementary Note 10
F_{ST} theory

This note describes theory related to F_{ST} and 'genetic drift' between populations. Suppose we have a biallelic marker with a variant allele with population frequency p_i in two populations (i = 1; 2). Set $q_i = 1 - p_i$. Then, we can define F_{ST} as $F_{ST} = N/D$, where $N = p_1(q_2 - q_1) + p_2(q_1 - q_2)$ and $D = p_1 q_2 + q_1 p_2 = N + p_1 q_1 + p_2 q_2$. This is not a standard definition, but is appropriate for our purposes here. This is a definition of F_{ST}, a parameter measuring divergence at a given locus, not a sample statistic.

Given k markers, let $N^{[i]}$, $D^{[i]}$ be defined as above for i=1, 2,.., k. F_{ST} can then be defined as $F = N/D$, where $N = \sum_{i=1}^{k} N^{[i]}$ and $D = \sum_{i=1}^{k} D^{[i]}$. Again, F is a parameter, not a statistic from data. Suppose we ascertain in population 1, or another population where there has been no significant gene flow with population 2 since the divergence with 1. At a given marker with variant allele frequencies p_1, p_2 as above, let z be the allele frequency in the root population. Then, $E(p_2 \mid z) = z$. Let $p_1^{[i]}, p_2^{[i]}, z^{[i]}$ be the corresponding frequencies for marker i. Then it follows that

$$E(D \mid p_1^{[i]}, z^{[i]}) = \sum_{i=1}^{k} \left(p_1^{[i]} + z^{[i]} - 2 p_1^{[i]} z^{[i]} \right)$$

which is in independent of the choice of population 2. A well known result (chapter 13 of ref. [24]) shows that

$$E((p_2 - z)^2 \mid z) = z(1-z)(1 - e^{-2N_2 t})$$

where t is the time back to divergence and N_2 is the effective population size of population 2 since divergence. We can write $(p_1 - p_2)^2 = (p_1 - z)^2 + (p_2 - z)^2 - 2(p_1 - z)(p_2 - z)$. The last term in this equation has zero mean.

If we consider another population 3 that also diverged from population 1 without significant gene flow since divergence, and define t to be the time since the older of the two divergences, then it follows that the difference F(1,2) – F(1,3) between F_{ST} for populations 1 and 2 on the one hand and 1 and 3 on the other has expected value

$$F(1,2) - F(1,3) = \left(\frac{\sum_{i=1}^{k} z^{[i]}(1 - z^{[i]})}{\sum_{i=1}^{k} \left(p_1^{[i]} + z^{[i]} - 2 p_1^{[i]} z^{[i]} \right)} \right) \left(e^{-2N_3 t} - e^{-2N_2 t} \right)$$

We use the following estimators for $N^{[i]}$, $D^{[i]}$:

$$\widehat{N}^{[i]} = \left(a_1/n_1 - a_2/n_2\right)^2 - h_1/n_1 - h_2/n_2$$
$$\widehat{D}^{[i]} = \widehat{N}^{[i]} + h + h'$$

where a_i, n_i are the allele counts and total number of alleles for population i and h_i is the heterozygosity estimate for population i:

$$h_i = \frac{a_i(n_i - a_i)}{n_i(n_i - 1)}$$

We have not found these exact estimators in the literature. By the Lehman-Scheffé theorem[25], $\widehat{N}^{[i]}$ and $\widehat{D}^{[i]}$ are uniformly minimum variance unbiased estimators. When the sample sizes n_1, n_2 are equal these estimators reduce to the estimators of F_{ST} given by Weir and Hill[26]. Following these estimators,

$$\widehat{F}(1,2) = \frac{\sum_{i=1}^{k} \widehat{N}^{[i]}}{\sum_{i=1}^{k} \widehat{D}^{[i]}}$$

with a similar expression for $\widehat{F}(1,3)$. We can then test for F(1,2) - F(1,3) > 0 using moving block bootstrap[20], with a significant result implying that N_2 is smaller than N_3. This argument applies whatever the ascertainment scheme, provided that it only involves populations not relevant to gene flow in populations 2 and 3 since the split from 1.

References

1. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
2. The International HapMap Consortium. The Phase II HapMap. *Submitted* (2007).
3. Hinds, D.A. et al. Whole-genome patterns of common DNA variation in three human populations. *Science* **307**, 1072-9 (2005).
4. Efron, B. *The Jacknife, the Bootstrap, and Other Resampling Plans*, (SIAM, Philadelphia, PA, 1982).
5. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-8 (2002).
6. Marth, G.T., Czabarka, E., Murvai, J. & Sherry, S.T. The allele frequency spectrum in genome-wide human variation data reveals signals of differential demographic history in three large world populations. *Genetics* **166**, 351-72 (2004).
7. Reich, D.E. et al. Linkage disequilibrium in the human genome. *Nature* **411**, 199-204 (2001).
8. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
9. Smith, M.W. et al. A high-density admixture map for disease gene discovery in African Americans. *Am J Hum Genet* **74**, 1001-13 (2004).
10. Parra, E.J. et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* **63**, 1839-51 (1998).
11. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-95 (1989).
12. Weir, B.S., Cardon, L.R., Anderson, A.D., Nielsen, D.M. & Hill, W.G. Measures of human population structure show heterogeneity among genomic regions. *Genome Res* **15**, 1468-76 (2005).
13. Lahiri, S.N. *Resampling methods for dependent data*, (Springer, New York, 2003).
14. Garrigan, D. & Hammer, M.F. Reconstructing human origins in the genomic era. *Nat Rev Genet* **7**, 669-80 (2006).
15. Harding, R.M. et al. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet* **60**, 772-89 (1997).
16. Yu, N. et al. Global patterns of human DNA sequence variation in a 10-kb region on chromosome 1. *Mol Biol Evol* **18**, 214-22 (2001).
17. Patterson, N. et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* **74**, 979-1000 (2004).
18. Patterson, N., Richter, D.J., Gnerre, S., Lander, E.S. & Reich, D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* **441**, 1103-8 (2006).
19. Becquet, C., Patterson, N., Stone, A., Przeworski, M. & Reich, D. Genetic structure of chimpanzee populations. *PLoS Genet* **in press**(2007).
20. Kunsch, H.R. The jackknife and the bootstrap for general stationary observations *The Annals of Statistics* **17**, 1217-1241 (1989).
21. Semino, O. et al. Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. *Am J Hum Genet* **74**, 1023-34 (2004).
22. Strimmer, K. & van Haeseler, A. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. *Mol Biol Evol* **13**, 964-969 (1996).

23. Estabrook, G.F., McMorris, F.R. & Meacham , C.A. Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. *Syst Zool* **34**, 193-200 (1985).
24. Nei, M. *Molecular evolutionary genetics*, (Columbia University Press, New York, 1987).
25. Bickel, P.J. & Doksum, K.A. *Mathematical statistics: Basic Ideas and selected topics*, (Holden-Day, San Francisco, 1977).
26. Weir, B.S. & Hill, W.G. Estimating F-statistics. *Annu Rev Genet* **36**, 721-50 (2002).