Supplementary Information

Accelerated genetic drift on chromosome X during the human dispersal out of Africa

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

Nature Genetics: doi:10.1038/ng.303

Supplementary Methods

Screening out regions known to be affected by natural selection: For all analysis (both of allele frequency and sequence diversity data), we excluded sections of the genome where there was a strong prior probability of natural selection. We removed exons and conserved non-coding sequences, using coordinates from the UCSC genome browser's "Known Genes" and comparative genomics conservation tracks from human genome Build 35. For analyses of the possible effect of selection (Figure 2; Supp. Note 4; Supp. Figure 2), we also removed genomic regions identified as having experienced recent positive natural selection by any of a panel of three different long range haplotype tests[1] (Supp. Note 4).

Fitting models of history to the allele frequency data: We used the procedure of ref. 2 to fit a demographic model of an out-of-Africa bottleneck to the allele frequency spectrum on chromosome X and the autosomes (separately and together). All analyses conditioned on ascertainment in two chromosomes, and used a moving block bootstrap (MBB) to obtain standard errors[3]. The strategy for chromosome X is identical, except effective population size is modeled as ¾ of that for the autosomes (Supp. Note 2).

Translating from sequence diversity to time: Mutation rates on chromosome X are different from those on the autosomes[4,5], and we therefore needed to adjust the raw sequence diversity estimates by a factor that adjusts for this difference to translate to elapsed time. To do this, we divided human diversity t_{HH}, on both chromosome X and the

autosomes, by human-chimpanzee diversity t_{HC}, which we obtained by including a chimpanzee in the alignment procedure. Using human-chimpanzee divergence time as a normalization is not appropriate, however, since it is strikingly different on chromosome X and the autosomes[6]. We therefore further multiplied by the ratio of human-chimpanzee to human-macaque divergence (t_{HC}/t_{HM}, obtained with standard errors from ref. 6 for both chromosome X and the autosomes). Thus, we obtained the quantity $t_{HH}/t_{HM} = (t_{HH}/t_{HC})\times(t_{HC}/t_{HM})$. We carried out this calculation in two steps, instead of directly adding a macaque into the alignment, since macaque is too diverged from human for ssahaSNP[7] to be appropriate for identifying human-human and human-macaque divergent sites with the same settings. Because of polymorphism in the population ancestral to human-macaque divergence, our estimate of t_{HM} on chromosome X is expected to correspond to a slightly smaller time than that for the autosomes. As a result, the chromosome X to autosome ratio of time divergence is slightly overestimated, which is conservative for our analyses.

Expectation of X-to-autosome sequence diversity ratio: Following ref. 8, we used models of human demographic history in which populations are assumed to have been constant in size over epochs. These predict the tMRCA of two chromosomes in a population to be $E(tMRCA) = 2\left\{N_1 + \sum_{m=1}^{M-1}\left[(N_{m+1} - N_m)e^{-\sum_{l=1}^{m}\frac{T_l}{2N_l}}\right]\right\}$. Here, N_i is the effective population size in epoch i and T_i is the duration in generations[8]. The expected tMRCA on chromosome X can then be obtained by multiplying all N_i values by ¾ (to adjust for differences in population size between chromosome X and the autosomes). To predict the X-to-autosome ratio, we used the models fitted to the autosomal SNP data in

Nature Genetics: doi:10.1038/ng.303

ref. 2. For West Africans, this consisted of a single population expansion, and for North Europeans and East Asians, two independent population bottlenecks and a constant population size at other times (Supp. Note 3 explores additional models).

Nature Genetics: doi:10.1038/ng.303

Supplementary Table 1: Bottleneck modeling estimates

	Autosomes		**Chromosome X**		**Test for difference**	
	North European	East Asian	North European	East Asian	North European	East Asian
Inbreeding coefficient *F*	0.151 (0.009)	0.201 (0.012)	0.567 (0.050)	0.579 (0.089)	$P<<10^{-12}$	P=0.0006
Time (kya)	32 (3)	23 (2)	35 (5)	29 (8)	P=0.59	P=0.46

Notes: We considered a bottleneck model of the history of non-African populations with two parameters: the time of a single bottleneck, and its inbreeding coefficient (defined as the number of generations the bottleneck lasted divided by twice the effective population size). The autosomal results are reproduced from the supplementary materials of ref. 2, while the chromosome X results are new. Mean and standard errors are based on 1000 moving block bootstraps (Methods). The table also presents p-values for a deviation of the chromosome X and autosome estimates (two-tailed two-sample z-test). To account for the different effective population size that is expected between chromosome X and the autosomes, this test first multiplies the chromosome X inbreeding coefficient by ¾, and then assess whether the two values are significantly different after this adjustment. While the bottleneck intensities are inferred to be significantly different on chromosome X and the autosomes, the times of the bottleneck are inferred to be consistent. This suggests that the modeling captures a real feature of history, with the only difference being accelerated chromosome X drift during the human dispersal out of Africa.

Supplementary Table 2: Within- and between-population genetic diversity

		African		**Non-African**	
		West African	Biaka Pygmy	East Asian	North European
(a) autosomal π ($\times 10^{-3}$)	West African	1.081 (0.005)	1.190 (0.024)	1.098 (0.004)	1.106 (0.004)
	Biaka Pygmy		n/a	1.186 (0.027)	1.212 (0.025)
	East Asian			0.772 (0.005)	0.892 (0.004)
	North European				0.827 (0.004)
(b) X-chromosome π ($\times 10^{-3}$)	West African	0.722 (0.017)	0.763 (0.025)	0.730 (0.013)	0.727 (0.012)
	Biaka Pygmy		n/a	0.786 (0.024)	0.802 (0.024)
	East Asian			0.414 (0.014)	0.511 (0.013)
	North European				0.460 (0.013)
(c) X:autosome tMRCA ratio	West African	0.763 (0.026)	0.732 (0.033)	0.759 (0.023)	0.751 (0.022)
	Biaka Pygmy		n/a	0.756 (0.034)	0.756 (0.033)
	East Asian			0.613 (0.026)	0.654 (0.023)
	North European				0.635 (0.024)

Notes: (a,b) For each pair of populations, we present autosomal and X-chromosome estimates of π, the fraction of divergent sites per base pair after removing sites that do not meet neighborhood quality score (NQS) thresholds (the diagonal gives within-population estimates). Standard errors in parentheses control for correlation among neighboring sites by jackknifing. (c) The ratio of X-to-autosome genetic diversity after normalizing each by human-macaque divergence to account for the different mutation rates in these two parts of the genome (Methods). Standard errors account for uncertainty in the estimates on chromosome X and the autosomes, as well as uncertainty in the human-macaque normalization.

Nature Genetics: doi:10.1038/ng.303

Supplementary Table 3: SNP data sets for chromosome X

Individual 1	Individual 2	Ancestry	Ascertained SNPs	After all corrections
Cor7340 [9]	Cor7340	North European	12,213	1,415
Cor7340	HuAA [10]	North European [a]	9,517	1,253
Cor11321 [9]	HuFF [10]	East Asian [a]	9,638	1,247
		West African [b]		1,087 [c]

Notes: Comparisons that we used to identify SNPs from chromosome X. The columns specify the two individuals from which the reads used for ascertainment were obtained, their ancestry, the number of SNPs identified, and the number of SNPs remaining after all filters and corrections were applied (Methods and ref. 2).

[a] The ancestry of HuAA and HuFF was determined using ancestry informative markers (ref. 2).

[b] We used four samples to identify SNPs between two West African chromosomes: NA18517, NA18507, NA19240 and NA19129 (Methods).

[c] These SNPs were genotyped in our lab since they are not necessarily in HapMap (Methods).

Nature Genetics: doi:10.1038/ng.303

Supplementary Table 4: DNA samples used in sequence diversity estimates

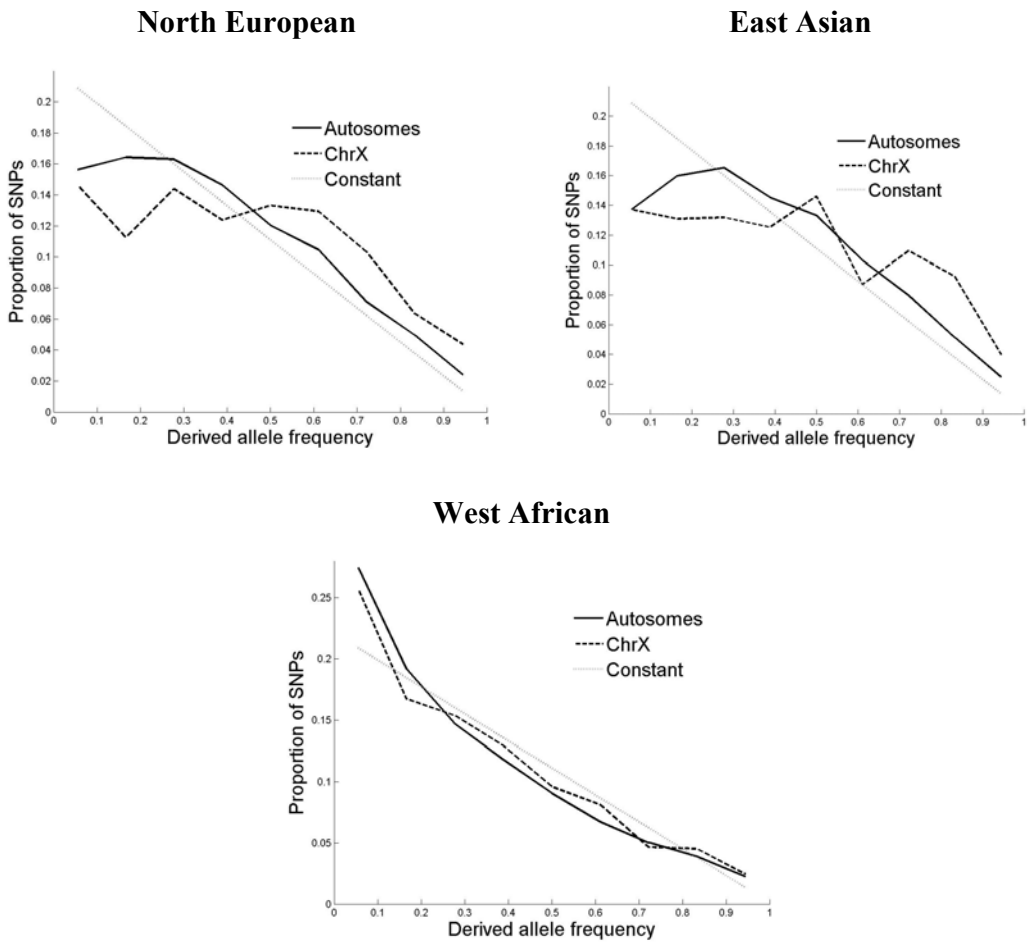
DNA sample	Sequencing center	Population, Coriell or Celera Sample ID	Gender	Traces*	Aligned autosomal bases†	Aligned chromosome X bases†
ABC7	Agencourt	YRI, NA18517	F	1,455,819	455,190,057	25,034,680
ABC8	Agencourt	YRI, NA18507	M	2,503,941	728,536,829	24,355,772
ABC9	Agencourt	JPT, NA18956	F	1,575,595	442,418,960	22,703,404
ABC10	Agencourt	YRI, NA19240	F	1,653,565	451,020,498	23,515,097
ABC11	Agencourt	CHB, NA18555	F	1,506,704	422,948,254	21,294,001
ABC12	Agencourt	CEU, NA12878	F	1,634,889	440,924,721	22,639,007
ABC13	Agencourt	YRI, NA19129	F	1,638,749	481,881,509	23,968,264
ABC14	Agencourt	CEU, NA12156	F	1,730,105	400,938,067	20,565,151
Cor7340	Sanger	CEU, NA07340	F	2,721,720	424,328,147	53,399,153
Cor10470	Sanger	Biaka Pygmy, NA10470	M	358,490	38,429,860	13,190,898
Cor11321	Sanger	East Asian, NA11321	M	1,828,769	550,965,952	62,393,229
Cor17109	Sanger	African American‡, NA17109	M	1,386,277	263,013,170	0
HuAA	Celera	European American, A	M	2,408,092	566,649,249	21,026,744
HuBB	Celera	European American, B	M	5,851,971	985,652,063	37,431,177
HuFF	Celera	East Asian, F	F	1,272,561	356,161,178	19,506,297

* Total number of traces uniquely aligned to the reference genome sequence.

† Total number of aligned bases used in analysis.

‡ Since African Americans are a mixture of African and European ancestry we used ANCESTRYMAP[11] to restrict our analysis to sections of the genome where we were >95% confident of the presence of two African chromosomes (Supp. Note 5).

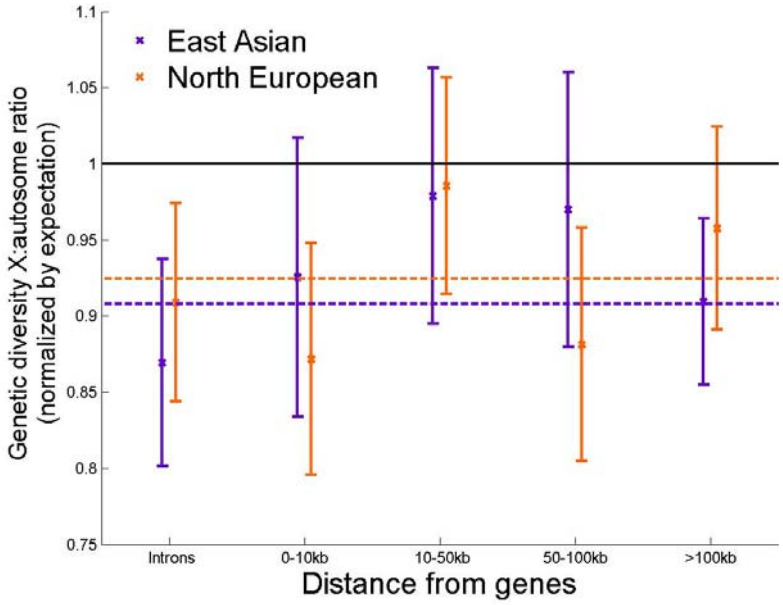
Supplementary Figure 1: Derived allele frequency distributions



Supplementary Figure 1: Derived allele frequency distributions (the proportion of SNPs of each possible derived allele frequency) for each of the HapMap populations, after discovery of SNPs in two reads of the same ancestry. The autosomal North European (CEU) spectrum is based on SNPs ascertained in both the Cor7340 and the HuAA libraries[2]; the X-chromosomal spectrum is based on SNPs ascertained in Cor7340 and between Cor7340 and HuAA (Supp. Table 3). The autosomal East Asian (CHB+JPT) spectrum is based on the Cor11321 and the HuFF libraries[2]; the X-chromosomal spectrum is based on SNPs ascertained between Cor11321 and HuFF (Supp. Table 3). The autosomal West African (YRI) spectrum is based on the Cor17109 library[2]; the X-chromosomal spectrum is based on SNPs ascertained in different Yoruba samples (Supp. Table 3). SNPs ascertained in individuals of the same ancestry are pooled together, as are allele frequency data from the two East Asian populations, CHB and JPT, since they result in very similar spectra. For comparison, the expected derived allele frequency spectrum for a population of constant size throughout history and the same ascertainment scheme is also shown (the spectrum is proportional to $1-x$ as the expected spectrum is proportional to $1/x$ for complete resequencing data, and is then multiplied by $2x(1-x)$, the probability of discovery in two chromosomes). Although all spectra are biased by discovery in two chromosomes, they are comparable visually since the bias is identical for all spectra (we fully account for this bias in analyses[2]). We note that chromosome X frequency spectra are noisier than the autosomal spectra since they are based on fewer SNPs.

Nature Genetics: doi:10.1038/ng.303

Supplementary Figure 2: Sequence diversity ratio binned by distance from genes



Supplementary Figure 2: No attenuation of the reduction in non-African sequence diversity with distance from genes. We divided both our chromosome X and autosomal data sets based on distance from the nearest gene, and found no attenuation of our signals with increasing distance, as would be expected if selection explained the results. The figure plots the ratio of X-to-autosome genetic diversity in non-Africans, normalized by the same quantity in West Africans, and normalized by the expectation from the best-fit models of history[2] (Supp. Note 3). The values are all below 1 (horizontal black line), reflecting the reduction in the X-to-autosome ratio outside of Africa, below demographic expectation. Dotted lines show the observed values for all bins together, and there is no evidence of a deviation of the individual bins from this average (error bars indicate ±1 standard error).

Nature Genetics: doi:10.1038/ng.303

Supplementary Note 1: Detailed allele frequency differentiation analysis

We estimated allele frequency differentiation between two populations using F_{ST} as defined in Supp. Note 10 of ref. 2. Under assumptions that are essentially satisfied for the SNP ascertainment and populations we are considering, the expected value is $F_{ST} = (1 - e^{-(\tau_1 + \tau_2)})/2$, where τ_1 and τ_2 are the scaled drift times of population 1 and population 2. The assumptions are:

(1) The SNPs are ascertained in population 1 using any ascertainment scheme.
(2) Population 1 has been panmictic and constant in size since the split from population 2.
(3) Population 2 has been panmictic (but not necessarily constant in size) since the split from population 1.
(4) No significant gene flow occurred between the two populations since their split.

Importantly, under these assumptions, our statistic is independent of population size changes in population 2, and only depends on the total amount of drift in that population, τ_2 (ref. 2). We note that while our F_{ST} definition is slightly different than the definitions most commonly used, for example that of Weir and Cockerham[12], it yields essentially identical estimates for the levels of differentiation between the human populations we studied (data not shown), and has the advantage that we can directly translate the observed values of F_{ST} to population genetic estimates of genetic drift.

Since $F_{ST} = (1 - e^{-(\tau_1 + \tau_2)})/2$ under our ascertainment scheme, we can compare the genetic drift on the autosomes and chromosome X using the equation $Q = \ln(1 - 2F_{ST}^{auto}) / \ln(1 - 2F_{ST}^{X})$. This ratio is expected to equal 3/4 if X chromosome effective population size has been ¾ that of the autosomes since the two populations split, and so allows a strict test of this assumption.

Due to the assumption of constant population size in population 1 (the ascertainment population) since the split from population 2, we only measured F_{ST} in scenarios where this assumption is reasonable. We further verified the reliability of our procedure using coalescent computer simulations[13], as detailed below.

To measure F_{ST} between West Africans and non-Africans, we used SNPs ascertained in two West African chromosomes (Supp. Note 2). The assumption of effectively constant population size is actually a reasonable assumption for West African history, since although the population expansions in the last tens of thousands of years have been quantitatively large, they have not had enough time to substantially affect the frequencies of more common variants[2]. To check that this procedure produces useful estimates of F_{ST}, we carried out coalescent simulations. We simulated the West African / non-African split to have occurred 60kya, a bottleneck on the ancestry of East Asians and North Europeans (estimates for autosomes in Supp. Table 1), and SNP ascertainment in two West African

Nature Genetics: doi:10.1038/ng.303

chromosomes. Using these parameters, we obtained autosome-to-X genetic drift ratios of Q=0.756 for the drift between West Africans and North Europeans and 0.746 for West Africans and East Asians, close to our theoretical expectation of ¾. To test the effect of gene flow on these expectations, we repeated the same simulations while allowing for 10% gene flow between the two populations after they split. These simulations resulted in an autosome-to-X genetic drift ratios of Q=0.749 for the drift between West Africans and North Europeans and 0.747 for West Africans and East Asians.

To measure F_{ST} between North Europeans and East Asians, we separately ascertained SNPs in each of the populations (Supp. Note 2). This analysis makes the simplifying assumption of a constant population size since the North European / East Asian split, but is not affected by prior changes in population size such as the out-of-Africa bottleneck. Using the coalescent simulations of the joint demographic history of North Europeans and East Asians from ref. 2, we obtained an autosome-to-X drift ratio of Q=0.745 (North European ascertainment) and 0.744 (East Asian ascertainment), again close to the theoretical expectation of ¾. When allowing for 10% gene flow between North Europeans and East Asians, Q=0.742 for both North European ascertainment and East Asian ascertainment.

While theory and coalescent simulations predict the autosome-to-chromosome X ratio of genetic drift to be ¾, and while East Asian-North European drift is consistent with this expectation (Table 1), the ratio since the West African-North European split and since the West African-East Asian split is observed in our data to be highly significantly lower than ¾ (Table 1). These results suggest that chromosome X experienced increased genetic drift compared with the autosomes after the split between Africans and non-Africans, but not after the East Asian-North European split.

Considering the frequency differentiation results by themselves, the additional genetic drift on chromosome X could have occurred on either the African or non-African lineages. However, the latter possibility is the only one consistent with the results based on frequency spectra (Supp. Note 2) and sequence diversity (Supp. Note 3). We note that even though F_{ST} between West Africans and East Asians is larger than F_{ST} between West Africans and North Europeans in both parts of the genome (Table 1)—which we have previously shown to be due to the increased drift in the history of East Asians compared with North Europeans[2]—the ratio between the two compartments of the genome is consistent across Europeans and Asians, 0.582±0.030 and 0.615±0.030, further supporting the inference that the increased X-chromosomal drift occurred before these two populations split.

Last, we also examined genetic drift among East Asians, estimating F_{ST} between Han Chinese (CHB) and Japanese (JPT) to be 0.0065±0.0003 on the autosomes and 0.0090±0.0016 on chromosome X, based on SNPs ascertained in Chinese. This resulted in an autosomal-to-chromosome X ratio of genetic drift of 0.720±0.133, consistent with the expectation of ¾ (P=0.82). This further highlights the fact that the deviation of the genetic drift ratio from expectation, comparing African and non-African populations, indicates a surprising and unusual feature of the out-of-Africa dispersal.

Nature Genetics: doi:10.1038/ng.303

F_{ST} and Q for different ascertainment schemes

East Asian – North European

Ascertainment population	Autosomal F_{ST}	Chromosome X F_{ST}	autosome-to-X genetic drift ratio *Q*	P-value versus expected ¾
East Asian	**0.098 (.002)**	**0.131 (.009)**	**0.715 (.050)**	**0.48**
North European	**0.106 (.002)**	**0.133 (.006)**	**.771 (.036)**	**0.57**
West African	0.107 (.003)	0.138 (.009)	0.753 (.054)	0.96

East Asian – West African

Ascertainment population	Autosomal F_{ST}	Chromosome X F_{ST}	autosome-to-X genetic drift ratio *Q*	P-value versus expected ¾
East Asian	0.158 (.002)	0.220 (.011)	0.653 (.036)	7.4×10^{-3}
North European	0.181 (.002)	0.276 (.012)	0.561 (.029)	4.7×10^{-11}
West African	**0.178 (.003)**	**0.256 (.010)**	**0.615 (.030)**	6.5×10^{-6}

North European – West African

Ascertainment population	Autosomal F_{ST}	Chromosome X F_{ST}	autosome-to-X genetic drift ratio *Q*	P-value versus expected ¾
East Asian	0.151 (.002)	0.226 (.013)	0.599 (.039)	1.1×10^{-4}
North European	0.141 (.002)	0.213 (.008)	0.598 (.027)	1.8×10^{-8}
West African	**0.144 (.003)**	**0.221 (.009)**	**0.582 (.030)**	3.0×10^{-8}

Notes: Similar to Table 1 in main text, but for each pair of populations, F_{ST} and *Q* are provided for each possible ascertainment population. The rows in bold are presented in Table 1 and used for analysis throughout since they are the only ones consistent with the theoretical assumptions for genetic drift analysis.

Nature Genetics: doi:10.1038/ng.303

Supplementary Note 2: Detailed derived allele frequency distribution analysis

The allele frequency spectrum holds important information about a population's demographic history, with differences between chromosome X and the autosomes indicative of different histories between these two parts of the genome.

We studied large data sets of tens of thousands of autosomal SNPs and thousands of chromosome X SNPs that were all identified in the same way: by comparison of two chromosomes of the same ancestry, followed by genotyping in a large number of samples from all the International Haplotype Map (HapMap) populations. The autosomal data sets are based on the data sets previously reported in ref. 2, with a few modifications (Methods). Here we repeated the same ascertainment procedure to obtain analogous data sets that allow estimation of the allele frequency spectrum of chromosome X in each population (Supp. Table 3). To reduce the effect of natural selection on our results, we excluded coding SNPs and SNPs in conserved non-coding regions (from UCSC hg17 genome browser's known genes and conservation tracks) from both the previously reported autosomal data sets, and our newly reported chromosome X data sets. The updated, combined data set is available at http://genepath.med.harvard.edu/~reich.

Supp. Figure 1 contrasts the chromosome X and autosome allele frequency distributions obtained by identifying SNPs in two chromosomes of known ancestry and studying their allele frequencies in samples of the same ancestry. Our analyses focus on derived alleles identified as the allele that is new relative to both chimpanzee and orangutan[2]. Both the chromosome X and autosomal West African allele frequency distributions exhibit more rare derived alleles than expected for a constant-size population, consistent with a population expansion[2,8,14-17], and there is a deficiency of rare alleles in chromosome X and autosomal spectra of both North Europeans and East Asians, compared with expectation, consistent with population contraction(s) dispersing out of Africa[2,8,14,17,18].

Importantly, the frequency spectrum of chromosome X differs from the frequency spectrum of the autosomes, particularly for the non-African populations (Supp. Figure 1). For both North Europeans and East Asians, chromosome X shows a higher proportion of SNPs of high derived allele frequency compared with the autosomes, with significant difference in the mean derived allele frequency ($P<<10^{-12}$ for North Europeans and $P=2x10^{-11}$ for East Asians; two-tailed two-sample t-test; for comparison, P=0.02 for West Africans). However, the two allele frequency distributions are not directly comparable since chromosome X SNPs are affected more by more recent demographic history due to the expected ¾ population size difference between chromosome X and the autosomes[19,20].

To examine whether known features of human demographic history can account for the differences between autosomal and chromosome X spectra, we examined the fit of our demographic models from ref. 2 to both allele frequency distributions (we use the same models of history as in Supp. Note 3 to follow, correcting for discovery of SNPs in two chromosomes from the population).

Nature Genetics: doi:10.1038/ng.303

The allele frequency distributions predicted by the models provide an excellent fit to the autosomal allele frequency distributions of all three populations (Figure 1a). If chromosome X experienced the same demographic history, these models should allow prediction of the chromosome X allele frequency distributions, after accounting for a difference in the effective population size by a factor of ¾. While the West African chromosome X allele frequency distribution matches this prediction reasonably well, both non-African chromosome X allele frequency distributions exhibit a deficiency of rare derived alleles compared with prediction (Figure 1b), with the predicted allele frequency distributions fitting the data much worse than is the case for the autosomes. These results suggest that non-African deviation of chromosome X SNP allele frequencies from their autosomal counterparts cannot be explained by the simple demographic events (bottlenecks) described in these models. Rather, the deviation from prediction suggests more drift on chromosome X than the autosomes in both non-African populations, consistent with chromosome X having experienced a more severe out-of-Africa population bottleneck.

To rigorously explore the hypothesis of differences in non-African demographic history between chromosome X and the autosomes—perhaps due to differences in how the out-of-Africa population bottleneck affected these two parts of the genome—we fit the same demographic model to both autosomal and chromosome X SNP allele frequencies. Due to the relatively small number of SNPs in the chromosome X data sets (Supp. Table 3), we could not make fine distinctions between models as was possible for the autosomal data set[2]. For example, we could not distinguish between one and two bottlenecks.

To approach the chromosome X data, we therefore took the strategy of identifying minimally complex models that could approximate important features in the data. We modeled a single population bottleneck independently in the history of both North Europeans and East Asians. These 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 as $F=T/2N$, the number of generations the bottleneck lasted divided by twice the effective population size, 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. The effect on the frequency spectrum depends primarily on this ratio and is practically independent of the predefined value of T, since a simultaneous scaling of both T and N does not change the results.[2]

A bottleneck model fits better than a model of constant population size, with likelihood ratio tests (LRTs) favoring a bottleneck at a significance of $P << 10^{-12}$ for both parts of the genome. The pattern is observed for both North Europeans and East Asians, with distinct peaks in the likelihood function (figure below). To account for the effect of correlation between SNPs, we also estimated standard errors of the maximum likelihood estimates by bootstrapping data sets using the Moving Block Bootstrap (MBB) approach[3,21,22], randomly resampling contiguous runs of SNPs from the data. Based on the MBB results (Supp. Table 1), the bottleneck model is significantly more likely than

Nature Genetics: doi:10.1038/ng.303

one of a constant-sized population ($P=8\times10^{-6}$ for East Asian chromosome X, and $P<<10^{-12}$ for the remaining three tests, one-tailed z-test). Importantly, the chromosome X model is not consistent with the autosome model after adjusting for the expected ¾ difference in population size ($P<<10^{-12}$ for North Europeans and $P=2\times10^{-6}$ for East Asians, LRT), with the chromosome X data suggesting a much higher underlying bottleneck intensity (Supp. Table 1). These differences in modeling between the two parts of the genome are consistent with increased chromosome X drift associated with the out-of-Africa event. We note that while the modeling suggests differences in bottleneck intensity, interestingly the bottleneck models applied to chromosome X and the autosomes separately estimate the same bottleneck time in both parts of the genome (Supp. Table 1): 35 to 23 thousand years ago (kya). While this time estimate is not realistic—reflecting the fact that a single-bottleneck model is oversimplified[2]—it is somewhat encouraging that the estimates from the two parts of the genome are consistent. The coincidence of the time estimates suggests that the two parts of the genome may really only be different with regard to the amount of genetic drift they experienced during the bottleneck(s).

The estimated bottleneck dates are much more recent than the 80-40kya date usually ascribed to the out-of-Africa expansion based on the archaeological record[23-25]. This reflects either the fact that a simple bottleneck model does not capture the complexity of the human dispersal out of Africa (which we think certainly explains part of the discrepancy; for example in a previous study we found a much better fit for a two bottleneck model[2]), or that scaling of the time estimates by human-chimpanzee genetic divergence is not accurate, although it seems unlikely that the time calibration to the fossil record will be inaccurate by more than 30%.[2] We emphasize, however, that the estimates of bottleneck intensity are independent of the time estimates or uncertainties due to calibration from the fossil record since they are not affected by the normalization by the population's sequence diversity, which also makes this result of a more severe bottleneck on chromosome X independent of the results based on sequence diversity (Supp. Note 3).

The results presented in this note provide an independent line of evidence showing acceleration of genetic drift on chromosome X associated with the human dispersal out of Africa of those based on allele frequency differentiation described in Supp. Note 1 since those were based on the data sets of SNPs ascertained in West Africans and compared between West Africans and non-Africans, while the results here are based on the data sets of SNPs ascertained in North Europeans and East Asians and are only studied in non-Africans.

We also analyzed SNP allele frequency data in West Africans by studying whether a 2-epoch expansion model that we had previously fitted to the autosomes also produced a better fit for the X chromosome data[2]. This model allows for one population size change in the history of a population, with two parameters capturing the time and magnitude of change. This model fits the chromosome X data significantly better than a model of constant population size (P=0.0046, LRT), and is consistent with estimates for the autosomes after adjusting for the expected ¾ difference in population size (P=0.34, LRT).

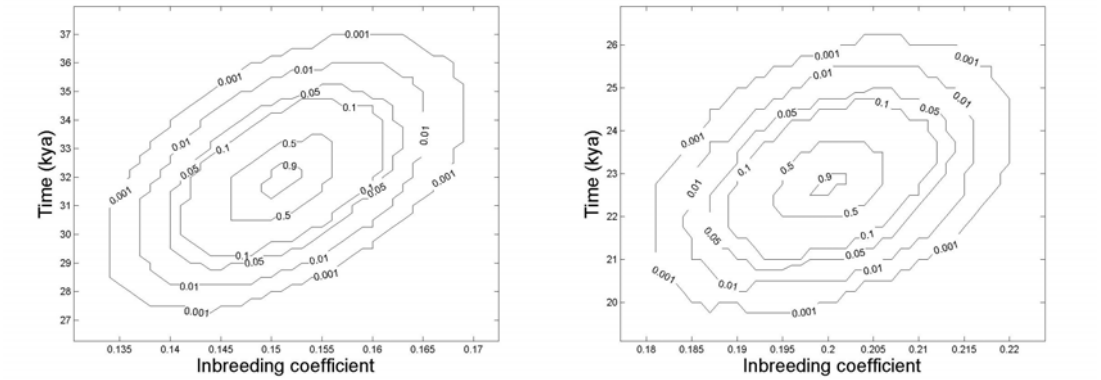
Nature Genetics: doi:10.1038/ng.303

This lends further support to the result that the acceleration of chromosome X genetic drift is unique to non-African history.

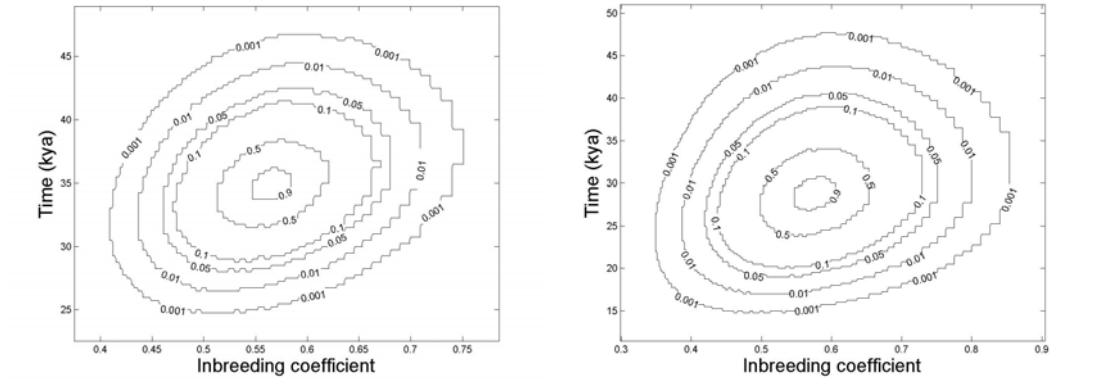
Bottleneck modeling likelihood surfaces

North European **East Asian**

a (Autosomes)



b (Chromosome X)



Contour of the likelihood surface for the bottleneck model as a function of the two model parameters, inbreeding coefficient *F* and time, for the autosomes (a) and chromosome X (b). Contour values (0.9, 0.5, 0.1, 0.05, 0.01, and 0.001) are P-values of testing the likelihood at each parameter values combination versus the maximum likelihood estimate (MLE; χ^2 test with 2 df). For example, the region between contours 0.05 and 0.01 includes all values for which the fit to the data is worse than for the MLE with significance $0.01<P<0.05$. The scales of both axes differ between the figures to allow focusing on the MLE (all values not presented satisfy $P<0.001$). The observed MLE slightly differ from the estimates in Supp. Table 1 since those estimates are based on bootstrapping, while this figure profiles the likelihood function from the full data set. The qualitative inferences, however, are identical.

Nature Genetics: doi:10.1038/ng.303

Supplementary Note 3: Detailed sequence diversity analysis

To estimate the average time since the most recent common ancestor (tMRCA) comparing DNA sequences within and between human populations, we aligned over a billion autosomal and chromosome X base pairs from individuals of known ancestry. We focused on individuals for whom shotgun genome sequence was available, using 7 African samples (6 West African-origin samples including 1 African American, and 1 Biaka Pygmy sample) and 9 non-African samples (5 North Europeans and 4 East Asians) (Supp. Table 4). We estimated π, which we refer to as sequence diversity, as the fraction of differences per base pair between chromosomes, both within- and between-populations. To minimize the effect of selection on our analyses, we excluded coding regions of genes and conserved non-coding sequences from this analysis (using information from UCSC hg17 genome browser's known genes and conservation tracks), which removed about 10% of our aligned sequence.

Throughout the genome we observe higher genetic diversity in African than in non-African populations (Supp. Table 2a-b), consistent with the out-of-Africa bottleneck that is well-known to have reduced diversity in non-African populations[26].

The ratio of chromosome X-to-autosome genetic diversity, following the normalization of each by human-macaque divergence in the same compartment of the genome (Methods), is expected to correspond to the ratio of effective population size of the two compartments, as this is the ratio of the tMRCA averaged over these two parts of genome. Assuming an equal effective population size of males and females, the ratio is expected to equal ¾. The ratio comparing between populations is expected to be slightly higher than ¾. As the time since the most recent common ancestor of two alleles from different populations is at least as old as the populations split, an equal amount of time is added to both the numerator and denominator, increasing the ratio.

The chromosome X-to-autosome ratio of both within-population and between-population genetic diversity in our data is very close to the ¾ expectation when estimated between African and non-African or between two African populations. The between-population estimates of genetic diversity ratio range between 0.732±0.033 and 0.759±0.023, while the within-population West African ratio is 0.763±0.026 (Supp. Table 2c), all consistent with 3/4. However, when both populations are non-African the ratio is much lower: North European diversity ratio is 0.635±0.024, East Asian ratio is 0.613±0.026, and the ratio of diversity between these two populations is 0.654±0.023 (Supp. Table 2c).

What could explain the reduced chromosome X-to-autosome ratio of genetic diversity in non-Africans? We first considered whether known features of non-African demographic history could explain the data. The ratio would be expected to deviate from ¾ for a population which demographic history deviates from a constant effective population size[27]. The fundamental reason for this is that chromosome X loci are expected to coalesce with higher probability in any generation than autosomal loci due to the smaller effective population size and hence, they are affected more by more recent demography. For a population that experienced an expansion, the ratio is expected to be greater than ¾

Nature Genetics: doi:10.1038/ng.303

since chromosome X more intensely "experiences" the recent, larger effective population size; for example, for a 10-fold expansion that occurred 1,000 generations ago, the ratio is expected to be 0.7607, and the ratio is expected to be larger the more extreme the expansion is (for a star-shaped genealogy, the ratio approaches 1). Similarly, the ratio is expected to be smaller than ¾ for a population contraction[27]. It is known that non-African human populations experienced a population bottleneck in their history—a temporary crash in population size, followed by expansion. A bottleneck in the history of a population has a mixed effect on chromosome X-to-autosomes diversity ratio, which is dominated by the effect of the population contraction, e.g. for a bottleneck with an inbreeding coefficient of F=0.25, that took place 1,000 generations ago, the ratio is expected to be 0.7044 (Methods).

To account for the effect of demographic history on the ratio of genetic diversity, we used the subsets of autosomal SNPs from HapMap[9,28] that allow us to overcome ascertainment biases and to provide accurate measurements of the allele frequency spectrum of West Africans, North Europeans, and East Asians[2]. We previously reported an analysis of these data on the autosomes, which identified demographic histories for these three populations that resulted in accurate fits to the observed allele frequency spectra (Figure 1a). These models also allow us to obtain an expectation for the chromosome X-to-autosomes diversity ratio under the assumption of an equal male and female effective population size. The best-fit model of West African history assumes a single ancient population expansion; the best-fit model of North European and East Asian history assumes two population bottlenecks, one associated with the out-of-Africa event, and another, more recent, possibly associated with the Last Glacial Maximum[2]. Since all models are idealizations, we verified that the predicted chromosome X-to-autosomes diversity ratio is not sensitive to the exact modeling assumed, specifically by showing that extremely similar predictions of the X-to-autosome diversity ratio are yielded by a model of a single bottleneck in the history of North Europeans and East Asians.

Table 2 compares the observed normalized chromosome X-to-autosomal genetic diversity ratio with the ratio predicted by the demographic history model of each population. While demography accounts for part of the lower non-African ratio, as captured by the fact that the model predicts a value below ¾, both non-African ratios are still significantly below prediction (P=0.005 and P=0.003 for North Europeans and East Asians; P=0.0004 and P=0.0002 when instead using a single bottleneck model for prediction). By contrast, the West African ratio closely matches prediction (P=0.514; Table 2). Combining evidence from all three populations, the significance of deviation from prediction is P=0.0007 (χ^2 test with 3 degrees of freedom; df) and it is P=0.0002 when combining evidence from only the two non-African populations (χ^2 test with 2 df). These results show that chromosome X is less diverse than the autosomes in non-African populations, beyond what is expected from demographic history as estimated by autosomal SNPs. Interestingly, the deviation of the tMRCA ratio from expectation is of approximately the same magnitude in the two non-African populations. The observation is 10.5% ± 4.2% less than expectation for the North Europeans and 12.6% ± 4.8% less than expectation for the East Asians (Table 2)—suggesting that the reduction in diversity

Nature Genetics: doi:10.1038/ng.303

is most likely due to a shared event in the two populations' histories, which was also supported by the allele frequency differentiation analysis (Supp. Note 1).

To more formally test whether the pattern of reduced chromosome X diversity in East Asians and North Europeans could be entirely explained by a shared event in their history, we assumed that the entire result can be explained by a different chromosome X to autosomes effective population size ratio during the out-of-Africa bottleneck. Specifically, we assumed that the X-to-autosome genetic drift ratio was ¾ both before and after this event, but deviated during the bottleneck. We then identified the ratio of autosome-to-X chromosome genetic drift during the bottleneck that matches the diversity ratio, carrying out the analysis separately for North Europeans and East Asians. The resulting X-to-autosome effective population size ratio is 0.525±0.071 for North Europeans and 0.509±0.064 for East Asians during the bottleneck. Based on our analysis with this simplified model, the amount of chromosome X increased drift during the out-of-Africa bottleneck is consistent between North Europeans and East Asians (P=0.87, two-tailed two sample z-test), with both deviating from the expected ratio of ¾ (P=0.001 and P=0.0002, two-tailed z-test).

These results suggest accelerated genetic drift of chromosome X in non-African human populations: after the split from Africans, but before the split of North Europeans and East Asians. This result is further supported by estimates of genetic diversity between pairs of populations that are significantly lower than ¾ only for comparisons involving two non-African populations (Supp. Table 2). We emphasize that the results presented here are independent of the results in Supp. Note 1 and Supp. Note 2 since they are based on independent data sets (sequence diversity rather than SNP allele frequencies).

Finally, we caution that the above analysis depends on the normalization of genetic diversity in both chromosome X and the autosomes by human-macaque divergence in these same parts of the genome. Although the estimates account for uncertainty in this normalization, the normalization could in principle be systematically biased due to gender-specific changes in generation time since the split of human and macaque. (The fact that chromosome X coalesces faster within the ancestral population of human and macaque, which is not accounted for by our normalization, is conservative for all our analyses: any differences of this type will raise the observed X-to-autosome ratio.) To account for possible errors in the macaque normalization, we also carried out analyses in which we directly compared X-to-autosome diversity ratios between Africans and non-Africans, without using human-macaque divergence so that uncertainties in this quantity do not affect our estimates. Specifically, we compared the ratio of non-African to West African X-to-autosome diversity ratio with the expectation based on the demographic history of both. Testing for a deviation between ratios of two random variables increases the standard errors, and so these results are less significant than the single-population results that do require normalization by macaque (Table 2). Nevertheless, we continue to observe a significant reduction in the X-to-autosome ratio outside of Africa compared with demographic expectation: P=0.028 for North Europeans and P=0.016 for East Asians using a two-tailed two sample z-test.

Nature Genetics: doi:10.1038/ng.303

Supplementary Note 4: Can natural selection account for the results?

One possible explanation for our results is natural selection differently affecting chromosome X and the autosomes. Although chromosome X is known to experience more selection than the autosomes because of the exposure of recessive alleles in hemizygous males, the only selection scenario that might be consistent with all results is that of increased X chromosome selection pressure in non-Africans, after the split from West Africans, but before the North European-East Asian split, which simultaneously affected many different loci on chromosome X. Such a change in selection pressure could have been associated with environmental change or an increased competition for resources during the dispersal out of Africa. (We note that the scenario of widespread selection focused on chromosome X associated with migration to a new environment seems less likely in light of the fact that when we analogously compare North Europeans to East Asians, or Japanese to Chinese, we do not find similar evidence.)

Natural selection on newly arising genetic variants

We began by considering two theoretical models of natural selection, both of which result in a reduction in genetic diversity on chromosome X: negative purifying selection, and positive selective sweeps.

Negative selection on newly arising mutations is known to reduce genetic diversity in the regions centered on these variants. However, this phenomenon is expected to produce the opposite pattern to what we observe in our data. Because recessive X-linked mutations are more efficiently purged in hemizygous males, negative selection on newly arising genetic variants is expected to produce a relatively greater reduction of linked neutral variation on the autosomes than on chromosome X[29,30], whereas the opposite is observed in our data of non-African populations (Table 2).

Reduction of sequence diversity due to positive selection on newly arising genetic variants is expected to have a greater effect on chromosome X than on the autosomes, because the hitchhiking effect associated with positive selection works more efficiently in hemizygous males when the positively selected allele is recessive.[31,32] However, a sweep also strongly affects the allele frequency spectrum, and so we can test whether the allele frequency spectrum on chromosome X in non-African populations is what is expected in the case of positive selection on newly arising variants. We note that to explain the chromosome X-wide effects we observe, there would have to have been not one, but multiple sweeps across chromosome X during the out-of-Africa dispersal, and these would have had to produce distributions of allele frequencies typical of positive selection.

The observed allele frequency spectra provide no support for positive selective sweeps explaining our data. A prediction of positive selective sweeps is that there will be an increase in the rate of alleles of low derived frequency for complete sweeps and an increase in the rate of alleles of high derived frequency for incomplete sweeps[33,34]. Since intense selective sweeps during the out-of-Africa dispersal are likely to have been

Nature Genetics: doi:10.1038/ng.303

completed, an increase of alleles of low derived frequency would be expected. However, it is not observed in our data (Supp. Figure 1). Moreover, while the average derived allele frequency is increased, consistent with incomplete sweeps, the effect is not confined to the very high end of the allele frequency spectrum as would be expected from this model of selection (Supp. Figure 1). We conclude that there is no evidence that positive selective sweeps on newly arising variants are explaining the gross patterns we observe on chromosome X.

Natural selection on pre-existing variants ("standing variation")

An alternative way that selection could explain the accelerated genetic drift on chromosome X is if the selection "turned on" after the dispersal of humans out of Africa – that is, many alleles that were previously nearly neutral or in mutation-selection equilibrium suddenly had an important phenotypic effect and were then subject to negative or positive selection in the non-African environment. In this scenario, the allele frequency distribution would not be expected to be substantially affected as the selected alleles would exist on diverse haplotypes. However, we caution that to explain our data this scenario would have to be rather extreme, with a very large number of loci affected, and most individuals subject to intense selection.

Filtering out regions known to be affected by selective sweeps does not diminish the evidence of accelerated genetic drift on chromosome X

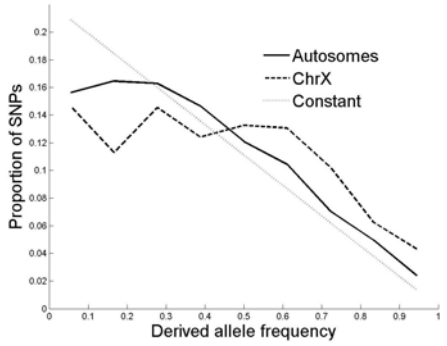
To control for the effect of selection on all our analyses, we not only filtered out coding sequences and regions of high conservation across species as for the other analyses, but also filtered both the sequence diversity and SNP data sets by a recent genome-wide survey of selective sweeps that used several different haplotype-based tests[1]. While a larger fraction of chromosome X than the autosomes was filtered as being under recent positive selection (6.9% for chromosome X and 2.4% for the autosomes), all our results hold even when the recent positive selection filter is applied, suggesting greater chromosome X genetic drift above and beyond recent selection. The following table repeats the results from Table 2 after the application of this filter. The observation of a lower ratio of chromosome X-to-autosomes genetic diversity in non-Africans continues to hold: P=0.029 and P=0.016 for North European and East Asian diversity, respectively, with a combined significance of P=0.004 for both populations (χ^2 test with 2 df).

Nature Genetics: doi:10.1038/ng.303

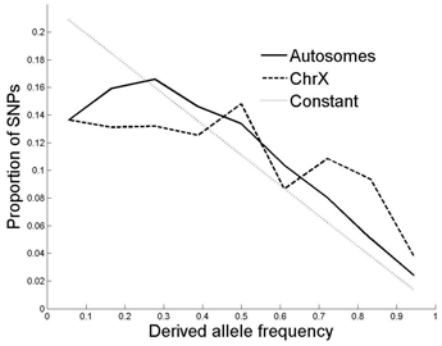
	Autosomes	**Chrom. X**	**Comparison of autosomes and chromosome X**	
	Divergent sites/base pair ($\times 10^{-3}$)	Divergent sites/base pair ($\times 10^{-3}$)	Observed X-to-autosome ratio normalized by macaque divergence	P-value for difference between observed and expected
West African	1.081 (0.005)	0.723 (0.016)	0.767 (0.026)	0.617
North European	0.827 (0.004)	0.470 (0.013)	0.649 (0.024)	0.029
East Asian	0.772 (0.005)	0.423 (0.014)	0.626 (0.026)	0.016

The allele frequency spectra also did not change substantially following the filtering of regions of putative selective sweeps (the figure below repeats Supp. Figure 1 after the application of this filter). The bottleneck modeling results (Supp. Note 2) also hold, with the chromosome X model inconsistent with the autosomal model ($P << 10^{-12}$ for North Europeans and $P = 3 \times 10^{-6}$ for East Asians, LRT). Finally, to be conservative, the reported results based on allele frequency differentiation analysis (Supp. Note 1; Table 1) were already after the filtering of putatively swept regions.

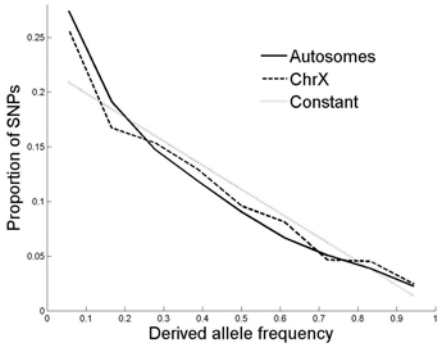
North European



East Asian



West African



Nature Genetics: doi:10.1038/ng.303

The signal of accelerated genetic drift on chromosome X is not correlated with the distance from genes

We tested whether the evidence for increased chromosome X genetic drift is attenuated with distance from genes by subdividing the data into bins based on distance from genes (introns, <10kb, 10-50kb, 50-100kb and >100kb). Our results are consistent across the different bins, showing no trend toward a weakened signal with longer distance from genes (Figure 2a; Supp. Figure 2). These results indicate that natural selection that unusually affected a large proportion of chromosome X genes is unlikely to explain our results. We caution, however, that if the selection was strong (selection coefficient >>0.001), regions >100kb would in effect still be effectively close to genes (selection would affect regions on the scale >>100kb). Thus, these results cannot rule out fast, intense selection on standing variation that affected many loci simultaneously.

The signal of accelerated genetic drift is widespread across chromosome X

We also estimated the genetic drift on chromosome X in 3 centimorgan windows, to study whether there was heterogeneity in genetic drift across the chromosome. The results demonstrate that the signal is widespread across chromosome X (Figure 2b,c), which is not what would be expected if the results we observe are due to natural selection on a few regions.

The signal of accelerated genetic drift is specific to the comparison of African and non-African populations

We analyzed the autosome-to-X genetic drift ratio Q in two other human population comparisons that like the dispersal out of Africa, also involved migration to new environments. If selection unusually affecting chromosome X is a characteristic of human dispersals into new environments or population founding events, it would be expected to similarly act in the context of the North European-East Asian split, and the Chinese-Japanese split. However, the ratio Q between both these pairs of populations is consistent with the expected ratio of ¾ (Table 1; Supp. Note 1).

The lack of a signal associated with the North European-East Asian population split (as well as in the Chinese-Japanese split) suggests that the patterns we observe are a unique feature of the history of the out-of-Africa period, and that if selection explains our results, there must be something biologically different about the dispersal to the non-African environment that did not occur to the same extent in these environments.

Nature Genetics: doi:10.1038/ng.303

Supplementary Note 5: African American SNP discovery

The DNA sample we used to ascertain autosomal SNPs between two West African chromosomes would ideally have had 100% African ancestry. However, since in practice we analyzed data from an African American sample (Cor17109), we had to deal with the complication that the individual had some European ancestry. While most SNPs discovered between the two chromosomes of this individual are expected to be between two chromosomes of African origin, some will be discovered between one African chromosome and one European origin chromosome, or even between two European origin chromosomes. The relative rates of each scenario will depend on the European ancestry proportion of this individual. For chromosome X, we ascertained SNPs in Yoruba samples and genotyped them in our own laboratory, so we were not challenged by this problem.

We used the software ANCESTRYMAP[11] to obtain probabilities in Cor17109 for the number of chromosomes of European ancestry at each point of the autosomes, at one centimorgan resolution. Based on the output of ANCESTRYMAP, we restricted SNP discovery to regions in which this individual is determined to have no European ancestry with probability >.95. Based on ANCESTRYMAP, we also estimated that Cor17109 has an overall European ancestry of only about 4%.

To validate this procedure, we repeated the main analysis that is based on SNPs ascertained in West African chromosomes—estimation of F_{ST} between West Africans and non-African—without using ANCESTRYMAP to exclude regions with potential European ancestry, but instead treating our data as a mixture of different ascertainments.

We defined the following quantities:

F_{ST}^{EE} For SNPs ascertained between two European chromosomes, which we could measure directly

F_{ST}^{EA} For SNPs ascertained between a European chromosome and Cor17109, which we could measure directly

F_{ST}^{AA} For SNPs ascertained between the two chromosomes of Cor17109, which we could measure directly

We then algebraically translated these quantities into:

F_{ST}^{YY} For SNPs ascertained between two African chromosomes, which we could not measure directly without using ANCESTRYMAP

F_{ST}^{EY} For SNPs ascertained between one African and one European chromosome, which we could not measure directly without using ANCESTRYMAP

Using the equations

$$F_{ST}^{EA} = (0.04)(F_{ST}^{EE}) + (1\text{-}0.04)(F_{ST}^{EY})$$

$$F_{ST}^{AA} = (0.04)^2(F_{ST}^{EE}) + 2(0.04)(1\text{-}0.04)(F_{ST}^{EY}) + (1\text{-}0.04)^2(F_{ST}^{YY})$$

We find that the algebraic estimate of F_{ST}^{YY} provides an accurate match to the ANCESTRYMAP analysis (table below). Further evidence for robustness came from

Nature Genetics: doi:10.1038/ng.303

also applying the ANCESTRYMAP analysis to Cor17119, a second African American sample with a larger proportion of European ancestry (20%), but whose results agree well with those from Cor17109.

	Algebraic calculation of F_{ST}^{YY} for Cor17109	Cor17109 based on ANCESTRYMAP (reproduced from Table 1)	Cor17119 based on ANCESTRYMAP
West African – East Asian	0.176 (0.004)	0.178 (0.003)	0.182 (0.003)
West African – North European	0.142 (0.003)	0.144 (0.003)	0.146 (0.002)

We note that while the ANCESTRYMAP filter eliminates most sites in the genome without two African chromosomes, it is inevitable that in a small proportion of sites the filter will fail and SNPs will be ascertained between one African and one European chromosome. This will result in an overestimation of autosomal F_{ST} since SNPs ascertained as divergent sites between an African and a European tend to be more differentiated between Africans and non-Africans. This overestimation, if it occurs, is conservative with regards to our result of reduced autosome-to-X genetic drift ratio.

It is worth pointing out that there were two African Americans samples that we could have used for SNPs discovery. We chose not to use Cor17119 for ascertaining SNPs in the present study, even though it was used to ascertain SNPs in a previous study[2], because this sample has a larger fraction of European ancestry than Cor17109, about 20% compared with 4%. In addition, in a principal component analysis[35] of data from an Affymetrix 6.0 array with approximately 1 million SNPs, we found that this individual is an ancestry outlier amongst African Americans (unpublished results).

To further explore the behavior of SNPs discovered from Cor17119, we compared the autosomal analyses for Cor17109 and Cor17119, and found that they gave highly concordant results (table above). However, two different runs of the ANCESTRYMAP software on chromosome X for Cor17119 produced unstable results, with a chunk of chromosome X in this individual indicated as being homozygous for African ancestry with high probability in one run and with low probability in a second run. For our X chromosome SNP ascertainment in two West African chromosomes, we therefore relied entirely on the results of our own genotyping of 1,087 SNPs ascertained between two West Africans (for the sake of consistency, we also excluded the autosomal Cor17119 data).

We emphasize that the results we previously obtained[2] based on SNPs ascertained in Cor17119 are not affected by this problem since these results were based on estimating autosomal F_{ST}, which as shown above is validated by both Cor17109 and our algebraic procedure based on European and African American data.

Nature Genetics: doi:10.1038/ng.303

References

1. Sabeti, P.C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-8 (2007).
2. Keinan, A., Mullikin, J.C., Patterson, N. & Reich, D. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. *Nat Genet* **39**, 1251-5 (2007).
3. Lahiri, S.N. *Resampling methods for dependent data*, (Springer, New York, 2003).
4. Lander, E.S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
5. McVean, G.T. & Hurst, L.D. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* **386**, 388-92 (1997).
6. 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).
7. Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome Res* **11**, 1725-9 (2001).
8. 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).
9. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299-320 (2005).
10. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
11. Patterson, N. et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* **74**, 979-1000 (2004).
12. Weir, B.S. & Cockerham, C.C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-1370 (1984).
13. Hudson, R.R. Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* **18**, 337-8 (2002).
14. Adams, A.M. & Hudson, R.R. Maximum-likelihood estimation of demographic parameters using the frequency spectrum of unlinked single-nucleotide polymorphisms. *Genetics* **168**, 1699-712 (2004).
15. Frisse, L. et al. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am J Hum Genet* **69**, 831-43 (2001).
16. Przeworski, M., Hudson, R.R. & Di Rienzo, A. Adjusting the focus on human variation. *Trends Genet* **16**, 296-302 (2000).
17. Voight, B.F. et al. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci U S A* **102**, 18508-13 (2005).
18. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585-95 (1989).

Nature Genetics: doi:10.1038/ng.303

19. Fay, J.C. & Wu, C.I. A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol* **16**, 1003-5 (1999).
20. Garrigan, D. & Hammer, M.F. Reconstructing human origins in the genomic era. *Nat Rev Genet* **7**, 669-80 (2006).
21. Kunsch, H.R. The jackknife and the bootstrap for general stationary observations *The Annals of Statistics* **17**, 1217-1241 (1989).
22. Liu, R.Y. & Singh, K. Moving blocks jackknife and bootstrap capture weak dependence. in *Exploring the limits of bootstrap* (eds. LePage, R. & Billard, L.) 225-248 (John Wiley, New York, 1992).
23. Ambrose, S.H. Chronology of the later stone age and food production in East Africa. *Journal of Archaeological Science* **25**, 377-392 (1998).
24. McBrearty, S. & Brooks, A.S. The revolution that wasn't: a new interpretation of the origin of modern human behavior. *J Hum Evol* **39**, 453-563 (2000).
25. Mellars, P. Going east: new genetic and archaeological perspectives on the modern human colonization of Eurasia. *Science* **313**, 796-800 (2006).
26. Yu, N. et al. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* **161**, 269-74 (2002).
27. Pool, J.E. & Nielsen, R. Population size changes reshape genomic patterns of diversity. *Evolution Int J Org Evolution* **61**, 3001-6 (2007).
28. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851-861 (2007).
29. Charlesworth, B. Background selection and patterns of genetic diversity in Drosophila melanogaster. *Genet Res* **68**, 131-49 (1996).
30. Charlesworth, B., Morgan, M.T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289-303 (1993).
31. Aquadro, C.F., Begun, D.J. & Kindahl, E.C. Selection, recombination and DNA polymorphism in Drosophila. in *Non-neutral evolution* (ed. Golding, B.) 46-56 (Chapman & Hall, New York, 1994).
32. Begun, D.J. & Whitley, P. Reduced X-linked nucleotide polymorphism in Drosophila simulans. *Proc Natl Acad Sci U S A* **97**, 5960-5 (2000).
33. Fay, J.C. & Wu, C.I. Hitchhiking under positive Darwinian selection. *Genetics* **155**, 1405-13 (2000).
34. Przeworski, M. The signature of positive selection at randomly chosen loci. *Genetics* **160**, 1179-89 (2002).
35. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904-9 (2006).

Nature Genetics: doi:10.1038/ng.303