nature genetics

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

Alon Keinan[1,2], James C Mullikin[3], Nick Patterson[2] & David Reich[1,2]

Comparisons of chromosome X and the autosomes can illuminate differences in the histories of males and females as well as shed light on the forces of natural selection. We compared the patterns of variation in these parts of the genome using two datasets that we assembled for this study that are both genomic in scale. Three independent analyses show that around the time of the dispersal of modern humans out of Africa, chromosome X experienced much more genetic drift than is expected from the pattern on the autosomes. This is not predicted by known episodes of demographic history, and we found no similar patterns associated with the dispersals into East Asia and Europe. We conclude that a sex-biased process that reduced the female effective population size, or an episode of natural selection unusually affecting chromosome X, was associated with the founding of non-African populations.

In a population with equal numbers of females and males, there are three copies of chromosome X for every four autosomes. As a consequence, in a population of constant size in the absence of natural selection, the average time since the most recent common genetic ancestor (tMRCA) of two unrelated individuals should be 3/4 on chromosome X of what it is on the autosomes, and allele frequency change on the autosomes should occur at 3/4 the rate of chromosome X. Population genetic studies in small datasets have found deviations from these expectations. For example, one study[1] found that the ratio of chromosome X–to-autosome diversity in non-African human populations was less than 3/4, whereas a later study observed an increased ratio in both African and non-African populations[2]. Genomic-scale human-variation datasets offer the possibility of more precise estimates[3,4], and have generally found more allele frequency differentiation on chromosome X between populations[3–6]. However, these observations have been difficult to interpret in terms of history or selection because the frequency distributions on chromosome X and the autosomes were biased by the different ways in which SNPs were selected from these two parts of the genome ('ascertainment bias')[3,7,8].

We compared patterns of variation on chromosome X and the autosomes using two uniformly collected, genome-scale datasets that we had originally assembled for the autosomes[8] and that we now extend to chromosome X. Both datasets are several orders of magnitude larger than previous datasets that have been available for comparing these two parts of the genome, allowing qualitatively new insights into history. The first dataset consists of about 130,000 SNPs that were discovered as differences between two chromosomes of samples of either West African, North European or East Asian ancestry, and then genotyped in more samples of all three ancestries[8]. Most of these data were mined from subsets of the International Haplotype Map (HapMap) using a strategy that produces allele frequency distributions that are indistinguishable from those obtained when the SNPs are discovered in two chromosomes of known ancestry[8]. We supplemented this with an additional 1,087 SNPs that we discovered between two West African copies of chromosome X and genotyped in our laboratory. The second dataset consists of over a billion base pairs of DNA that we compared between West Africans, North Europeans and East Asians to estimate sequence diversity within and between populations. Both datasets exclude exons and conserved noncoding sequences, as we were interested in learning about features of human variation that are not due to known effects of natural selection. These two datasets provide complementary information about history. Allele frequency data are fundamentally population based, whereas sequence diversity data reflect the history of individual DNA sequences (the time that has elapsed since they shared a common ancestor; tMRCA). For example, allele frequency differences between two populations are affected only by the history after the populations split, whereas sequence diversity within and between these populations is also affected by the history of the ancestral population.

For our first line of analysis, we used the uniformly collected SNP data and allele frequency differentiation measurements (F_{ST}) to estimate the amount of genetic drift that occurred between pairs of populations since they split. By analyzing SNPs discovered between two chromosomes from population A, it is possible to estimate the genetic drift that has occurred between population A and a second population B in a way that is unbiased by the history of expansions and contractions in B (Methods and **Supplementary Note** online). The only requirement for this theory to hold is that population A has been effectively of constant size since the split from B, a requirement that is approximately satisfied in the context of our analysis

[1]Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [3]Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. Correspondence should be addressed to A.K. (akeinan@genetics.med.harvard.edu) or D.R. (reich@genetics.med.harvard.edu).

Received 10 July; accepted 3 November; published online 21 December 2008; doi:10.1038/ng.303

Table 1 Frequency differentiation between Africans and non-Africans is higher on chromosome X than is expected from the autosomes

	Autosomes		Chromosome X		Comparison of autosomes and chromosome X	
	SNPs	F_{ST} (error)	SNPs	F_{ST} (error)	Observed autosome-to-X genetic drift ratio Q (error)	P value for observed versus expected 3/4
North European	62,830	0.106	2,668	0.133	0.771	0.57
– East Asian		(.002)		(.006)	(.036)	
East Asian	45,423	0.098	1,247	0.131	0.715	0.48
– North European		(.002)		(.009)	(.050)	
West African	13,606	0.178	1,087	0.256	0.615	6.5×10^{-6}
– East Asian		(.003)		(.010)	(.030)	
West African	13,606	0.144	1,087	0.221	0.582	3.0×10^{-8}
– North European		(.003)		(.009)	(.030)	

Analyses are based on SNPs ascertained in two chromosomes of individuals from the ancestry indicated in boldface. With this ascertainment strategy, $Q = \ln(1 - 2F_{ST}^{auto}) / \ln(1 - 2F_{ST}^{X})$ estimates the ratio of autosome-to-X genetic drift. This estimate is independent of the demographic history of the population not used to ascertain SNPs, so that unlike the case of the sequence diversity ratio (**Table 2**), we do not need to carry out modeling of demographic history to determine the autosome-to-X expected relationship. For the comparisons involving West Africans, the autosome-to-X genetic drift ratio is significantly below the expected 3/4 (standard errors are based on 1,000 moving block bootstraps and account for linkage disequilibrium between SNPs; **Supplementary Methods** online).

(**Supplementary Note**). To verify these theoretical predictions, we carried out simulations of scenarios of population expansion and contraction (including scenarios of gene flow) that fit the allele frequencies in all three populations[8], and found that the simulated autosome-to-X genetic drift ratio is consistent with the expected 3/4 (Methods and **Supplementary Note**). Empirically, we applied this approach to our data and estimated that the autosome-to-X genetic drift ratio between North Europeans and East Asians is consistent with the expected 3/4 (**Table 1**; $P \sim 0.5$). However, it is significantly lower than 3/4 between North Europeans and West Africans (0.582 ± 0.030; $P = 3.0 \times 10^{-8}$), and between East Asians and West Africans (0.615 ± 0.030; $P = 6.5 \times 10^{-6}$) (**Table 1**), a reduction that is also significant when compared with simulations of realistic demographic histories[8] (**Supplementary Note**). These results point to a period of accelerated drift on chromosome X that largely occurred after the split of West Africans and non-Africans, but before the separation of North Europeans and East Asians.

For our second line of analysis, we focused on SNP allele frequency distributions within populations (no longer comparing across populations). The shape of the chromosome X allele frequency distribution differs significantly from that on the autosomes, particularly for the non-African populations ($P < 10^{-11}$; **Supplementary Fig. 1** and **Supplementary Note** online). To examine whether known features of demographic history can account for these differences[9,10], we examined models of history that we previously showed provided a good fit to autosomal data[8]. Although these models fit the West African data after adjustment by a 3/4 ratio of X-to-autosome population size (**Fig. 1**), the fit is poor for the North European and East Asian populations, with both populations harboring more high-frequency derived alleles on chromosome X than expected (**Fig. 1**). To formally test for stronger chromosome-X drift in these populations, we separately fit a model of a population bottleneck to the allele frequencies of the autosomes and chromosome X and found evidence for a much more intense bottleneck on chromosome X than is expected on the basis of adjustment for a 3/4 difference in population size ($P \ll 10^{-12}$ for North Europeans and $P = 6 \times 10^{-4}$ for East Asians; **Fig. 1**, **Supplementary Table 1** and **Supplementary Note** online). These results provide independent support for an acceleration in the rate of genetic drift on chromosome X since the separation of African and non-African populations, and also show that the pattern is observed independently of any data from African populations. Putting the two lines of evidence together, we conclude that the accelerated genetic drift on chromosome X occurred in the ancestral population of non-Africans, after the split from West Africans.

For our third line of analysis, we examined the sequence diversity data from both chromosome X and the autosomes by counting the number of differences per base pair between two unrelated DNA sequences and then translating to estimates of time (tMRCA) by

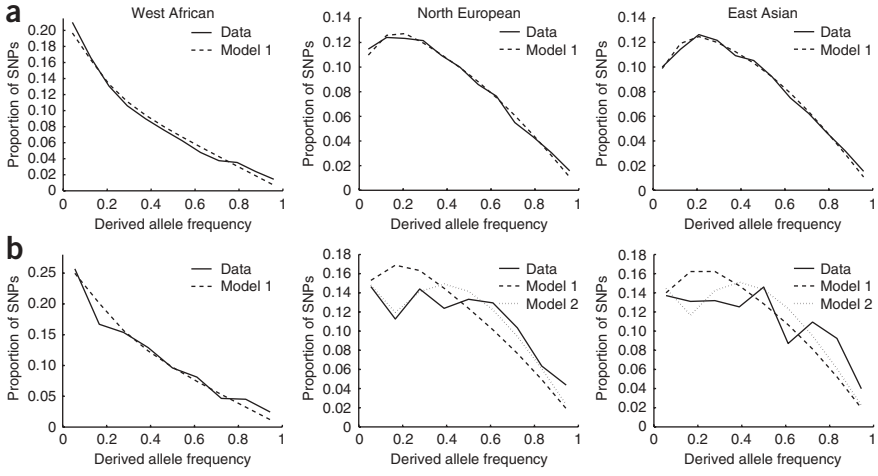

Figure 1 The distribution of allele frequencies on chromosome X does not match the expectation from the autosomes in non-African populations. (**a**) Distribution of derived allele frequencies on the autosomes compared with the expectation for our best-fit models of history[8], for SNPs discovered in two chromosomes of an ancestry and then genotyped in a larger number of samples of the same ancestry from HapMap. Derived allele refers to the new mutation, which we infer by requiring that both chimpanzee and orangutan match the other allele. (**b**) Analogous plots for chromosome X compare the observed data to the expectation from the autosomal best-fit models after rescaling all effective population sizes by 3/4 ('Model 1'). Although the fit of the model to the chromosome X data is excellent for West Africans, the fit of the models is poor in North Europeans and East Asians, both of which show more high-derived allele frequencies than expected from the models. 'Model 2' separately fits the out-of-Africa bottleneck on chromosome X, resulting in a more intense bottleneck than is expected from the autosomes even after adjustment for the 3/4 expected difference in population size (**Supplementary Note**). The fit to the observed data is greatly improved in Model 2 compared to Model 1, with mean squared errors reduced by 79% for North Europeans and 34% for East Asians. (The panels in this figure use different scales reflecting the numbers of sampled chromosomes used in each analysis; modeling adjusts for the differences in sample size[8].)

Table 2 Ratio of sequence diversity on chromosome X to the autosomes is reduced outside Africa

	Autosomes		Chromosome X		Comparison of autosomes and chromosome X		
	Aligned bases (millions)	Divergent sites/base pair ($\times 10^{-3}$) (error)	Aligned bases (millions)	Divergent sites/base pair ($\times 10^{-3}$) (error)	Observed X-to-autosome ratio normalized by macaque divergence	Expected X-to-autosome ratio from best-fit demographic model	*P* value for difference between observed and expected
West African	641.7	1.081 (.005)	22.0	0.722 (.017)	0.763 (.026)	0.780 (.001)	0.514
North European	657.2	0.827 (.004)	34.6	0.460 (.013)	0.635 (.024)	0.702 (.004)	0.005
East Asian	296.8	0.772 (.005)	23.5	0.414 (.014)	0.613 (.026)	0.690 (.004)	0.003

We aligned sequence from unrelated individuals from each population to calculate genetic diversity per base pair (Methods and **Supplementary Table 4**). Standard errors are obtained by a jackknife method to account for correlated diversity in neighboring regions of the genome (Methods). The X-to-autosome ratio is obtained by dividing the chromosome X diversity column by the autosome diversity column, normalizing each by human-macaque divergence in the same section of the genome, and combining all standard errors appropriately (**Supplementary Methods**). We also compared the observed data to the expectation from the best-fit autosomal demographic model[8] (**Supplementary Note**; standard errors reflect uncertainty in model fitting). Tests for significance use a two-sided *z* test between the observed and expected value.

normalizing by an outgroup (human-macaque divergence) to adjust for differences in mutation rates between these two parts of the genome. (The tMRCA of human and macaque is expected to be slightly less for chromosome X because of ancestral polymorphism, which upwardly biases our estimate of the X-to-autosome tMRCA ratio and is conservative for our analyses.) The ratio of the tMRCA between chromosome X and the autosomes in West Africans, 0.763 ± 0.026, is consistent with the expected 3/4, but it is lower than 3/4 in non-African populations: 0.635 ± 0.024 in North Europeans and 0.613 ± 0.026 in East Asians (**Table 2**, **Supplementary Note** and **Supplementary Table 2** online). We assessed whether this lower ratio could be explained by known features of human history, including the out-of-Africa bottleneck that occurred during this period[11]. Although the demographic models[8] predict reductions below 3/4 (0.702 ± 0.004 for North Europeans and 0.690 ± 0.004 for East Asians), the observed ratios are significantly below these values ($P = 0.005$ for North Europeans, $P = 0.003$ for East Asians; **Table 2**). The reduction is significant for several models of history we considered, further supporting the hypothesis of accelerated chromosome X drift, and showing that it occurred in non-African history after the split from Africans (**Supplementary Note**).

We have used three independent lines of evidence to show that there was a period of intense chromosome X drift in the history of non-Africans, during which the effective population size on chromosome X was transiently reduced below the expected 3/4 of the autosomes. This process seems to have largely occurred after the ancestors of West Africans split from the ancestors of non-Africans, but before the divergence of North Europeans and East Asians. We found no similar acceleration of chromosome X drift associated with other major human migrations to new environments: the autosome-to-X drift ratio comparing North Europeans and East Asians (**Table 1**) and Chinese and Japanese (**Supplementary Note**) are both consistent with 3/4.

Deviations in the X-to-autosome ratio from 3/4 have been documented in several species, especially in *Drosophila*, where natural selection is usually hypothesized to be the explanation[12–14] as recessive

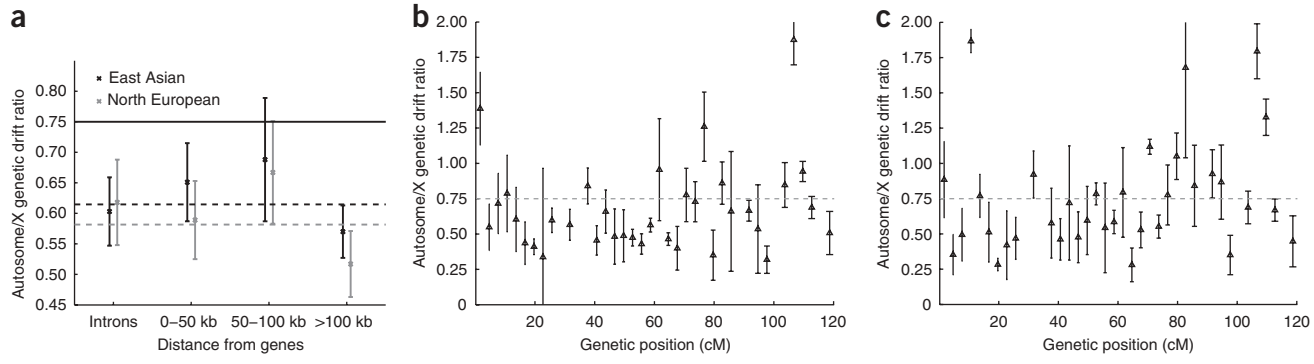

Figure 2 Gene-centric natural selection, or natural selection localized to specific regions of chromosome X, fail to explain the signal of accelerated genetic drift. (**a**) Dividing the chromosome X and autosome datasets on the basis of distance from the nearest gene, we find no evidence that the ratio of autosome-to-X genetic drift, *Q*, between West Africans and East Asians and between West Africans and North Europeans increases with distance, as would be expected if selection explains our results. In particular, the bin furthest from genes (>100 kb) is significantly below the expectation of 3/4 (horizontal solid black line; error bars indicate one standard error below and above *Q*). Dotted lines show the values for all data for comparison to the individual bins. A related plot for the sequence diversity data also shows no significant effect of distance from genes (**Supplementary Fig. 2**). (**b**,**c**) The ratio of autosome-to-X genetic drift, *Q*, between West Africans and North Europeans (**b**) and between West Africans and East Asians (**c**) is presented spatially across chromosome X. For each of 40 bins 3 cM in size on chromosome X, *Q* is presented as the reciprocal of the ratio of genetic drift in that bin to the average across the autosomes. Each bin contains between 15 and 60 SNPs, with bins with fewer SNPs excluded from the figure. Error bars indicate one standard error below and above *Q*, which we obtain by a jackknife over 5 windows of 0.6 cM. The reduction of *Q* below the expectation of 3/4 (horizontal line) has a complex pattern. However, it is widely distributed across chromosome X rather than localized to a few regions. This rules out the possibility that the signal of accelerated drift on chromosome X can be explained by selection at a small number of loci.

alleles on chromosome X are exposed in hemizygous males[15,16], although demographic processes have also been explored[14,17]. In humans, the dispersal out of Africa was a period when humans were moving to new environments, and thus is a time when selective pressures could have changed[18].

To search for additional features of the data that could shed light on natural selection, we carried out three analyses. First, we marked all the SNPs by their distance from genes, but found no attenuation of the signal with increased distance (**Fig. 2a** and **Supplementary Fig. 2** online), indicating that selection affecting a large proportion of chromosome X genes is unlikely to explain our results. However, our results could potentially still be explained by intense selection that acted over a larger distance scale than we could measure (**Fig. 2a** and **Supplementary Note**). Second, we studied the physical distribution of accelerated drift across chromosome X, and found that it is widespread across the chromosome (**Fig. 2b**,**c**), ruling out the possibility that selection was focused on just a few loci. Third, we did not find any evidence that the difference between chromosome X and the autosomes is explained by selection on mutations that newly arose after the split of Africans and non-Africans, as this would be expected to produce characteristic effects on the allele frequency distributions and diversity data in North Europeans and East Asians, which we did not observe (**Supplementary Note**). We conclude that if selection explains our results, it is likely to have been on preexisting alleles across chromosome X ('standing variation') that were nearly neutral in the ancestral environment and only came under selection after the split of non-Africans from West Africans. Although this is an extreme scenario (and we see no similar signal associated with other major human dispersals into new environments such as North Europe and East Asia), we cannot rule out this possibility.

Sex-biased demographic processes provide an alternative class of explanations for the observed skew from an autosome-to-X genetic drift ratio of 3/4, as men carry only half the number of X chromosomes as women. Sex-biased migration is particularly plausible: anthropological studies of hunter gatherers have shown that female migration usually dominates at short distances[19,20] and male migration at longer distances[19,21], and sex-biased migration is also documented in modern populations[22–25]. Our observations could be explained if after the ancestral population of non-Africans was established, it received long-range male migration from an African source (either quickly or over thousands of years), which retarded drift twice as effectively on the autosomes as on chromosome X. Another sex-biased demographic process that could contribute to our observations is if women had a much different generation time than men, which might reduce the effective population size on chromosome X relative to the autosomes. Our observations rule out variability in reproductive success as sufficient to explain our observations[26,27]. Although polygyny (males having multiple female mates) is a sex-biased process that has been observed in some human populations[2], it would predict a rise in the ratio above 3/4, which is opposite to what we observe. The alternative of a tiny fraction of women having almost all the offspring during the out-of-Africa dispersal could in principle explain our observations[26,27], but is inconsistent with the pattern seen in human populations including hunter gatherers[21], and is implausible considering the large investment females place in childbirth and child rearing. Our observations could also be consistent with other sex-biased demographic processes, and future work should explore these scenarios.

We have shown that there was a period of accelerated genetic drift on chromosome X associated with the human dispersal out of Africa, which was qualitatively different from what occurred during the subsequent human dispersals into Northern Europe and East Asia. Our results are also of methodological interest. Chromosome Y and mitochondrial DNA (mtDNA) are usually analyzed to study sex-biased demographic events. However, these loci provide limited resolution about ancient demographic processes. For example, few independent non-African lineages stretch back all the way to the time of the out-of-Africa dispersal, and estimates of sex-biased demographies at that time therefore have large errors. By contrast, chromosome X and the autosomes encompass thousands of independent genetic loci, each of which probes more ancient times than the mtDNA and the Y chromosome. By averaging measurements over these loci, it is possible to obtain high resolution measurements of sex-biased processes, revealing previously undocumented events in history.

METHODS

SNP data mined from HapMap. For analyses of allele frequency differentiation across populations and allele frequency distributions, we used subsets of SNPs from HapMap (Public Release #21a), which were ascertained uniformly across the genome so that the datasets were appropriate for population genetic analysis[8]. All the SNPs in our study are ascertained as divergent sites in exactly two chromosomes of the same ancestry (two each of either West African, North European or East Asian ancestry) and genotyped in all HapMap samples, including 120 unrelated West African chromosomes from Ibadan, Nigeria (YRI), 120 unrelated European American chromosomes from Utah, USA (of North European ancestry; CEU), and 180 unrelated East Asian chromosomes (90 Han Chinese from Beijing, China (CHB) and 90 Japanese from Tokyo, Japan (JPT), which we pooled for most analyses). As males carry a single copy of chromosome X, the counts for chromosome X are at most 90, 90, and 135 (our modeling adjusts appropriately for the sample size of every SNP used in our analysis[8]). We removed all sites that were in hypermutable CpG dinucleotides, and determined the ancestral allele by requiring a match to both the chimpanzee and orangutan sequence[8]. In addition to expanding the data from ref. 8 to chromosome X (**Supplementary Table 3** online), we made two improvements. First, we no longer required SNPs to be discovered in two chromosomes from the same individual; instead, we allowed SNPs to be discovered by comparison of two chromosomes, one from each of two individual of the same ancestry, which we found generates an indistinguishable frequency distribution. Second, for the autosomal SNPs discovered in West Africans, we used data from an African American sample (NA17109) who we determined had 4% European ancestry on average using the ANCESTRYMAP software[28]. We restricted the SNPs used for analysis to sections of this individual's genome where we were $>95\%$ confident of African ancestry in both chromosomes, as determined by an analysis with ANCESTRYMAP. In the **Supplementary Note**, we present analyses showing that this procedure generates results that are indistinguishable from those obtained using two chromosomes from a West African.

SNP genotyping. As the African American sample (NA17109) used for mining SNPs in HapMap was male, we could not use this individual to identify sites that were different between two West African copies of chromosome X. To fill this gap, we used four West African (YRI) samples (NA18517, NA18507, NA19240 and NA19129) for whom shotgun sequencing data were available in public databases. We randomly dropped sequencing reads until we had no more than two unrelated chromosomes at each site, and then used ssahaSNP[29] to identify 4,884 SNPs on chromosome X for which we could confidently identify the ancestral allele by comparison to both chimpanzee and orangutan, and for which we were able to successfully design primers, and which passed all the other filters we applied to the SNPs mined from HapMap. We attempted to genotype a randomly chosen subset of 1,366 of these SNPs in all HapMap samples using the Sequenom iPLEX method[30]. After removal of SNPs with $<85\%$ genotyping completeness or four or more heterozygous genotypes in males, or that were out of Hardy-Weinberg equilibrium ($P < 0.01$ in a statistic combined across populations) or monomorphic, we were left with 1,087 SNPs. From the 210 unrelated samples in HapMap, we filtered to 189 after dropping

samples that had <85% genotyping completeness, an excess of heterozygous genotypes in males or a deficiency of heterozygous genotypes in females compared with others from the same population, or that were one of six YRI samples related to those used in SNP discovery. This left us with 77 YRI, 82 CEU and 122 CHB + JPT X chromosomes for analysis.

Sequence diversity data. We obtained DNA sequence reads from public databases for five individuals of North European ancestry (European Americans), four of East Asian ancestry (from China, Japan and the United States) and five of West African ancestry (four Nigerians and one African American, whose genome was only analyzed in sections where we were >95% confident of two African-origin chromosomes, as determined by the ANCESTRYMAP software[28]) (**Supplementary Table 4** online). To identify divergent sites using these sequence reads, we aligned them to Build 35 of the human reference sequence by ssahaSNP[29] with the settings Qsnp $\geq$ 40, Qneighbor $\geq$ 15, Nneighbor = 5, maxNeighborhoodDiffs = 1, maxSNPs/kb = 15. A subset of nonoverlapping sequence reads for each individual was selected at random, providing a single mosaic, haploid genome that was not biased according to the strand of DNA from which a read derived. Within-population sequence diversity was estimated by only analyzing bases where there were two or more individual haploid genomes, and then counting differences by selecting two haploid genomes at random. To similarly compute between-population diversity (**Supplementary Table 2**), we combined the individual haploid genomes of each population so that at any base only one individual was represented. To estimate standard errors correcting for correlation between neighboring sites, we used a jackknife analysis, dividing the genome into blocks of 100,000 aligned bases, and removing each block in turn[8].

Allele frequency differentiation analysis. To estimate allele frequency differentiation across populations, we used the F_{ST} statistic as formulated in ref. 8. Briefly, when (i) a SNP is discovered as polymorphic in population A, and (ii) population A has been of effectively constant size since the split from population B, the expected value of F_{ST} is $E(F_{ST}^{auto}) = (1 - e^{-(\tau_A+\tau_B)})/2$, where τ_A and τ_B are scaled drift times. Multiplying τ_i by 4/3, the equivalent expression for chromosome X is $E(F_{ST}^{X}) = (1 - e^{-4/3(\tau_A+\tau_B)})/2$, thus $Q = \ln(1 - 2E(F_{ST}^{auto}))/\ln(1 - 2E(F_{ST}^{X})) = 3/4$. To test this expectation, we simulated models of history for each pair of populations and found all values to be close to 3/4 (**Supplementary Note**).

URLs. SNP datasets, http://genepath.med.harvard.edu/~reich.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank C. Aquadro, O. Bar-Yosef, M. Bernstein, B. Charlesworth, A. Helgason, E. Lander, D. Lieberman, K. Lohmueller, S. Myers, S. Pääbo, A. Price, S. Schaffner and C. Stringer for comments, and J. Neubauer and A. Waliszewska for genotyping SNPs discovered in two West African X chromosomes. The orangutan sequence reads were generated by the Washington University genome sequencing center (ftp.ncbi.nih.gov/pub/TraceDB/ pongo_pygmaeus_abelii); we thank R. Wilson for permission to use these data. J.C.M. was supported by the Intramural Research Program of the US National Human Genome Research Institutes (NHGRI), N.P. by a K-01 career transition award from the US National Institutes of Health (NIH) and D.R. by a Burroughs Wellcome Career Development Award in the Biomedical Sciences. A.K., N.P. and D.R. were also supported by NIH grant U01 HG004168.

AUTHOR CONTRIBUTIONS

A.K. and D.R. designed the study; J.C.M. and A.K. assembled datasets; D.R., A.K. and J.C.M. conducted genotyping; A.K., J.C.M. and D.R. performed analyses; N.P. provided guidance on statistical analyses; A.K. and D.R. interpreted results and wrote the manuscript, which was edited by all co-authors.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hammer, M.F. *et al.* Heterogeneous patterns of variation among multiple human x–linked loci: the possible role of diversity-reducing selection in non-Africans. *Genetics* **167**, 1841–1853 (2004).
2. Hammer, M.F., Mendez, F.L., Cox, M.P., Woerner, A.E. & Wall, J.D. Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet.* **4**, e1000202 (2008).
3. The International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
4. Li, J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
5. Akey, J.M., Zhang, G., Zhang, K., Jin, L. & Shriver, M.D. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* **12**, 1805–1814 (2002).
6. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
7. Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. & Nielsen, R. Ascertainment bias in studies of human genome-wide polymorphism. *Genome Res.* **15**, 1496–1502 (2005).
8. 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**, 1251–1255 (2007).
9. 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**, 1003–1005 (1999).
10. Garrigan, D. & Hammer, M.F. Reconstructing human origins in the genomic era. *Nat. Rev. Genet.* **7**, 669–680 (2006).
11. Pool, J.E. & Nielsen, R. Population size changes reshape genomic patterns of diversity. *Evolution Int. J. Org. Evolution* **61**, 3001–3006 (2007).
12. Andolfatto, P. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila* simulans. *Mol. Biol. Evol.* **18**, 279–290 (2001).
13. Betancourt, A.J., Kim, Y. & Orr, H.A. A pseudohitchhiking model of X vs. autosomal diversity. *Genetics* **168**, 2261–2269 (2004).
14. Wall, J.D., Andolfatto, P. & Przeworski, M. Testing models of selection and demography in *Drosophila* simulans. *Genetics* **162**, 203–216 (2002).
15. Charlesworth, B., Coyne, J.A. & Barton, N.H. The relative rates of evolution of sex chromosomes and autosomes. *Am. Nat.* **130**, 113–146 (1987).
16. Charlesworth, B., Morgan, M.T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303 (1993).
17. Haddrill, P.R., Thornton, K.R., Charlesworth, B. & Andolfatto, P. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res.* **15**, 790–799 (2005).
18. Storz, J.F., Payseur, B.A. & Nachman, M.W. Genome scans of DNA variability in humans reveal evidence for selective sweeps outside of Africa. *Mol. Biol. Evol.* **21**, 1800–1811 (2004).
19. Kayser, M. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
20. Seielstad, M.T., Minch, E. & Cavalli-Sforza, L.L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280 (1998).
21. Marlowe, F.W. Hunter-gatherers and human evolution. *Evol. Anthropol.* **14**, 54–67 (2005).
22. Bedoya, G. *et al.* Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. *Proc. Natl. Acad. Sci. USA* **103**, 7234–7239 (2006).
23. Hammer, M.F. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
24. Helgason, A. *et al.* Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am. J. Hum. Genet.* **67**, 697–717 (2000).
25. Parra, E.J. *et al.* Estimating African American admixture proportions by use of population-specific alleles. *Am. J. Hum. Genet.* **63**, 1839–1851 (1998).
26. Caballero, A. On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**, 1007–1011 (1995).
27. Charlesworth, B. The effect of life-history and mode of inheritance on neutral genetic variability. *Genet. Res.* **77**, 153–166 (2001).
28. Patterson, N. *et al.* Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000 (2004).
29. Ning, Z., Cox, A.J. & Mullikin, J.C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
30. Tang, K. *et al.* Chip-based genotyping by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **96**, 10016–10020 (1999).