Y chromosome diversity, human expansion, drift, and cultural evolution

Jacques Chiaroni[a], Peter A. Underhill[b], and Luca L. Cavalli-Sforza[c,1]

[a]Unité Mixte de Recherche 6578, Centre National de la Recherche Scientifique, and Etablissement Français du Sang, Biocultural Anthropology, Medical Faculty,Université de la Méditerranée, 13916 Marseille, France; [b]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA 94304-5485; and [c]Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305-5120

Contributed by Luca L. Cavalli-Sforza, September 26, 2009 (sent for review May 26, 2009)

The relative importance of the roles of adaptation and chance in determining genetic diversity and evolution has received attention in the last 50 years, but our understanding is still incomplete. All statements about the relative effects of evolutionary factors, especially drift, need confirmation by strong demographic observations, some of which are easier to obtain in a species like ours. Earlier quantitative studies on a variety of data have shown that the amount of genetic differentiation in living human populations indicates that the role of positive (or directional) selection is modest. We observe geographic peculiarities with some Y chromosome mutants, most probably due to a drift-related phenomenon called the surfing effect. We also compare the overall genetic diversity in Y chromosome DNA data with that of other chromosomes and their expectations under drift and natural selection, as well as the rate of fall of diversity within populations known as the serial founder effect during the recent "Out of Africa" expansion of modern humans to the whole world. All these observations are difficult to explain without accepting a major relative role for drift in the course of human expansions. The increasing role of human creativity and the fast diffusion of inventions seem to have favored cultural solutions for many of the problems encountered in the expansion. We suggest that cultural evolution has been subrogating biologic evolution in providing natural selection advantages and reducing our dependence on genetic mutations, especially in the last phase of transition from food collection to food production.

migration | natural selection | serial founder effect | isolation by distance

The discussion of the relative importance of natural selection and random genetic drift (the effect of population size N in determining genetic diversity of individuals and populations) is as old as population genetics. A book published by the Hagedoorns in 1921 (1) proclaimed the evolutionary importance of what we now call drift. It was critically reviewed by Fisher (2), who stated that what he called the Hagedoorn effect might be observed perhaps only in small island populations, but species sizes are too large to show drift effects. Wright made other major contributions to the mathematic theory of drift, and introduced the name (3).

Kimura (and later many others) extended the theory greatly, suggested the explanatory expansion name "random genetic drift," and showed (4) that if mutations are (largely) free of selective effects, the rate of molecular evolution would be determined by the mutation rate. This caused controversy, and a dedicated symposium (5) was convened to discuss it. Although there is no expressed agreement that most mutations are selectively neutral or nearly so, well-ascertained phenomena like the high occurrence of silent nucleotide substitutions in genes, and the likelihood that large DNA sections in Eukaryotes are inactive on the phenotype or nearly so, lend credibility to the basic correctness of Kimura's statement. Drift, measured by population size N and migration m among populations, has similar effects on genetic diversity among populations, and it simplifies matters to consider their global effect Nm. The smaller the Nm of a population, the smaller is the "component of variation" of gene frequencies, and the larger the "between." Our species allows accurate estimates of the relevant demographic parameters, as well as of the intensity of natural selection by measurements of survival and fecundity, and is thus an excellent target for a quantitative analysis of the relative importance of natural selection vs. Nm, and thus of genetic adaptation vs. chance.

In one of the earliest studies of drift, 3 blood group systems were tested on 2,875 individuals of an Italian population, coming from 74 villages of the Parma Valley, whose demography was known for almost 5 centuries thanks to the availability of parish books (6–8). Areas with smallest village size (in the mountains) showed the highest heterogeneity of gene frequencies among villages and the smallest within-population variance, which could be predicted very accurately by computer simulations using only N and m, thus leaving practically no evidence of natural selection on the loci analyzed. Variation among larger villages or towns showed no significant variance among populations.

The most easily visible effect of drift is decrease of heterozygosity of a population with respect to that expected under random mating, with decreasing Nm. It was observed that the average heterozygosity calculated for each of 51 aboriginal populations from the 5 continents (9) showed a linear fall of genetic diversity with the geographic distance of each population from the origin in East Africa. The fall of this linear decay was extremely regular: the correlation between average heterozygosity and geographic distance from Africa was $r = -0.89$, with a total of 783 microsatellites. With 650,000 Illumina SNPs the correlation was $r = -0.91$ (10).

It was hypothesized that this regular reduction of genetic diversity within populations, measured by average heterozygosity of all genes in each population, was the consequence of the "Out of Africa" expansion, which most likely consisted of the succession of founders of small new colonies into virgin territory near the source populations and was called the serial founder effect (9). For each of these colonizing episodes the impact of drift would have been inversely proportional to the size of the colonizing group but would be reduced by successive migration, and these founder effects must have accumulated over distance from the origin.

Simulations (10, 11) involving repetitions of this founder effect at each of the hundreds or thousands of colonizing steps that occurred in the expansion of modern humans from East Africa to the most distant place (Chile), using anthropologically credible demography, showed that the magnitude of the observed slope of the decay of genetic diversity was reasonable.

Further evidence of the greater relative importance of drift vs. natural selection became available when the difference in average gene frequency of all pairs of the same populations used in refs. 10 and 11, calculated by the standard measure Fst, showed a very high linear correlation with the geographic distance between the members of each population pair. The correlations were $r = 0.87$ and $r = 0.90$. This leaves very little room for the effect of natural selection, which was estimated to be on the order of 20% (10) and should be

Author contributions: L.L.C.-S. designed research; J.C. performed research; J.C., P.A.U., and L.L.C.-S. analyzed data; and J.C., P.A.U., and L.L.C.-S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: cavallisforza@gmail.com.

This article contains supporting information online at www.pnas.org/cgi/content/full/0910803106/DCSupplemental.

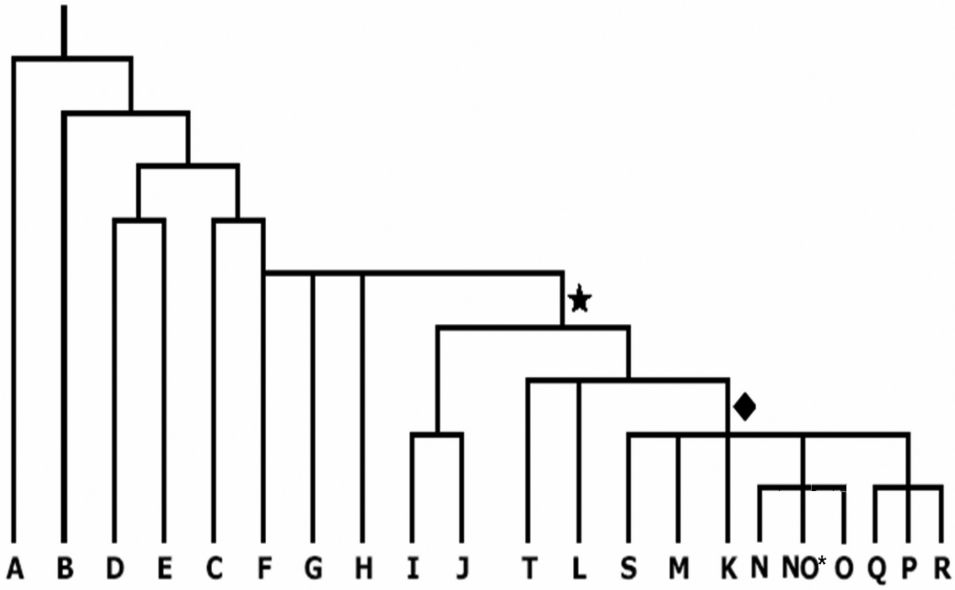

Fig. 1. Current phylogenetic relationships of the 20 major haplogroups of the global Y chromosome gene tree. Star denotes the new topology formed by the M522 and M523 SNPs that now join the previously independent IJ-M429 and KT-M9 haplogroups. Diamond indicates position of M526 SNP that now unifies haplogroups KMNOPS. M522, M523, and M526 marker specifications are given in Table S3.

smaller considering that this estimate should be decreased by considering errors due to smallness of the samples and other error sources.

This research is dedicated to using the genetic variation of the Y chromosome, from published data [supporting information (SI) Table S1] on 45,864 individuals from 937 global populations, for bringing more light to these and related problems.

Results and Discussion

Haplogroups: Their Phylogenesis and Geographic Distributions. The original nomenclature system of Y chromosome genotypes standardized in 2002 defined 18 haplogroups, labeled A through R (12). There are now 20 main haplogroups, designated A through T (13). Their basic phylogenetic relationships are shown in Fig. 1. Two deep branches, unifying haplogroups IJK and MNOPS, are shown. The geographic distributions of frequencies of each haplogroup were calculated on the basis of published data (listed in Table S1) and are shown in Fig. 2 for the 15 numerically and geographically most important haplogroups. The K, M, and S haplogroup distributions are not shown because they occupy rather unique space, mostly in Oceania. The F* and P* haplogroups are paraphyletic and often rare, except for F* in India, which codistributes with H, and are also omitted.

Haplogroups A and B are the deepest branches in the phylogeny and are essentially restricted to Africa, bolstering the evidence that modern humans first arose there (14, 15). Haplogroup A is mainly found in the Rift Valley from Ethiopia to Cape Town, mostly but not exclusively in some of the oldest hunter-gatherers who still survive and speak Khoikhoi and San languages, proposed by some to be the oldest languages. The interruption of its distribution in the middle of the Rift Valley is possibly the consequence of replacement by Bantu-speaking farmers who settled the region starting in the first millennium of the Christian era. Haplogroup B is found mainly among African Pygmies, who live in the central African forest and are still predominantly hunters-gatherers but speak Bantu languages borrowed from farmers who arrived in the area between 2,000 and 3,000 years ago. The third predominantly African haplogroup, E, diversified some time afterward, probably descending from the East African population that generated the Out of Africa expansion. The geographic distributions of the major branches of this haplogroup, given in Fig. S1*b*, suggest that most of the settlement outside of Africa by haplogroup E members involves the later mutant E-M35 varieties like M78, M81, and M123 that extended to Arabia and the northern Mediterranean coast.

Although the precise phylogenetic relationships among haplogroups F–H remain uncertain (Fig. 1), 3 phylogenetic unifying relationships in the interior framework of the phylogeny reveal shared patterns of common molecular heritage. First, the CF-P143 common ancestor shows an ancient split from the DE-YAP clade, clarifying the earliest diversification events outside Africa (Fig. S2*a*). Second, the unification of haplogroups IJK by M522 and M523 now creates evolutionary distance from F–H representatives, as well as supporting the inference that both IJ-M429 (Fig. S3*a*) and KT-M9 arose closer to the Middle East than central Asia or eastern Asia. Third, the very successful and widespread MNOPS-M526 (Fig. S3*a*) ancestor indicates that haplogroups L and T have individual ancestry consistent with their current distinctive geography (Fig. S3*a*). The phylogeographic patterns shown in the 4 last rows in Fig. 2 are consistent with the Out of Africa expansion approximately 60,000 years ago, characterized by rapid dispersal across Eurasia and Oceania and followed by subsequent isolation (16).

Haplogroups G–J, T, and L generally are constrained geographically near Europe, the Middle East, and parts of western Asia with various extensions, especially to Arabia and India. Haplogroups D, H, T, and L have more localized distributions and often lower frequencies. The haplogroups with the greatest spatial distribution are those associated with widespread haplogroups NO and PQR ancestry. Haplogroup R peopled central and western Asia and a great part of Europe. Haplogroup N is frequent across boreal and northwest Asia and O in southeast Asia including the islands and part of New Guinea.

Haplogroups C and Q display Asian ancestry and hold the unique privilege of having settled America. Not surprisingly their origin seems to have been in northeast Asia. The absence of haplogroup N in the Americas indicates that its spread across Asia happened after the submergence of the Bering land bridge. It is likely that haplogroup C entered America after Q, even though C originated phylogenetically earlier than Q. In fact, C is found only in the northern part of North America, limited to the area where the

ANTHROPOLOGY

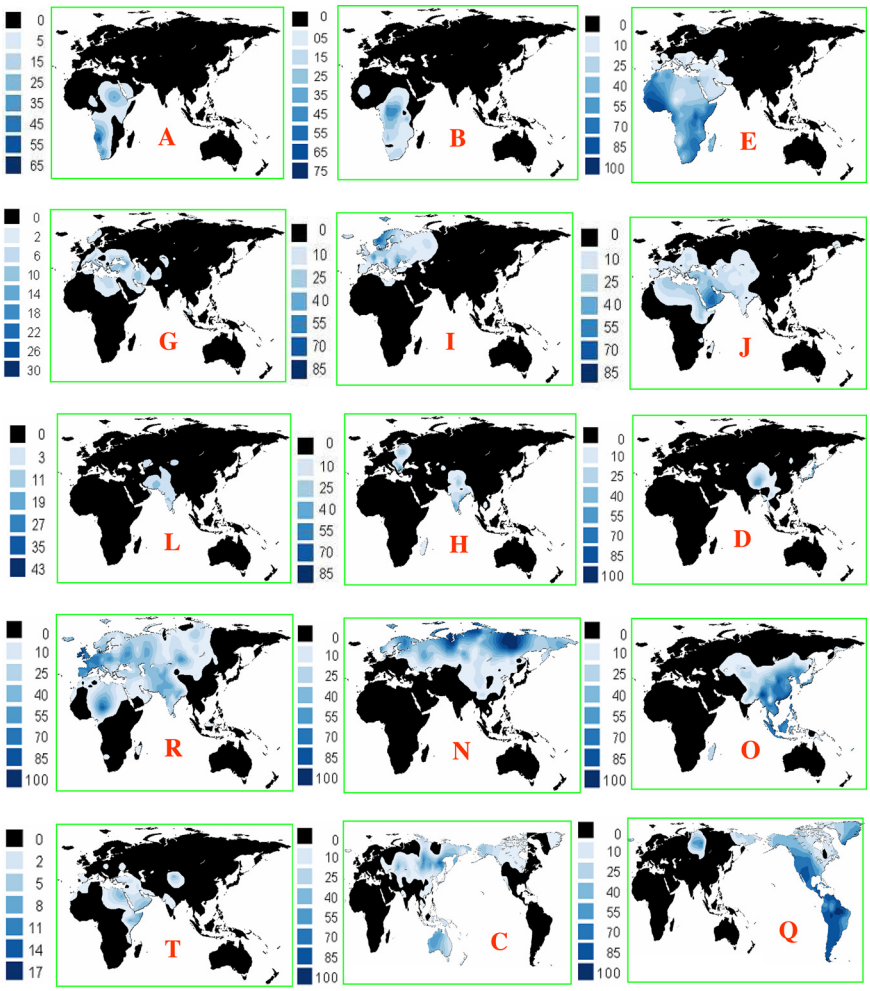

Fig. 2. Y chromosome haplogroup geographic frequency distribution maps.

family of languages by American Indians called Na-Dene (17) is spoken.

Inferring the Place of Origin of Haplotypes or Haplogroups. Haplogroup centroids with their standard deviations are given in Tables S2 and S3 and their location in Fig. S4*b*. Standard deviations measure the spread of each haplogroup as average distance from the mean and are calculated on straight-line distances from the means.

Several geographic distributions have multiple modes. These may be due to geographic barriers and/or secondary migrations or expansions, but also to later arrivals of other haplogroups that replaced/displaced earlier settlements in the regions between 2 widely separated modes. This may have happened specifically, for instance for haplogroup A, which probably at the beginning occupied continuously the Ethiopia–South Africa region, but in the first millennium after Christ much of it was subsequently replaced because of occupation by various waves of expanding Bantu farmers from Cameroon across all southern Africa. Another phenomenon is the presence of small propagules far away from the rest of the distribution, as in haplogroups B, G, I, J, L, H, and Q. In part this phenomenon is due to the small size of many samples. Sometimes there is complete geographic separation between areas showing fairly important presence of the haplogroup, as in G, H, Q, and T. For example, haplogroup H in Europe is often confined to ethnic Roma. Another phenomenon possibly responsible for some of these irregularities, especially in the early part of the expansion, is the surfing effect described later.

If migrations were random, the geographic distribution of individuals with a specific haplogroup would be approximately normal (Gaussian) around the place of origin of the oldest mutation defining the haplogroup, apart from irregularities due to vagaries of the environment: obstacles, like mountains and deserts, or favored routes, like coasts and rivers. Below we list the most prominent 16 haplogroups by the number of major modes (i.e., the local geographic maxima of the gene frequencies of each haplogroup), which are easily observable on the maps as the darkest blue areas (Fig. 2): haplogroups B, D, G, J, L, M, and O are largely unimodal geographically, apart from isolated propagules (e.g., in B). The presence of haplogroup O in remote Madagascar is consistent with a well-known part of the Polynesian expansion. Most of these haplogroups have a clear, major mode and a relatively symmetric geographic distribution around it. Some like G, L, S, M, and T do not reach very high frequencies and/or have a very limited geographic distribution. Haplogroups A, E, and H are sharply bimodal, whereas C, I, N, Q, R, and T are clearly multimodal and several of them have a wide and abundant world representation. By subdividing each haplogroup into a tree showing the geographic distributions of its subclades, one can distinguish within-haplogroup paths of expansions and their mutational hierarchy. Examples are shown in Figs. S1–S3, S4*a*, and S5 for the most interesting cases of complicated distributions: A, B, E, I, J, O, and R. Even for

superficially unimodal haplogroups like J and R this analysis reveals interesting complexity. The similarity of patterns of different mutants indicates some secondary expansions. It is also interesting to sum the distributions of different haplogroups descending from the same mutation, as for example D and E, which both descend from DE-YAP, the first mutation that split into the E branch that perhaps returned to Africa (or arose there), whereas the other branch, D, is found today mainly in the Himalayas and Japan.

For unimodal haplogroups, centroids are not far from the modes of the geographic distribution of the population represented by the haplogroup. Their approximately spatial (bidimensional) Gaussian distribution suggests that the populations carrying the haplogroup may have expanded relatively slowly and relatively homogenously in all directions around their place of origin. This is more likely to happen if the local environment has no major barriers or irregularities. Under these assumptions, the centroid or the mode or some intermediate position are reasonable estimates of the place of origin of the earliest mutation(s) marking the haplogroup. Such a geographic assignment of the place of origin of the mutation leading each haplogroup allows superimposing a phylogenetic tree on a geographic map of the origin of each haplogroup, producing an approximate space and time display of major migrations of the human species in the past.

The Surfing Effect. Fisher (18) studied the rate of expansion of an advantageous mutant and showed that a "wave of advance" forms and proceeds at a constant rate in which its spread pattern is dictated by migration rate and the degree of selective advantage. The same formula is valid also for the expansion of a demographically growing population that enters new territory under a constant rate of migration, and a constant population density at saturation during the whole advance: thus, populations will move away from their origin by a "wave of advance" that soon reaches a constant shape from beginning to end of the expansion.

Mutants that arise in the wave front of an expanding population have an interesting advantage over mutants arising behind the expansion front, in the fully saturated portions of the expansion. This advantage is due to random genetic drift, because, as is well known, the probability of success (final fixation) by drift of a just-arisen mutant is equal to $1/N$, where N is the population size. Therefore a mutant arisen within the expansion front finds itself in a position of advantage over mutants arisen back of the expansion front, where populations have reached saturation levels and therefore higher Ns. This is due to the fact that these mutants find themselves for some time in a locally small population, which is smaller the closer the new mutation is located to the extreme line of the advancing front. The smaller the local population in the immediate neighborhood of the mutant's origin within the advancing front, the higher will be the relative local frequency of the mutant, and therefore the mutant's probability of final success by drift alone is greater. The result is that they have a greater chance of final success, even if they are not favored by natural selection (i.e., are selectively neutral). The advantage increases the closer the just-born mutant is to the extreme of the advancing front, where the population density is thinnest. The faster the population expansion, the greater the probability of success of a mutant that arises in the wave front, because then the wave front is longer.

The existence of this phenomenon was detected by simulations, which showed that it affects in a characteristic way the shape of the geographic distribution of selectively neutral mutants that originate in the wave front. It has been called the surfing effect because it superficially resembles the fact that a swimmer surfing on the wave crest travels with the speed of the wave (19, 20). Observations on the geographic distribution of several Y chromosome mutants show that some of the patterns strongly suggest that the surfing effect is at play. The ocean coasts and other barriers will stop expansions, and several of the observed geographic distributions show specific increases of the frequency of the mutant in the direction of migration that can reach highest values up to 100%, at or before barriers most distant from their likely place of origin. Such examples include (*i*) the haplogroup A-M6 mutant that reaches a very high frequency near the Atlantic Ocean between Angola and Namibia; (*ii*) the E-PN2 variant that probably arose near the Red Sea and has its highest frequencies near the Atlantic coast of West Africa; (*iii*) I-M170 on the Atlantic coast of Scandinavia; (*iv*) J1-M267 in southern Arabia; (*v*) O-M122 in China; (*vi*) N near the Arctic; (*vii*) Q in South America; and (*viii*) R in western Europe.

One consequence of the surfing effect is that for mutants showing such geographic patterns, the place of origin of the mutants is neither at the spatial mode of the mutant frequency nor at the spatial average frequency, which seem reasonable for most other mutants. This is also intuitively suggested by the observed spatial distributions of the frequencies of the mutants just listed, which are highly asymmetric: the shape of the distribution is such that the mode would often be extrapolated on the coast line or close to it, whereas the arithmetic average is inside land but close to the coast.

A deterministic approach (21) would place the origin of the mutant midway between the average frequency of the mutant and the place of origin of the expansion, assuming it proceeds at a constant rate. But drift tends to shift the place of origin of a mutant even closer to the origin of the expansion, by 10% in the simulations shown in ref. 19. Other sources of uncertainty are due to lack of knowledge of the exact routes followed by the expansion, considering the vagaries of the Earth's surface; the estimates used in our graphic presentations of expansion paths are tentative. Another source of uncertainty is the possibility that the evolutionary success of some of the mutants presented above is due at least in part to natural selection for the mutants. The absence of recombination within most of the Y chromosome (and especially for the haplogroups that we study) means that the particular mutant used to label the haplogroups is not necessarily responsible for the natural selection of any other mutant(s) that appeared after the one used for labeling the haplogroup.

One might surmise that the geographic distributions observed and imputed to surfing are entirely due to natural selection. Although we have not made specific computations, which would demand extensive simulations, it seems likely that the high gradients of gene frequencies observed would request unusually high selection coefficients. One cannot exclude that mutations affecting male fertility (or perhaps less probably, survival) might be involved and determine relatively high selection, but they seem unlikely. One indication in favor of surfing is that most of the examples of potential surfing mentioned above are in regions where gradients of genetic diversity have already been noted. These are indications of important partial expansion routes during the Out of Africa one, or secondary ones, connected with agro-pastoral transitions or other less well known major movements. Fast expansions favor surfing.

Genetic Diversities Among Populations for Autosomes, X Chromosome, and Y Chromosome, and the 1:4/3:4 Expected Ratio of Diversities. As mentioned, the estimates of relative genetic diversity for the average autosome (henceforth called A), for the X chromosome, and for the Y chromosome under drift alone are expected to vary inversely to the number of chromosomes per couple of parents, which are 1 for the Y chromosome, 3 for the X chromosome, and 4 for autosomes. An earlier article (11) showed that the variations among autosomes A and X chromosome are in a ratio very close to the expected: 4/3 (= 1.33). Another article (22)* has partially confirmed the rule, with the exception that West Africans showed a slightly but significantly greater ratio than Europeans or Asians.

There are several reasons that can explain the difference between

*The paper of Keinan A, et al. (22), very relevant, shows, with a greater number of individuals and populations than ours, that the X chromosome/autosome ratio is slightly aberrant, suggesting special evolutionary events at the founding of the non-African populations.

Table 1. Variation among populations by *F*st

Markers	Observed	Expected
Autosomes, average of 22 chromosomes	A = 11.7% ± 0.11%	
X chromosome	15.6% ± 0.53%	1.33 A = 15.6% ± 0.14%
Y chromosome	38.9% ± 2.5%	4 A = 46.8% ± 0.44%

these 2 articles. The first article (11) examined 948 individuals from 51 world populations from all 5 continents, whose DNA is available at the Human Genome Diversity Project (HGDP)–Foundation Jean Dausset collection (23), for 650,000 SNPs (1 of the 52 original HGDP populations was excluded because the number of individuals is too small). The second article (22) examined 3 populations: 1 from West Africa (Nigeria), 1 from northern Europe, and 1 made of equal numbers of Chinese and Japanese for a total of 240 individuals, and for a variable number of SNPs. Because Li et al. (11) examined many populations from all continents with a larger number of SNPs, and the average data show no significant deviation from the expected 4/3 ratio, it is likely that the anomaly observed by Keinan et al. (22) has a simple explanation, which they themselves suggest: one African population they examined is slightly biased. The Li et al. (11) data included 6 African populations certainly different from theirs, and the anomaly was absent. A likely explanation is that the southern half of West Africa used by Keinan et al. underwent a recent agricultural expansion, some 4,000–5,000 years ago, whereby a new, strong founder effect probably altered considerably the local genetic diversity (18, 23).

One of our interests in collecting Y chromosome data was to test their usefulness in evolutionary studies. An analysis of early data indicated that the Y chromosome showed considerable world variation (24). Many later data accumulated since that early beginning gave stronger proof of the correctness of this hypothesis (25). The published data we collected for this article allowed us to calculate an Fst among populations (Table 1), which we compare with autosome and X chromosome estimates by Li et al. (11).

The expectations for X and Y are calculated with respect to the autosome value A, which being an average of all of the autosomes whose individual Fsts vary very little, has a much smaller statistical error. The X chromosome expectation (1.33 A) is practically identical to the observed value, and that of the expected Y chromosome value is significantly higher (*P ca.* 0.001) than the observed one, but the difference in value is only 17%. There is no indication that Y chromosome SNPs used for population analysis are subject to natural selection. It is practically certain that many autosomal or X genes do show some natural selection differences among populations. Taken at face value, these data suggest that there is somewhat more natural selection on average in A+X than in the Y chromosome, but the difference from the drift expectation based on the 1:4/3:4 rule the difference for the *F*st value is modest.

Slopes of the Serial Founder Effect in Different Situations and with Different Markers. Fst calculates the genetic difference between 2 or more populations. The total diversity among individuals is usually partitioned into 2 variances, *among* populations and *within* populations. A simple approach to studying the within-population fraction is to calculate as genetic variance among individuals forming a given population, the "average heterozygosity," which is the frequency of heterozygotes for each gene averaged over all genes. This is obtained using the average heterozygote frequency for every gene in each population expected under random mating. Although this quantity has full genetic meaning only in autosomal genes, it can also be calculated from the frequencies of haplogroups of the Y chromosome and has the same meaning for studying genetic diversity within populations. Between-populations variance g regularly increases proportionately to their geographic distance from the population of origin. Conversely diversity within populations decreases regularly because it is $1\text{-}g$ in relative values. In fact, the within-population variance does fall linearly with the distance from the population of origin, an observed phenomenon called serial founder effect (10) that is explained almost entirely by drift and migration. In Table 2 we compare data available for the slope of the decrease of genetic diversity within populations for autosomes, X, and Y chromosome for 3 different types of markers (SNPs, microsatellites, or haplogroups), as a function of the geographic distance of the population from that of origin of modern humans.

Although there is no general theory about the expected slope of g under the serial founder effect, it seams reasonable to assume that also for different chromosomes the slope of the within-population variance will vary proportionately to the expectations under drift, which are given for A:X:Y as 1:4/3:4. But there may be other factors affecting the slope that differ according to marker type or for other reasons. Simulations conducted for microsatellites and SNPs (11, 25) indicate that the serial founder effect slope depends on mutation rates, decreasing in absolute value for increases in certain mutation rates. Because no satisfactory theory of the dependence of the slope on the mutation with different types of markers is available, accurate comparisons can be made only for marker types that are subject to the same mutation rates.

Table 2 shows that present data do not afford making comparisons using the same type of markers, but at least qualitatively some expectations of the simple drift model are satisfied. In particular the highest value of the slope is obtained with the Y chromosome (25.7), and it is near the expectation of 4 times the autosome value, but only when it is compared with the microsatellite value (25.7/6.5 = 3.9). The comparison between the slopes obtained for A with 783 microsatellites (6.5) and 650,000 SNPs (3.8) shows that microsatellites have significantly flatter slope than SNPs, as expected because of their much higher mutation rate. The only comparison among different types of chromosomes using the same type of SNPs

Table 2. Slopes of the relative linear decrease of genetic variation within populations, estimated from the average heterozygosity (or equivalent quantity for Y chromosome) of each population, as a function of its geographic distance from Addis Ababa (see also Fig. S6 and Table S4)

Chromosome	Markers	Source	Slope (absolute value)
All autosomes + X	783 microsatellites	Ref. 11	6.5×10^{-6}
All autosomes + X	650,000 SNPs	Ref. 12	3.8×10^{-6}
X chromosomes	Haplotypes	Ref. 12	18.8×10^{-6}
Y chromosome	NRY haplogroups	This article	25.7×10^{-6}

The decrease has been postulated to be due to drift caused by the serial founder effect (11). Geographic distance in km. All slopes are negative: their absolute values are given in the table.

is from Li et al. between X and A, but the 2 values differ significantly from the 4/3 ratio, X being much higher than A in absolute value. There is a good explanation for this disagreement: SNPs of A are a random choice, whereas the X data refer to selected haplotype frequencies and are likely to be closer to 50% than those of random SNPs. This will undoubtedly increase the expected absolute value for the *Xm* and may explain the discrepancy. In conclusion, it is worth pursuing further the approach of the serial founder effect slope. In general these results from the analysis of the serial founder effect are in general agreement with the idea that drift is responsible for a very high proportion of the observed genetic variation among human populations, but presently available data can offer only qualitative support for it. Further work aimed more directly to it can supply stronger evidence.

Conclusions

There is a substantial literature related to methods and conclusions of genetic variation and geography that discusses the merits and difficulties of the various genetic systems (20, 26–31).

DNA variation among human populations seems to have less adaptive significance than drift, reinforcing earlier results (10, 11). On the other hand, natural selection has favored an enormous advance to our species in competition with other species, allowing a sheer numbers increase by a factor of $\approx$1,000 (32) between approximately 60,000 and 10,000 years ago, because of expansion of a small African tribe to the whole world, and another 1,000$\times$ factor because of the transition from food collection to food production initiated probably independently in several parts of the world, beginning 12,000 to 6,000 years ago, which involved many new challenges probably very different from place to place. There is wide agreement that the first phase was due to the generalization of linguistic skills, favoring sociological developments and, with it, very rapid spread of innovations, and another type of cultural evolution that became even stronger, although more localized, with the exploitation of local domesticated plants and animals. One would expect little increase of genetic variation between populations in the first phase, except for some adaptation to different climates (which was, however, often met by cultural adaptations like housing and clothing). There probably was more need of biologic adaptations, and therefore of natural selection in the second phase, to meet challenges due to profound differences in diet and increased contact with domesticated animals and their contagious diseases. Examples in different parts of the world are pigmentation (33), lactose tolerance (34–36), and increase of genetic resistances to malaria (37, 38). Cultural evolution became much more effective than biologic evolution in meeting perceived needs, and may have almost replaced it, but every novelty has costs in addition to benefits, and thus much of natural selection may now be directed to take care of specific costs generated by cultural evolution, which were, however, not too severe, at least so far.

Materials and Methods

Centroid Estimation. The center of gravity and the weighted average standard deviation from the center of gravity are calculated as described by Forster et al. (39).

Geographic Maps. Haplogroup frequency data from the literature were normalized to common levels of haplogroup resolution. The geographic distributions of the different haplogroup frequencies across the world were created using the Kringing method in MapViewer (version 7, Golden Software).

ACKNOWLEDGMENTS. We thank Gianna Zei and Antonella Lisa for providing Fig. S6 *a* and *b*; Chris Gignoux for technical advice regarding spatial frequency mapping and centroid computation and alerting us to SNPs appearing to unify the IJK and NMOP haplogroups, whose relationships were confirmed by Alice A. Lin; Roy J. King for his comments regarding the phylogeny; and Alice A. Lin, Cheryl-Emiliane T. Chow, Gianpiero L. Cavalleri, and Cengiz Cinnioglu for technical assistance regarding the binary data for the Centre d'Étude du Polymorphisme Humain–HGDP samples. This investigation was supported by a research grant from the Sorenson Molecular Genealogy Foundation.

1. Hagedoorn AL, Hagedoon AC (1921) *The Relative Value of the Processes Causing Evolution* (Martinus Nijho, The Hague).
2. Fisher RA (1922) On the dominance ratio. *Proc R Soc Edinburgh* 42:321–341.
3. Wright S (1969) *Evolution and the Genetics of Populations. Volume 2: The Theory of Gene Frequencies* (University of Chicago Press, London)
4. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217:624–626.
5. Le Cam LM, Neyman J, eds (1966) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health Conference* (University of California Press, Berkeley).
6. Cavalli-Sforza LL, Moroni A, Zei G (2004) *Consanguinity, Inbreeding and Genetic Drift in Italy* (Princeton University Press, Princeton).
7. Cavalli-Sforza LL (1967) Human populations. *Heritage from Mendel*, ed Brink A (University of Wisconsin Press, Madison, WI), pp 309–331.
8. Cavalli-Sforza LL (1969) Genetic drift in an Italian population. *Sci Am* 221:30–37.
9. Prugnolle F, Manica A, Balloux F (2005) Geography predicts neutral genetic diversity of human populations. *Curr Biol* 8:159–160.
10. Ramachandran S, et al. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci USA* 1:15942–15947.
11. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 22:1100–1104.
12. Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y chromosomal binary haplogroups. *Genome Res* 12:339–348.
13. Karafet TM, et al. (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 18:830–838.
14. Goebel T (2007) The missing years for modern humans. *Science* 315:194–196.
15. Stringer C (2002) Modern human origins: Progress and prospects. *Philos Trans R Soc Lond B Biol Sci* 29:563–579.
16. Deshpande O, Batzoglou S, Feldman MW, Cavalli-Sforza LL (2009) A serial founder effect model for human settlement out of Africa, *Proc Biol Sci* 276:291–300.
17. Greenberg JH (1987) *Language in the Americas* (Stanford University Press, Palo Alto, CA)
18. Fisher RA (1937) The wave of advance of advantageous genes. *Ann Eugenics* 7:355–369.
19. Edmonds CA, Lillie AS, Cavalli-Sforza LL (2004) Mutations arising in the wave front of an expanding population. *Proc Natl Acad Sci USA* 101:975–979.
20. Klopfstein S, Currat M, Excoffier L (2006) The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol* 23:482–490.
21. Vlad MO, et al. (2002) Neutrality condition and response law for nonlinear reaction-diffusion equations, with application to population genetics. *Phys Rev E Stat Nonlin Soft Matter Phys* 65:6–14.
22. Keinan A, Mullikin JC, Patterson N, Reich D (2009) Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* 41:66–70.
23. Cann HM, et al. (2002) A human genome diversity cell panel. *Science* 296:261–262.
24. Kuper R, Kröpelin S (2006) Climate-controlled Holocene occupation in the Sahara: Motor of Africa's evolution. *Science* 11:803–807.
25. Ammerman AJ, Biagi P (2003) *The Widening Harvest (The Neolithic Transition in Europe)* (Archaeological Institute of America, Boston).
26. Weale ME, et al. (2003) Rare deep-rooting Y chromosome lineages in human: Lessons for phylogeography. *Genetics* 165:229–234.
27. Garrigan D, Hammer MF (2006) reconstructing human origins in the genomic era. *Nat Rev Genet* 7:669–680.
28. Mitchell-Olds T, Willis JH, Goldstein DB (2007) Which evolutionary process influence natural genetic variation for phenotypic traits? *Nat Rev Genet* 8:845–856.
29. Coop G, et al. (2009) The role of geography in human adaptation. *PLoS Genet* 5:e1000500.
30. Diamond J (2002) Evolution, consequences and future of plant and animal domestication. *Nature* 418:700–707.
31. Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848.
32. Cavalli-Sforza LL, Feldman MW (2003) The application of molecular genetic approaches to the study of human evolution. *Nat Genet Suppl* 33:266–275.
33. Mc Evoy B, Beleza S, Shriver MD (2006) The genetic architecture of normal variation in human pigmentation: An evolutionary perspective and model. *Hum Mol Genet* 5:176–181.
34. Enattah NS, et al. (2008) Independent introduction of two lactase-persistence alleles into human populations reflects different history of adaptation to milk culture. *Am J Hum Genet* 82:57–72.
35. Gerbault P, Moret C, Currat M, Sanchez-Mazas A (2009) Impact of selection and demography on the diffusion of lactase persistence. *PLoS One* 4:e6369.
36. Enattah NS, et al. Evidence of still-ongoing convergence evolution of the lactase persistence T-13910 alleles in humans. *Am J Hum Genet* 81:615–625.
37. Rosenberg R (2007) *Plasmodium vivax* in Africa: Hidden in plain sight? *Trends Parasitol* 23:193–196.
38. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *History and Geography of Human Genes* (Princeton University Press, Princeton).
39. Forster P, et al. (2002) Continental and subcontinental distributions of mtDNA control region types. *Int J Legal Med* 116:99–108.