The application of molecular genetic approaches to the study of human evolution

L. Luca Cavalli-Sforza[1] & Marcus W. Feldman[2]

doi:10.1038/ng1113

The past decade of advances in molecular genetic technology has heralded a new era for all evolutionary studies, but especially the science of human evolution. Data on various kinds of DNA variation in human populations have rapidly accumulated. There is increasing recognition of the importance of this variation for medicine and developmental biology and for understanding the history of our species. Haploid markers from mitochondrial DNA and the Y chromosome have proven invaluable for generating a standard model for evolution of modern humans. Conclusions from earlier research on protein polymorphisms have been generally supported by more sophisticated DNA analysis. Co-evolution of genes with language and some slowly evolving cultural traits, together with the genetic evolution of commensals and parasites that have accompanied modern humans in their expansion from Africa to the other continents, supports and supplements the standard model of genetic evolution. The advances in our understanding of the evolutionary history of humans attests to the advantages of multidisciplinary research.

Reconstructing human evolution requires both historical and statistical research. Although conclusions are not experimentally verifiable because the process cannot be repeated, various disciplines such as physical and social anthropology, archaeology, demography and linguistics provide complementary approaches to researching questions of human evolution. The existence of molecular genetic variation among human populations was first demonstrated by Hirszfeld and Hirszfeld[1] in a classic study published in 1919 of the first human gene to be described—*ABO*, which determines ABO blood groups. The subsequent identification of blood group protein markers, such as MNS and Rh expanded the repertoire of polymorphic markers that could be analyzed using antibodies. R.A. Fisher showed that evolution could be reconstructed by analyzing the multilocus genotypes on a chromosome observed in populations and their inheritance within families[2]. The term 'haplotype' for the multilocus combination of alleles on a chromosome was introduced by Ceppellini *et al.*[3] during early research on the major histocompatibility complex. Immunological methods remained the only satisfactory technique for detecting genetic variation until Pauling *et al.*[4] introduced electrophoresis to separate different mutants of hemoglobin, a technique that was rapidly adapted to analyze variation in other blood proteins .

It was soon obvious that genetic variation was not rare but, on the contrary, that almost every protein had genetic variants[5,6]. These variants became useful markers for population studies. The first book of allele frequencies in populations, published in 1954, was limited almost completely to serological variation[7], and books listing genetic variation increased rapidly in size and number[8–10]. In 1980, a method for studying variation in DNA[11] identified mutants of restriction sites by using radioisotopes and generated several new markers. But it was only with the development of PCR in 1986 that the study of more general DNA variation became possible. The development of automated DNA sequencing in the early 1990s paved the way for the application of systematic study of genome variation to human evolutionary biology.

Data from protein markers (sometimes called 'classical' markers) are still more abundant than are data from DNA, although this situation is rapidly changing. For example, Rosenberg *et al.*[12] studied 377 autosomal microsatellite polymorphisms in 1,065 individuals from 52 populations producing a total of 4,199 different alleles, about half of which were found in all principal continental regions. Another study[13] of 3,899 single-nucleotide polymorphisms (SNPs) in 313 genes sampled in 82 Americans self-identified as African American, Asian, European or Hispanic Latino found that only 21% of the sites were polymorphic in all four groups—a fraction that would be expected to increase with more sampled individuals. It is interesting to note, however, that so far no conclusions derived from the earlier studies of classical polymorphisms[14] have been found to be in disagreement with those obtained with DNA markers. Nonetheless, molecular genetic markers have provided previously unavailable resolution into questions of human evolution, migration and the historical

[1]*Department of Genetics, Stanford Medical School, Stanford University, Stanford, California 94305-5120, USA.* [2] *Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA. Correspondence should be addressed to M.W.F. (e-mail: marc@charles.stanford.edu).*

relationship of separated human populations. In this review we discuss the evolutionary and historical forces that have shaped genomic variation and how its interpretation has led to a deeper understanding of the evolution of our species.

Evolutionary events affecting genomic variation

All genetic variation is caused by mutations, of which there are many different types. The most common and most useful for many purposes are SNPs, which can be detected by DNA sequencing and other recently developed methods, such as denaturing high performance liquid chromatography[15], mass spectrometry[16] and array-based resequencing[17].

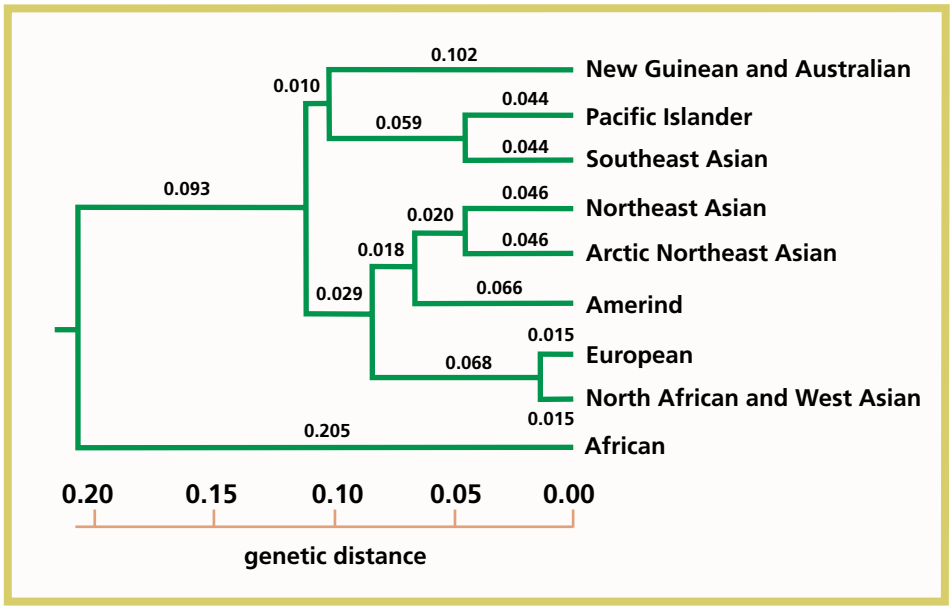
Allelic frequencies change in populations owing to two factors: natural selection, which is the result of population variation among individual genotypes in their probabilities of survival and/or reproduction, and random genetic drift, which is due to a finite number of individuals participating in the formation of the next generation. Both natural selection and genetic drift can ultimately lead to the elimination or fixation of a particular allele. In the presence of mutation and in the absence of selection (that is, under neutral conditions), the rate of neutral evolution of a finite population is equal to the reciprocal of the mutation rate[18].

The earliest evidence of selection acting on a human gene was the discovery that heterozygotes of the hemoglobin *A/S* polymorphism have greater resistance to malaria than do *AA* or *SS* homozygotes. In malarial environments, this results in a balanced polymorphism that maintains the *S* allele even though *SS* individuals are severely ill with sickle-cell anemia. Recent studies of DNA variation have focused on detecting signatures of selection, either balancing or directional[19]. This has produced many different statistical tests using DNA diversity[20,21] and comparisons of nucleotide substitutions that do or do not affect the amino acid sequence of proteins[22,23].

Strong molecular evidence of balancing selection, also in malarial environments, has been found for the *G6PD* locus, the low-activity alleles of which seem to confer resistance to malaria[24,25]. Other analyses[26] have found evidence for positive selection at both *G6PD* and another gene *TNFSF5*, which is also implicated in the response to infectious agents. Strong directional selection has also been proposed[27] for *FOXP2*, which shows a two amino-acid difference between the human protein and the monomorphic form in primates. It has been suggested that these changes may have been selectively important for the evolution of speech and language in modern humans[27]. In other genes, however, the agent of selection is not at all obvious; for example, the *CCR5* gene[28] seems to be related to HIV resistance, and mutations in the *BRCA1* gene[29] produce an increased risk of female breast cancer. In such cases it is often very difficult to disentangle the effects of population dynamics or structure from selective pressures. These complications can be clearly observed in a thorough analysis of the *HFE* locus[30], mutations of which result in hemochromatosis. In this study no evidence of selection on single SNPs or on haplotypes was detected, but significant between-continent variation was found. Unlike other studies[12,13], African samples showed only slightly more rare SNPs than Europeans or Asians. This suggests the possibility that different evolutionary models are relevant to the different continents.

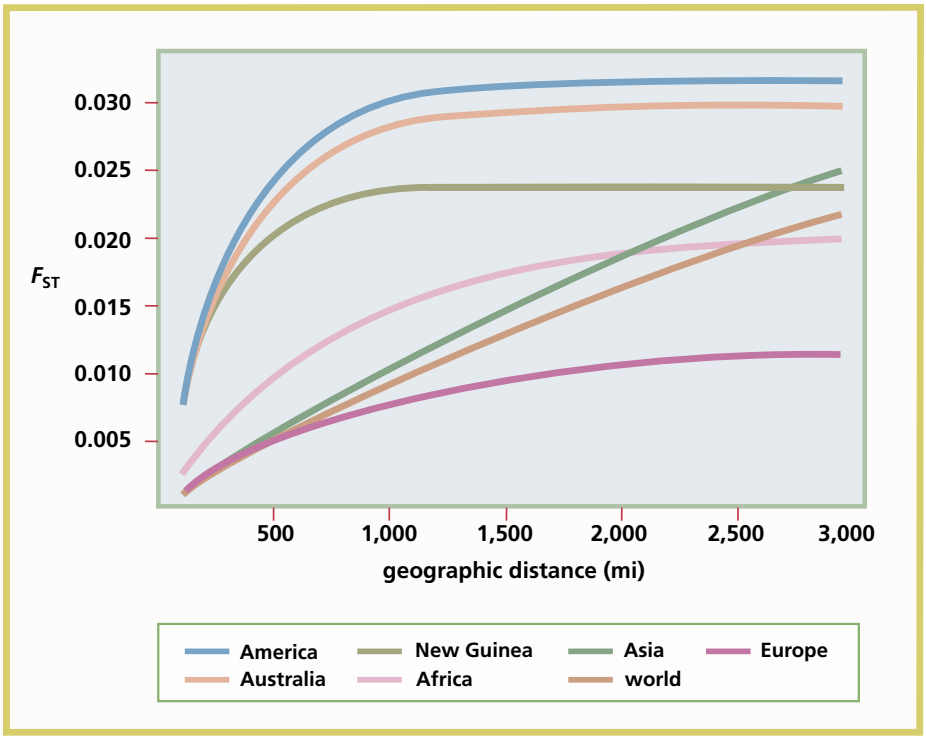
Genetic statistics of the substructure underlying human populations may also suggest which genes are candidates to have been under selection. The idea, originally proposed by Cavalli-Sforza[31] and expanded by Lewontin and Krakauer[32], is to compare the expected and observed values of F_{ST} statistics (a measure of the relevant amount of genetic diversity among populations)[33] for a large enough number of genes and focus on those loci that produce extreme values. In a recent study of 8,862 SNPs mapped to gene-associated regions[34], 156 genes for which the F_{ST} value was exceptionally high and 18 for which it was exceptionally low were identified, suggesting that these 174 genes are candidates for having been under selection. Similar approaches have been applied to specific genes such as *G6PD*[24], the Duffy blood group locus[35], lactase haplotypes[36], *MAOA*[37] and skin pigmentation[38]; in each case, unusually high variation among populations has been invoked as a signature for the action of selection. The interactions among population substructure, demography and phenotypic variation are discussed in a recent review[39].

Migration is another important factor in human evolution that can profoundly affect genomic variation within a population. Most populations are relatively isolated, however, although rare exchange of marriage partners between groups does occur. An average of one immigrant per generation in a population is sufficient to keep drift partially in check and to avoid complete fixation of alleles. Sometimes a whole population (or a fraction of it) migrates and settles elsewhere. If the migrant group is initially small but subsequently expands, by chance alone the frequencies of alleles among the founders of the new population will differ from those of the original population and even more so from those among which it settles. In this situation, group migration has an effect that in some respects is opposite to that of



BOB CRIMI

Fig. 1 Summary tree of world populations. Phylogenetic tree based on polymorphisms of 120 protein genes in 1,915 populations grouped by continental sub-areas and F_{st} genetic distances[14]. Root placed assuming a constant rate of evolution.



BOB CRIMI

Fig. 2 Relationship between genetic and geographic distance. Genetic distance of population pairs measured by F_{st} as a function of geographic distance between members of the pairs[14]. Only samples from indigenous people were included. Continents where primitive economies predominate (hunting-gathering or tropical gardening) show highest asymptotes. Asia and the world do not asymptote within the range shown.

individual migration among neighboring populations: it creates more chances for drift and therefore divergence[40]. The effect will be intergroup variation in allele frequencies.

Genome structure and population history

A complete description of human genetic variation requires more than just properties of isolated genes, microsatellites or SNPs. How these vary simultaneously within a part or whole chromosome requires statistics of correlation between the variation at different positions, and these are usually described by patterns of linkage disequilibrium (LD). The stronger the LD, the more likely that alleles at each of two positions will be found in association with one another. Using studies of protein variants in the 1970s and 1980s it was rare to identify strong LD in populations of outcrossing diploids. However, as more details become available on variation in human DNA across populations, LD between polymorphic DNA sites is increasingly being detected.

Standard population genetic theory suggests that LD between pairs of genetic markers should decrease as the recombination between them increases. But early studies of short segments of DNA did not show this relationship for SNPs[41]. Because the pattern of LD is expected to vary both with local effects, such as the extent of selection, the degree to which pairs of sites interact in response to selection (epistasis), and with population-scale forces, such as drift (reviewed in ref. 42), migration and non-random mating[43], genomic patterns of LD can be expected to be fairly complex. Recent studies of relatively long (200–500 kb) stretches of DNA, however, have produced a picture of blocks of high LD interspersed by short intervals of low LD. Within the blocks of high LD there is evidence of lack of recombination, whereas the regions between the blocks seem to be 'hot spots' in which recombination occurs frequently[44–47]. It has been therefore suggested that the next phase of research into human variation should focus on these blocks of high LD, for which haplotypes, rather than single markers, will become the unit of variation[44]. Although it has been known for many years that the extent of LD among specific sets of genes shows great variation around the world[48]—for example, it is usually much weaker in African than in European populations[49]—genome-wide studies covering representative worldwide populations remain to be done[50,51].

Interpreting evolutionary history

The history of population differentiations using genetic data was initially inferred from phylogenetic trees[52–54] and from multivariate statistical methods such as principal components[53,55] (of which multidimensional scaling is a derivative) that use allele frequencies. Population trees are especially useful for reconstructing history if population differences can be assumed to result from fissions that occur randomly in time, with a constant rate of neutral evolution in each population between fissions. This is likely to be roughly true for data on several autosomal genes from large populations that are geographically and genetically distant, as illustrated in Figure 1, which shows nine such groups from around the world. Completely different types of DNA variation provide the same basic conclusion regarding the relationships between these populations (refs. 56, 57; and L.A. Zhivotovsky, N.A. Rosenberg and M.W. Feldman, manuscript in preparation).

Violation of the above assumptions, such as the presence of migration or selection, affects the interpretation of population trees. However, when migration between geographic neighbors is frequent, principal components displayed in two dimensions reflect the geographical distribution of populations. Under the simple evolutionary model described above, trees and

Table 1 • Variation components within and between populations

Polymorphism	Loci (*n*)	Group (*n*)	Average variance components (%) Within samples	Among samples within group	Among groups	References
protein	17–25	3–7	83.8–87	2.8–8.3	6.3–11.2	60, 61,141,142
DNA and protein	109	4–5	84.4	4.7	10.8	62
DNA	377	5–7	93.2–94.1	2.4–2.5	3.6–4.3	12

principal components give similar results[58]. For populations that are geographically close, genetic and geographic distances are often highly correlated (Fig 2), with an asymptote for the genetic distance at about 1,000–2,600 miles on average (but higher for Asia and the world, which are not at equilibrium). Recent statistical developments in detecting clustering among populations based on highly polymorphic autosomal markers[59] have been valuable for analyzing very large population genetic data sets[12]. It is important that this completely different approach produces the same primary continental clusters as the earlier methods. In its application to data sets with numerous polymorphic loci, however, it does seem to be more sensitive in detecting and assessing individual ancestry.

Early studies showed that genetic differences between populations are relatively small as compared with those within populations[60,61]. Subsequent analyses, including molecular polymorphisms of 14 populations representing all continents, confirmed that the within-population variance was about 85% of the total (Table 1)[62]. A recent analysis of 377 autosomal microsatellite markers[12] in 1,065 individuals from 52 worldwide populations found that only 5–7% of the variation was between populations. It is the remaining 5–15%—the between-population component—that can be used to reconstruct the evolutionary history of populations.

Dating the origin of our species using genetic data

Archeological evidence is generally considered to support the initial spread of humans within Africa from an East African origin during the first half of the last 100 kya and the spread from the same origin to all the world in the last 50–60 kya. Analyses of numerous classical markers under this assumption have estimated the dates of first occupation by anatomically modern humans of Asia, Europe and Oceania at 60–40 kya, in agreement with archeological and fossil data. Dates for the first occupation of America are estimated at 15–35 kya. Thus, genetically derived dates are consistent with evidence from physical anthropology, providing support for the use of population trees[63]. Below we discuss how recent analysis of DNA polymorphisms supports this timing of the earliest split between Africans and non-Africans.

Studies of variation in DNA became possible in the early 1980s (refs. 64, 65). Subsequent estimates for the emergence of modern humans from Africa using autosomal restriction fragment length polymorphisms were consistent with earlier estimates[56,66]. From the analysis of several mitochondrial DNA (mtDNA) polymorphisms, Cann *et al.*[67] derived two important conclusions: the first major separation in the evolutionary tree of modern humans was between Africans and non-Africans; and the time back to the most recent common ancestor (TMRCA) of modern human mtDNA was 190,000 years (however, with a large error). After early doubts about the statistical validity of these interpretations of the data[68], the order of magnitude was confirmed[69–71]. It is important to note that TMRCA is usually significantly earlier than the first archaeologically observable divergence among a set of populations[72,73]. Also, TMRCA does not necessarily coincide with the onset of population expansion. The 'mismatch' method[74] to analyze mtDNA, which analyzes the distribution of between sequence differences, gives estimates that are more compatible with the beginning of expansions inferred from archeology.

Because mitochondria are transmitted along only female lineages and mtDNA is genetically haploid, the effective size of a population of mtDNAs is a quarter of that of the corresponding autosomes. The mutation rate of the mitochondrial genome is about ten times higher than that of nuclear DNA[75], which provides an abundance of polymorphic sites, but creates difficulties in reconstructing genealogies owing to repeated and reverse mutations. Like the non-recombining part of the Y chromosome (NRY), there is no evidence for recombination in mtDNA although low-frequency rearrangements of somatic mtDNA have been observed in heart muscle[76].

The mutation rate of the NRY is comparable to that of nuclear DNA, which means that polymorphisms are more difficult to find but genealogies are easier to reconstruct. The greater length of DNA on the NRY (perhaps 30 million bases of euchromatic DNA) relative to mtDNA compensates in data analyses for its lower mutation rate. Even though the NRY behaves effectively as a single locus, which is usually insufficient for evolutionary analyses, it has provided results that are consistent across many studies and in agreement with many archeological findings. In fact, the NRY genealogy constructed from 167 mutations[77] has been replicated with a totally independent set of 114 mutations[75] and confirmed independently using mostly different population samples[78,79].

Statistical analysis of Y chromosome data have been carried out using coalescent theory devised by Kingman[80]. Coalescent-based techniques using numerical methods to study complex likelihood functions derived from Bayesian analyses were developed subsequently[81–82] and have facilitated estimation of key parameters in the Y chromosome genealogy (ref. 83; and H. Tang *et al.*, manuscript in preparation) under specific assumptions about demographic history. Tang *et al.*[75] have shown that important evolutionary properties of the Y chromosome TMRCA, which is close to 100 kya, can be derived under few demographic assumptions.

Two recent estimates of TMRCA from mtDNA have been made using different methods. From complete mtDNA sequences (excluding the D loop) in a sample of 53 individuals, 516 segregating sites were seen and a TMRCA was estimated at 171 ± 50 kya[70]. From a sample of 179 individuals with 971 SNPs, the TMRCA was estimated at 200–281 kya using a generation time of 25 years, and 160–225 kya using a generation time of 20 years[75]. Corresponding estimates for the NRY-based TMRCA are 60–130 kya and 72–156 kya, with generation times of 25 and 30 years, respectively[75].

It is important to stress that such estimates of TMRCAs do not imply that the human population contained only one woman at 230 kya (the time of the mtDNA-based TMRCA, assuming constant mutation rates) or only one man at 100 kya (the time of the NRY-based TMRCA). The only implication is that all human mitochondria existing today descend from that of a single woman living 230 kya, and all NRYs descend from that of a single man living 100 kya. In both cases, it is likely that there were many more human individuals alive at the TMRCA—whether they were of the same species as *Homo sapiens* is hard to determine, but descendants of other species are either absent or extremely rare.

Although the reconstructed genealogies of mtDNA and NRY are broadly similar, there are some notable differences, probably owing to social differences in migration customs. For example, patrilocal marriage has historically been more common than matrilocal[84], which can explain differences in mtDNA and Y chromosome data in a number of populations[85–93]. Demographic differences between the sexes, such as greater male than female mortality, the greater variance in reproductive success of males than females and possibly the greater frequency of polygyny than polyandry, may explain the discrepancy between the NRY and mtDNA dates. These factors reduce the effective number of males and may explain the more than twofold difference between the NRY-based and the mtDNA-based TMRCA. Another attractive alternative explanation is that mutation rates in mtDNA are very variable, and when this variation is taken into account TMRCA of mtDNA could become closer to that of NRY.

Estimates of TMRCAs from autosomal genes are higher than those from mtDNA or NRY. In theory, they should be higher by a factor of four and the estimates are in this direction, although the number of autosomal genes studied is small and estimates of TMRCAs vary considerably[94]. For analyses of autosomal and X chromosomes, recombination can complicate genealogies and make TMRCAs impossible to estimate. There is also the possibility of heterozygote advantage, which has the potential to increase estimates of TMRCA. Heterozygote advantage may be widespread throughout the human genome but has been very difficult to show unequivocally, and the only fully confirmed example is sickle cell anemia, for which very large samples were required. There is some optimism, however, that the development of techniques that can detect heterosis for some genes in yeast[95] may lead to greater success in other organisms, including humans.

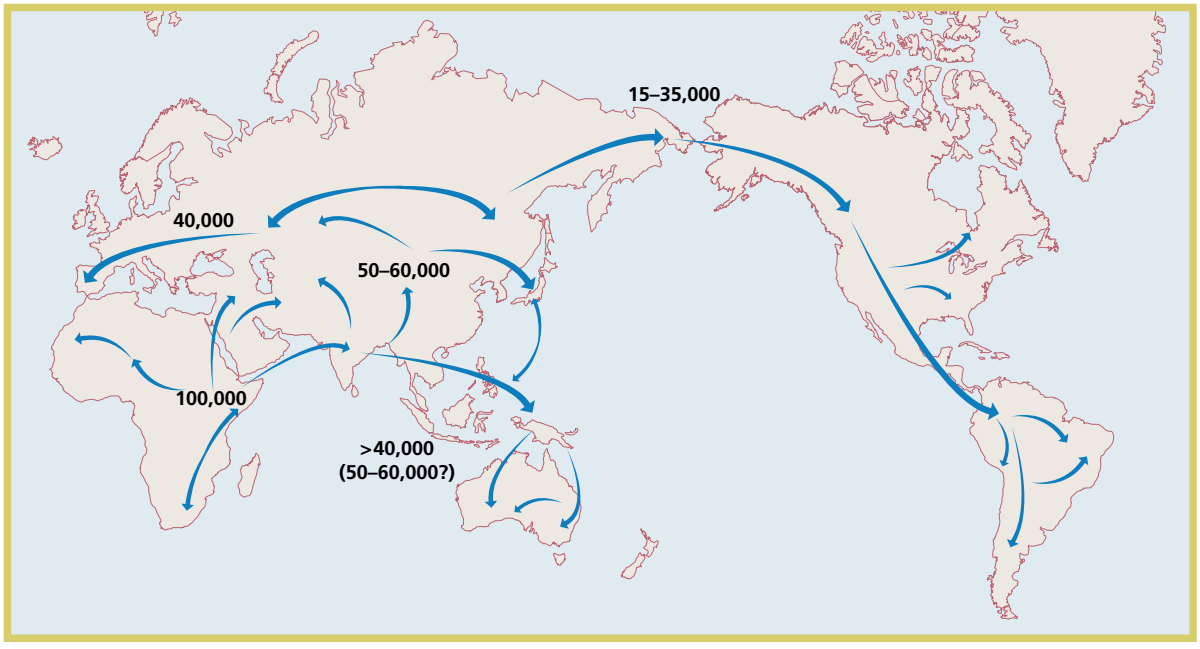
Tracking migrations of our species using DNA

A recent synthesis of Y chromosome phylogeography, paleoanthropological and paleoclimatological evidence suggests a possible hypothesis for the evolution of human diversity[96–98]. Around 100 kya or shortly after, a small population of about 1,000 individuals (that is, a tribe), most probably from East Africa, expanded throughout much of Africa. Then, between 60 and 40 kya there was a second expansion, most probably from a descendant population, into Asia and from there to the other continents (Fig. 3). This may be referred to as the 'standard model of modern human evolution'; it is also called 'out of Africa 2' in recognition of an earlier expansion of *Homo erectus* from Africa into Eurasia around 1.7 million years ago and assumes that anatomically modern humans (also called *Homo sapiens sapiens*) replaced earlier poorly known species of *Homo* that descended from the first migrants of *H. erectus*[98]. Genetic data provide some indication that the spread of humans into Asia occurred through two routes. The first was a southern route, perhaps along the coast to south and southeast Asia, from where it bifurcated north and south[99]. In the south, these modern humans reached Oceania between 60 and 40 kya, whereas the northern expansion later reached China, Japan and eventually America (this might represent the second migration to America, associated with the NaDene languages, postulated by Greenberg[100]). The second was a central route through the Middle East, Arabia or Persia to central Asia, from where migration occurred in all directions reaching Europe, east and northeast Asia about 40 kya, after which the first and principal migration to America suggested by Greenberg occurred not later than 15 kya[101].

It is still unresolved whether the divergence between these two expansion routes occurred in Africa or after entry into west Asia, and, if the latter, where it happened. Most literature accepts without discussion that the entry to Europe and central Asia was through the Levant. It is not at all certain that this was the only or the earliest route. These two initially divergent routes converged later, especially in the extreme East and America.

An alternative to the out of Africa 2 hypothesis, originated by Weidenreich[102] and expanded and called 'multiregional' by Wolpoff[103], maintains that all human populations living today originated in their various continents and evolved in parallel into modern humans. The main basis of this hypothesis is the claim that most ancient fossils (essentially those from Europe and Asia but not Oceania and America, where the human fossils found are all very recent and of modern human type) show a continuous morphological transition to modern humans. An extreme example of parallel evolution that included the doubling of brain volume is invoked to explain this scenario. In later versions of the multiregional model, parallelism is claimed to be the result of substantial intermigration[100,104].

Recent quantitative anthropological research on several human skulls has shown no morphological continuity in the various continents[85]. In addition, in the only part of the world where there existed a human type with some clear similarity to modern humans—namely Neandertals in Europe and west Asia—this purported ancestor of modern Europeans disappeared shortly after the appearance of modern humans (40–30 kya). MtDNA analysis of three Neandertals from Germany[105,106],



BOB CRIMI

Fig. 3 The migration of modern *Homo sapiens*. The scheme outlined above begins with a radiation from East Africa to the rest of Africa about 100 kya and is followed by an expansion from the same area to Asia, probably by two routes, southern and northern between 60 and 40 kya. Oceania, Europe and America were settled from Asia in that order.

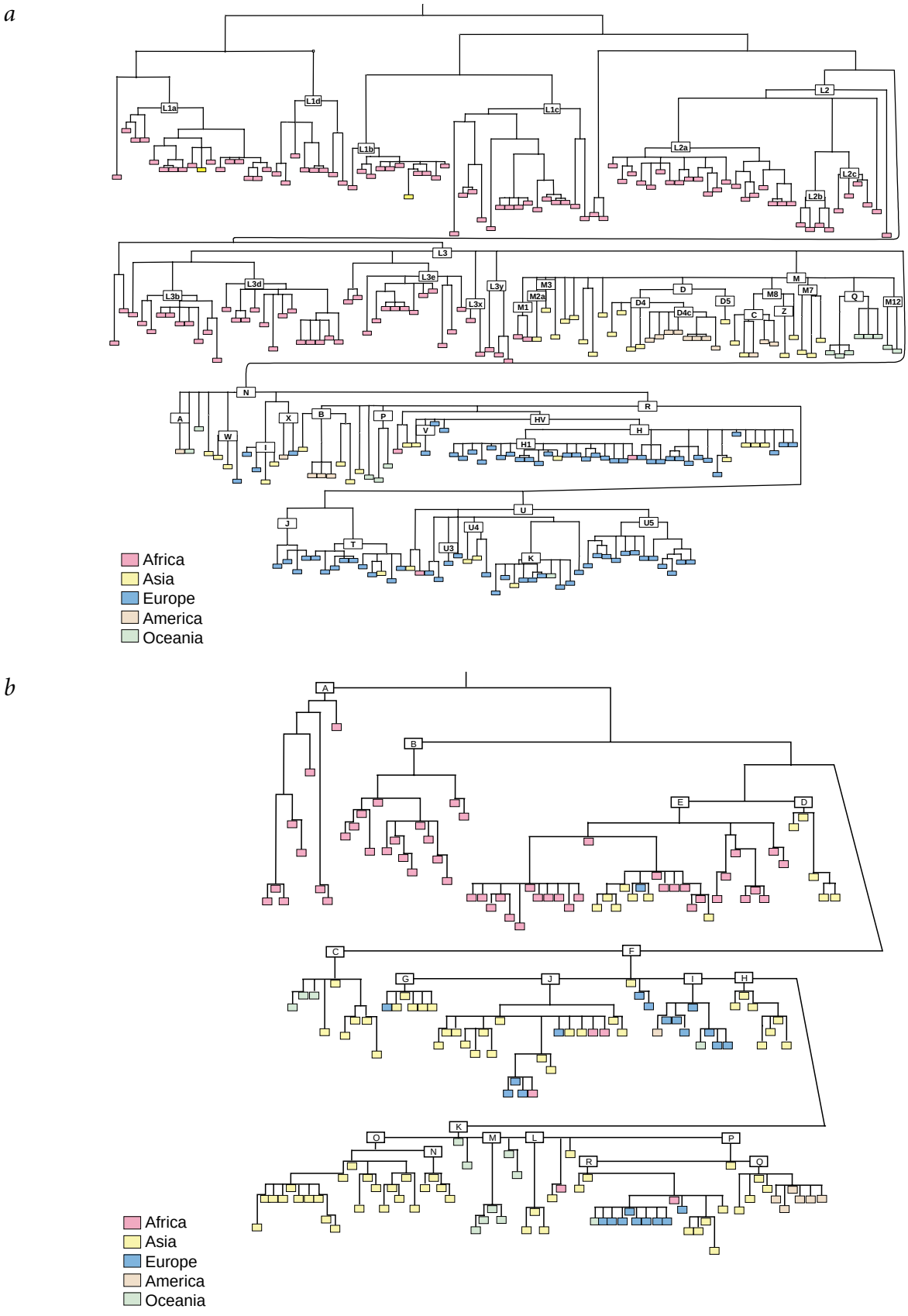


Fig. 4 High resolution molecular phylogeny to study human history. ***a***, Phylogeny of human mtDNA haplogroups and their continental affiliation composed from resequencing of 277 individuals[143,144]. The length of the branches corresponds approximately to the number of mutations. ***b***, Phylogeny of human Y chromosome haplogroups and the continental affiliation of their most frequent occurrence, composed from population genotyping of over 1,000 individuals and resequencing of over 100[145,146]. The length of the branches corresponds approximately to the number of mutations.

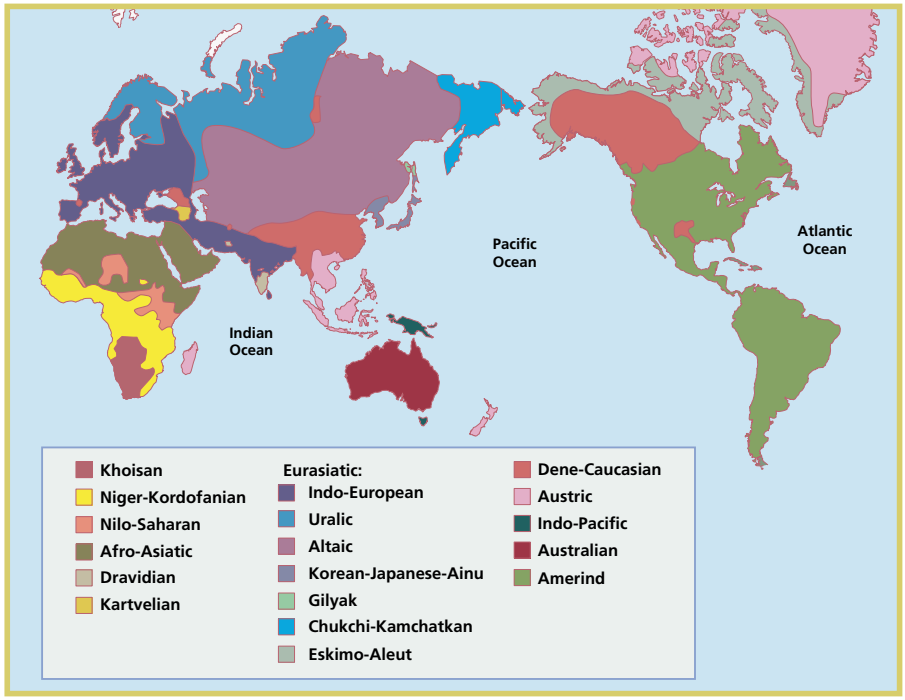
Croatia[107] and the Caucasus[108] detected no similarity with modern humans and indicated that the evolutionary separation of Neandertal from modern humans took place at least 500 kya.

It has been claimed that the age of TMRCA derived from the few human autosomal genes examined (between 500 and 1,000 kya) is proof of early expansions that have not been detected in NRY and mtDNA but are compatible with the multiregional hypothesis[109,110]. Templeton[100] proposes that this ancient TMRCA of autosomal genes is due to multiple migrations from Asia of *H. erectus* types before out of Africa 2 and the origin of modern humans. There is no evidence for such early migrations; even small populations tend to maintain high genetic variation (Table 1), and the amount of variation observed between human populations today is so small relative to the average variation within populations that it could have easily accumulated in the 100–200 ky before the present.

Recent simulation-based tests of the nested-clade method used by Templeton have found that it may produce an inference of long-term recurrent gene flow where this is specifically excluded from the simulation[111]. It is also important that the extent of LD in autosomal genes is much lower in African than non-African populations[48,49,112,113], suggesting that non-African populations represent a small genetic subset of the Africans. LD has had a long time to dissipate in Africa, and the polymorphisms of the autosomal genes from which their (expected) long TMRCAs are calculated are much more likely to have arisen in Africa than in Asia.

High resolution history using haploid markers

The identification in recent years of a large number of SNPs on the NRY and mtDNA has afforded higher resolution of population history through the reconstruction of the phylogenetic relationships of extant Y chromosomes and mtDNA (Fig. 4). Using the nomenclature developed by the Y Chromosome Consortium[114], the first two haplogroups (Fig. 4*a*; A and B) are almost completely African and even today represent mostly hunter-gatherers or their descendants, who have never reached high population densities or undergone high rates of increase. Slow growth is indicated by the accumulation of many mutations within a branch, as in most descendants of haplogroup A and B and in those of the earliest branches of haplogroups C, D, E and F. By contrast, when there are many branches (called a starburst) after a specific mutation or group of mutations, we can infer rapid growth[115,116]. The major expansions are those of haplogoup F (seven branches) after an initial lag in population growth, and even more remarkable is the later expansion of haplogroup K (nine branches). These began in the last 40 kya and led to the major settlement of all continents from Africa, first to Asia, and from Asia to the other three continents. The tree of mtDNA (Fig. 4*b*) is more bushy, but there are more haplogroups because of the higher mutation rate. The general structures of the male and female genealogies in Figure 4 are the same. The earliest branches all remain in Africa; in both trees they clearly refer to the slowly growing hunter-gatherers. In both trees the major growth in Africa is due to a late branch, taking place in the second part of the last 100,000 years and clearly connected with the expansion to Asia. The M, N branches of the mtDNA phylogeny indicate the separation of the expansion from Africa to Asia into a southern and a northern branch. In the NRY genealogy, the southern branch is on average earlier than the northern, and includes mostly haplogroups C, D, H, M and L. Of these, H and L remained in India and part of C went to Oceania, the rest to Mongolia, Siberia, and eventually to northwest America (Na-Dene speakers). D went as far as southeast Asia and Japan. The northern Asian expansion remains mostly in East Asia (haplogroup O—its branch N has a major propagule to N.E. Europe, among Uralic speakers). Haplogroup I from north Asia generates what is probably the first major Paleolithic expansion to central Europe. G and J are found today in the Middle East and from there expanded to Europe, mostly in the south and probably with Neolithic farmers. R is found in Europe, India, Pakistan, and America, but an early branch seems to have returned to the central part of the Sahel in North Africa. Haplogroup Q generates most Amerinds, except for Na-Dene speakers and Eskimos. Haplogroup



BOB CRIMI

Fig. 5 Language families of the world. The 12 families of the Greenberg classification[88,122–125]. The Eurasiatic superfamily includes six families (most of which are recognized by most linguists) and an isolate, Gilyak, listed in the central column. The oldest family is the Khoisan that includes Bushmen and Hottentots, many of whom also belong genetically to the oldest haplogroups of both mtDNA and NRY. Australian and Indopacific are also old families. Other African languages are Niger-Kordofanian (mostly west Africa), Nilo-Saharan and Afroasiatic (that includes Semitic languages like Arab and Hebrew). American languages belong to three families: Amerinds were the first to migrate from Asia, according to some (Fagan, ref. 89) as late as 15 kya, and Amerind shows affinities with Eurasiatic. One of the other two American families is Na-Dene (belonging to Dene-Caucasian), a family that probably spread to Eurasia before Eurasiatic and includes Sinotibetan, spoken in almost all of China, as well as some isolated, probably relic, languages (Basque, a few Caucasian languages and Burushaski, spoken in N. Pakistan) that all survived the later spread of Eurasiatic languages. The third American family is Eskimo-Aleut, the last to spread to America from N.E. Siberia. The Austric family is very large and is spoken in S.E. Asia, Indonesia, all of Polynesia to the east and Madagascar to the west.

I is also found in north and central Europe, where it probably originated around 20,000 kya. A few indigenous individuals in America and Australia probably inherited European Y chromosomes.

Parallel developments to human evolution

What were the causes of the expansions that increased the number of modern humans by a million times or more over the past 100 kyr? Many capabilities distinguish modern humans from our predecessors (especially our closest relative, Neandertal): sophistication of stone tools, art, religion and, above all, language. We cannot totally exclude art or religion among Neandertals, but it is usually claimed that modern humans showed a very early, sudden development of art, with common themes related to magic, religion and an afterlife linked to the making of tombs[117,118], although there is evidence that many of these aspects of modern human behavior have a long history in Africa[119].

It has been rejected that Neandertal could speak languages like ours for anatomical reasons, but the evidence offered is considered inconclusive[120,121]. Modern human languages are mutually incomprehensible and superficially unrelated to each other. A general classification based on 12 language families has been suggested by Greenberg (Fig. 5)[88,122–125]. For geneticists like us, it seems natural to think that modern languages derive mostly or completely from a single language spoken in East Africa around 100 kya, given that today's genes also derive from that population. This does not mean that this was the only language in existence at the time; in parallel with genetic TMRCAs, it was the only language then existing that survived and evolved with rapid differentiation and transformation. Evidence supporting the existence of a common single language include the shared lexicon, sounds and grammar of present-day languages. Language, like many other forms of cooperation, must have originated as intrafamilial communication[126].

The expansion of modern humans may have been stimulated by the development of a new, more sophisticated culture of stone tools (called Aurignacian), which developed at the time of the expansion[127]. It is also very likely that navigation became available (or else the passage from southeast Asia to Oceania would have been impossible) and may even have been used earlier, such as in coastal south Asia[87], or later along the Pacific American coast.

Innovations that increased food availability may have then allowed groups to remain in the same area and to increase in size. This apparently happened in many parts of the world on a massive scale starting 10–13 kya with the adoption of agriculture and pastoralism. From the beginning of food production to the present, there must have been a thousand-fold population increase. Demographic growth in the well identified, specific areas of origin of agriculture must have stimulated a continuous peripheral population expansion wherever the new technologies were successful. 'Demic expansion' is the name given to the phenomenon (that is, farming spread by farmers themselves) as contrasted with 'cultural diffusion' (that is, the spread of farming technique without movement of people). Innovations favoring demographic growth would be expected to determine both demic and cultural diffusion[55,128,129]. Recent research suggests a roughly equal importance of demic and cultural diffusion of agriculture from the Near East into Europe in the Neolithic period[130,131].

Demic diffusion also results in the spread of the language of the initiators of the expansion. This probably occurred for Indo-European languages spreading from the Middle East to Europe and India[132], or for Austronesian languages spreading to Polynesia[133]. There is generally a strong correlation between linguistic families and the genetic tree of major populations[14,63], with some important exceptions.

There is generally a strong correlation of genetic tree clusters with language families[63,134], but there are also clear examples of historically dated language replacements. It is likely that these language shifts have become more common recently, with massive colonizations made possible by development of transportation and military technology.

Knowledge, which forms the basis of human behavior, is accumulated by 'cultural transmission' over generations and is subject to rapid change within generations. We have developed a theory of cultural transmission, in which the most important feature is 'duality': culture is transmitted either 'vertically' from parents to children or 'horizontally' between people with no particular age or genetic relationship[135].

Evolution under vertical transmission is slow, although faster than genetic evolution, and its time unit of one generation is the same. In assessing the importance of vertical transmission, we note that children are more prone to accept parental education because of specific susceptibilities during 'critical periods' of maturation[135,136]. For example, most 'mother tongues' are learned without accent only in the first 4–5 years. But under coercion or other special circumstances, the language of a whole population can be fully replaced in 3–4 generations. Although complete rapid replacement of languages may occur, such events are probably rare. Evolution under horizontal cultural transmission is usually much faster than under vertical transmission, and modern means of communication have made it exceptionally fast. Present-day humans are a 'cultural animal', but even today old customs may persist because some vertical cultural transmission remains important.

Humans carry many parasites or commensal organisms, some of which began their relationship with humans more than 100 kya. If their transmission is even partly vertical—as it is for hepatitis B virus—then their evolution is similar to that of humans, with origins in Africa and a spread first to Asia and then, independently, from Asia to the other three continents. It has been suggested that this is true of other viruses, such as polyomavirus[137], and also of the bacterium *Helicobacter pylori*[138], which was recently found to be the causative agent of gastric ulcer. It is likely that the same evolutionary properties will be detected for other commensals and parasites, indicating that at least part of their transmission is vertical.

Summary and outlook

Late twentieth century population genetic research was marked by a significant expansion in the available research tools through a greater appreciation of the level of polymorphism in the human genome. The development of assays for loci that allowed inferences about female (mtDNA)– or male (Y chromosome)–specific histories yielded new insights into human history. A growing appreciation of the importance of the genetic structure of human populations has seen the scope and application of population genetic studies expand.

In many ways we are currently hampered by the limited range of populations from which samples are available for detailed analysis. The World Cell Line Collection of 1,064 individuals from 52 populations is a beginning, but at least 5,000–10,000 from a more representative sampling of all continents would be preferable. Inferences about human history from small samples[13,17] are invariably fallible. Most published analyses concern genes chosen because of a putative relation to some phenotype, but sampling of DNA variation should be random with respect both to coding and non-coding regions[71].

Current statistical procedures to estimate the extent of migration or to measure the strength of selection from patterns of nucleotide variation are still primitive. New computational and analytical methods are needed for both if we are to increase our

confidence in the calculation of ages of mutations and TMRCAs. A key requirement here is the ability to separate selection from demographic effects. Comparative sequencing of primates may facilitate the detection and estimation of selection.

For haplotype determination, large samples of trios—father, mother, son—would be useful but expensive to obtain on a worldwide scale. Thus, improved algorithms for estimating haplotypes are required. Systems that combine SNPs and microsatellites may provide a way to map haplotypes more finely, to assess erosion of LD and to reconstruct the evolutionary history of gene regions[139]. Construction of somatic cell hybrids might, in the future, enable individual chromosomes to be isolated and made available for haplotypic analysis.

There is great scope for more interaction among anthropologists and population geneticists. Recent work by Hewlett *et al.*[140] suggests that correlation of microcultural variation and genetic variation in the same groups can be very informative about population interactions on various timescales. In the same vein, there are still few studies that compare patterns of variation in representative populations of human pathogens with those in their hosts. Perhaps this is a symptom of our focus on the genetics and diseases of developed countries and of the tiny fraction of available resources allocated to studying genetic variation in those populations about whom we have the least knowledge.

Acknowledgments

We thank P. Underhill and P. Oefner for comments on early drafts of the manuscript.

1. Hirszfeld, L. & Hirszfeld, H. Essai d'application des methods au probléme des races. *Anthropologie* **29**, 505–537 (1919).
2. Race, R.R. & Sanger, R. *Blood Groups in Man* (Blackwell Scientific, Oxford, 1975).
3. Ceppellini, R. *et al.* Genetics of leukocyte antigens. A family study of segregation and linkage. In *Histocompatibility Testing* (eds. Curtoni, E.S., Mattiuz, P.L. & Tosi, R.M.) (Munksgaard, Copenhagen, 1967).
4. Pauling, L., Itano, H.A., Singer, S.J. & Wells, I.C. Sickle-cell anemia, a molecular disease. *Science* **110**, 543–548 (1949).
5. Harris, H. Enzyme polymorphisms in man. *Proc. R. Soc. Lond. B* **164**, 298–310 (1966).
6. Lewontin, R.C. & Hubby, J.L. A molecular approach to the study of genetic hetrozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. *Genetics* **54**, 595–609 (1966).
7. Mourant, A.E. *The Distribution of Human Blood Groups* (Blackwell Scientific, Oxford, 1954).
8. Mourant, A.E., Kopec, A C. & Domaniewska-Sobczak, K. *The Distribution of the Human Blood Groups and Other Polymorphisms* (Oxford Univ. Press, London, 1976).
9. Mourant, A.E., Kopec, A C. & Domaniewska-Sobczak, K. *Blood Groups and Diseases* (Oxford Univ. Press, Oxford, 1978).
10. Nei, M. & Roychoudhury, A.K. *Human Polymorphic Genes: World Distribution* (Oxford Univ. Press, New York, 1988).
11. Botstein, D., White, R.L., Skolnick, M. & Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **32**, 314–331 (1980).
12. Rosenberg, N.A. *et al.* Genetic structure of human populations. *Science* **298**, 2381–2385 (2002).
13. Stephens, J.C. *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489–493 (2001).
14. Cavalli-Sforza, L.L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
15. Xiao, W. & Oefner, P.J. Denaturing high-performance liquid chromatography. *Hum. Mutat.* **17**, 439–474.
16. Oberacher, H. *et al.* Re-sequencing of multiple single nucleotide polymorphisms by liquid chromatography—electrospray ionization mass spectrometry. *Nucleic Acids Res.* **30**, e67.
17. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
18. Kimura, M. Evolutionary rate at the molecular level. *Nature* **217**, 624–626 (1968).
19. Przeworski, M., Hudson, R.R. & Di Rienzo, A. Adjusting the focus on human variation. *Trends Genet.* **16**, 296–302 (2000).
20. Hudson, R.R., Kreitman, M. & Aquadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
21. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
22. Muse, S.V. & Gaut, B.S. A likelihood approach for comparing synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* **11**, 715–724 (1994).
23. Yang, Z. & Nielsen, R. Synonymous and non-synonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.* **46**, 409–418 (1998).
24. Tishkoff, S.A. *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462 (2001).
25. Verrelli, B.C. *et al.* Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am. J. Hum. Genet.* **71**, 1112–1128 (2002).
26. Sabeti, P.C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837 (2002).
27. Enard, W. *et al.* Molecular evolution of *FOXP2*, a gene involved in speech and language. *Nature* **418**, 869–872 (2002).
28. Bamshad, M.J. *et al.* A strong signature of balancing selection in the 5′ *cis*-regulatory region of *CCR5*. *Proc. Natl. Acad. Sci. USA* **99**, 10539–10544 (2002).
29. Huttley, G.A. *et al.* Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. *Nat. Genet.* **25**, 410–413 (2000).
30. Toomajian, C. & Kreitman, M. Sequence variation and haplotype structure at the human HFE locus. *Genetics* **161**, 1609–1623 (2002).
31. Cavalli-Sforza, L.L. Population structure and human evolution. *Proc. R. Soc. Lond. B* **164**, 362–379 (1966).
32. Lewontin, R.C. & Krakauer, J. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195 (1973).
33. Weir, B.W. *Genetic Data Analysis II* (Sinauer, Sunderland, MA, 1996).
34. 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).
35. Hamblin, M.T., Thompson, E.E. & Di Rienzo, A. Complex signatures of natural selection at the Duffy blood group locus. *Am. J. Hum. Genet.* **70**, 369–383 (2002).
36. Hollox, E.J. *et al.* Lactase haplotype diversity in the Old World. *Am. J. Hum. Genet.* **68**, 160–172 (2001).
37. Gilad, Y., Rosenberg, S., Przeworski, M., Lancet, D. & Skorecki, K. Evidence for positive selection and population structure at the human *MAO-A* gene. *Proc. Natl. Acad. Sci. USA* **99**, 862–867 (2002).
38. Rana, B.K. *et al.* High polymorphism at the human melanocortin 2 receptor locus. *Genetics* **151**, 1547–1557 (1999).
39. Goldstein, D.B. & Chikhi, L. Human migrations and population structure: what we know and why it matters. *Annu. Rev. Genom. Hum. Genet.* **3**, 129–152 (2002).
40. Cavalli-Sforza, L.L. Some current problems in human population genetics. *Am. J. Hum. Genet.* **25**, 82–104 (1973).
41. Clark, A.G. *et al.* Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. *Am. J. Hum. Genet.* **63**, 595–612 (1998).
42. Ewens, W.J. in *Mathematical Population Genetics* 98–104 (Springer, Berlin, 1979).
43. Feldman, M.W. & Christiansen, F.B. The effect of population subdivision on two loci without selection. *Genet. Res.* **24**, 151–162 (1974).
44. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**, 229–232 (2001).
45. Jeffreys, A.J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nat. Genet.* **29**, 217–222 (2001).
46. Goldstein, D.B. Islands of linkage disequilibrium. *Nat. Genet.* **29**, 109–111 (2001).
47. Reich, D.E. *et al.* Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat. Genet.* **32**, 135–142 (2002).
48. Payne, R., Feldman, M.W., Cann, H. & Bodmer, J.G. A comparison of HLA data of the North American black with African black and North American caucasoid populations. *Tissue Antigens* **9**, 135–147 (1977).
49. Kidd, K.K. *et al.* A global survey of haplotype frequencies and linkage disequilibrium at the *DRD2* locus. *Hum. Genet.* **103**, 211–227 (1998).
50. Reich, D.E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
51. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
52. Edwards, A.W.F. & Cavalli-Sforza L.L. Reconstruction of evolutionary trees. In *Phenetic and Phylogenetic Classification* (eds. Heywood, V.E. & McNeill, J.) 67–76 (The Systematics Association, London, 1964).
53. Cavalli-Sforza, L.L. & Edwards, A.W.F. Analysis of human evolution. *Proc. 11th Int. Congr. Genet.* **2**, 923–933 (1964).
54. Cavalli-Sforza, L.L. & Edwards, A.W.F. Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Genet.* **19**, 223–257 (1967).
55. Menozzi, P., Piazza, A. & Cavalli-Sforza, L.L. Synthetic maps of human gene frequencies in Europe. *Science* **201**, 786–792 (1978).
56. Bowcock, A.M. *et al.* Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* **88**, 839–843 (1991).
57. Bowcock, A.M. *et al.* High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457 (1994).
58. Cavalli-Sforza, L.L. & Piazza, A. Analysis of evolution: evolutionary rates, independence, and treeness. *Theor. Popul. Biol.* **8**, 127–165 (1975).
59. Pritchard, J.K., Stephens, M. & Donnelly, P.J. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
60. Lewontin, R.C. The apportionment of human diversity. In *Evolutionary Biology* Vol. 6 (eds. Dobzhansky, T.H., Hecht, M.K. & Steere, W.C.) 381–398 (Appleton-Century-Crofts, New York, 1972).
61. Nei, M & Roychoudhury, A.K. Genic variation within and between the three major races of Man, Caucasoids, Negroids, and Mongoloids. *Am J. Hum. Genet.* **26**, 421–443 (1974).
62. Barbujani, G., Magagni, A., Minch, E. & Cavalli-Sforza, L.L. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* **94**, 4516–4519 (1997).
63. Cavalli-Sforza, L.L., Piazza, A., Menozzi, P. & Mountain, J. Reconstruction of human evolution; bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* **85**, 6002–6006 (1988).
64. Brown, W.M., George, M. Jr. & Wilson, A.C. Rapid evolution of animal mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* **76**, 1967–1971 (1979).
65. Johnson, M.J., Wallace, D.C., Ferris, S.D., Rattazzi, M.C. & Cavalli-Sforza, L.L. Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* **19**, 255–271 (1983).
66. Mountain, J.L., Lin, A.A., Bowcock, A.M. & Cavalli-Sforza, L.L. Evolution of modern humans: evidence from nuclear DNA polymorphisms. *Phil. Trans. R. Soc. Lond. B.* **337**, 159–165 (1992).
67. Cann, R.L., Stoneking, M. & Wilson, A.C. Mitochondrial DNA and human evolution. *Nature* **325**, 31–36 (1987).

68. Templeton, A.R. Human origins and analysis of mitochondrial DNA sequences. *Science* **255**, 737 (1992).
69. Chen, Y.-S., Torroni, A., Excoffier, L., Santachiara-Benerecetti, A.S. & Wallace, D.C. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* **57**, 133–149 (1995).
70. Ingman, M. Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).
71. Shen, P. *et al.* Population genetic implications from DNA polymorphism in random human genomic sequences. *Hum. Mutat.* **20**, 209–217 (2002).
72. Satta, Y., Klein, J. & Takahata, N. DNA archives and our nearest relative: the trichotomy problem revisited. *Mol. Phylogenet. Evol.* **14**, 259–275 (2000).
73. Rosenberg, N.A & Feldman, M.W. The relationship between coalescent times and population divergence times. In *Modern Developments in Theoretical Population Genetics* (eds. Slatkin, M. & Veuille, M) 130–164 (Oxford Univ. Press, Oxford, 2002).
74. Rogers, A.R. & Harpending, H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* **9**, 552–569 (1992).
75. Tang, H., Siegmund, D.O., Shen, P., Oefner, P.J. & Feldman, M.W. Frequentist estimation of coalescence times from nucleotide sequence data using a tree-based partition. *Genetics* **161**, 447–459 (2002).
76. Kajander, O.A., Karhunen, P.J., Holt, I.J. & Jacobs, H.T. Prominent mitochondrial DNA recombination intermediates in human heart muscle. *EMBO Rep.* **2**, 1007–1012 (2001).
77. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
78. Hammer, M.F. *et al.* Hierarchical patterns of global human Y-chromosome diversity. *Mol. Biol. Evol.* **18**, 1189–1203 (2001).
79. Paracchini, S., Arredi, B., Chalk, R. & Tyler-Smith, C. Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res.* **30**, e27 (2002).
80. Kingman, J.F.C. The coalescent. *Stochastic Processes and their Applications.* **13**, 235–248 (1982).
81. Hudson, R.R. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**, 203–217 (1990).
82. Griffiths, R.C. & Tavaré, R.C. Ancestral inference in population genetics. *Stat. Sci.* **9**, 307–319 (1994).
83. Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J. & Feldman, M.W. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**, 7360–7365 (2000).
84. 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).
85. Salem, A.H., Badr, F.M., Gaballah, M.F. & Pääbo, S. The genetics of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai Peninsula. *Am. J. Hum. Genet.* **59**, 741–743 (1996).
86. Sajantila, A. *et al.* Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc. Natl. Acad. Sci. USA* **93**, 12035–12039 (1995).
87. Finnilä, S., Hassinen, I.E., Ala-Kokko, L. & Majamaa, K. Phylogenetic network of the mtDNA haplogroup U in northern Finland based on sequence analysis of the complete coding region by conformation-sensitive gel electrophoresis. *Am. J. Hum. Genet.* **66**, 1017–1026 (2000).
88. Richards, M. *et al.* Tracing European founder lineages in the near eastern mtDNA pool. *Am. J. Hum. Genet.* **67**, 1251–1276 (2000).
89. Zerjal, T. *et al.* Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**, 1174–1183 (1997).
90. Richards, M., Oppenheimer, S. & Sykes, B. mtDNA suggests Polynesian origins in eastern Indonesia. *Am. J. Hum. Genet.* **62**, 1234–1236 (1998).
91. Kayser, M. *et al.* Melanesian origin of Polynesian Y chromosomes. *Curr. Biol.* **10**, 1237–1246 (2000).
92. Underhill, P.A. *et al.* Maori origins, Y chromosome haplotypes and implications for human history in the Pacific. *Hum. Mutat.* **17**, 271–280 (2001).
93. Oota, H., Settheetham-Ishida, W., Tiwaweck, D., Ishida, T. & Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nat. Genet.* **29**, 20–21 (2001).
94. Templeton, A.R. Out of Africa again and again. *Nature* **416**, 45–51 (2002).
95. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002)
96. Underhill, P. *et al.* The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann. Hum. Genet.* **65**, 43–62 (2001).
97. Lahr, M.M. & Foley, R. Multiple dispersals and modern human origins. *Evol. Anthropol.* **3**, 48–60 (1994).
98. Lahr, M.M. & Foley, R.A. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am. J. Phys. Anthropol.* **27**, 137–176 (1998).
99. Stringer, C. Coasting out of Africa. *Nature* **405**, 24–27 (2000).
100. Greenberg, J. *Language in the Americas* (Stanford Univ. Press, Stanford, CA, 1987).
101. Fagan, B.M. *The Great Journey: The Peopling of Ancient America* (Thames and Hudson, London, 1987).
102. Weidenreich, F. *Apes, Giants, and Man* (Univ. Chicago Press, Chicago, IL, 1946).
103. Wolpoff, M.H. Multiregional evolution: The fossil alternative to Eden. In *The Human Revolution: Behavioural and Biological Perspectives on the Origins of Modern Humans* (eds. Mellar, P. & Stringer, C) 62–108 (Princeton Univ. Press, Princeton, NJ, 1989).
104. Weiss, K.M. & Maruyama, T. Archeology, population genetics and studies of human racial ancestry. *Am. J. Phys. Anthropol.* **44**, 31–50 (1976).
105. Krings, M. *et al.* Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30 (1997).
106. Krings, M., Geisert, H., Schmitz, R.W., Krainitzki, H. & Pääbo, S. DNA sequence of the mitochondrial hypervariable region II from the Neandertal type specimen. *Proc. Natl. Acad. Sci. USA* **96**, 5581–5585 (1999).
107. Krings, M. *et al.* A view of Neandertal genetic diversity. *Nat. Genet.* **26**, 144–146 (2000).
108. Ovchinnikov, I.V. *et al.* Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* **404**, 490–493 (2000).
109. Harding, R.M. *et al.* Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**, 772–789 (1997).
110. Harris, E.E. & Hey, J. X-chromosome evidence for ancient human histories. *Proc. Natl. Acad. Sci. USA* **96**, 3320–3324 (1999).
111. Knowles, L.L. & Madison, W.P. Statistical phylogeography. *Mol. Ecol.* **11**, 2623–2635 (2002).
112. Kidd, J.R. *et al.* Haplotypes and linkage disequilibrium at the phenylalanine hydroxylase locus, PAH, in a global representation of populations. *Am. J. Hum. Genet.* **66**, 1882–1899 (2000).
113. Osier, M.V. *et al.* A global perspective on genetic variation at the ADH genes reveals unusual patterns of linkage disequilibrium and diversity. *Am. J. Hum. Genet.* **71**, 84–99 (2002).
114. The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
115. Slatkin, M. & Hudson, R.R. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562 (1991).
116. Donnelly, P. Interpreting genetic variability: the effects of shared evolutionary history. In *Variation in the Human Genome*, Ciba Foundation Symposium No. 197 (Wiley, Chichester, UK, 1996).
117. Anati, E. *The Intellectual Expressions of Prehistoric Man: Art and Religion.* Acts of the Valcamonica Symposium '79. Centro Camuno di Studi Preistorici, Capo di Ponte, Brescia, Italy, and (Editoriale Jaca Book SpA, Milano, Italy, 1983).
118. Conkey, M.W., Soffer, O., Stratmann, D. & Jablonski, N.G. (eds.) *Beyond Art: Pleistocene Image and Symbol*, Watts Symposium Series in Anthropology: Memoirs of the California Academy of Sciences No. 23 (California Academy of Sciences, San Francisco, CA, 1997).
119. 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).
120. Lieberman, P. & Crelin, E. S. On the speech of Neanderthal man. *Linguistic Inquiry.* **2**, 203–222 (1971).
121. Arensburg, B., Schepartz, L.A., Tillier, A.M., Vandermeersch, B. & Rak, Y. A reappraisal of the anatomical basis for human speech in Middle Paleolithic hominids. *Am. J. Phys. Anthropol.* **83**, 137–146 (1990).
122. Greenberg, J.H. *The Languages of Africa* (Bloomington, Indiana, 1963).
123. Greenberg, J.H. The Indo-Pacific Hypothesis. *Current Trends in Linguistics*, Volume **8**, 809–871 (1971).
124. Greenberg, J.H. Indo-European and Its Closest Relatives: The Eurasiatic Language Family: Grammar (Stanford Univ. Press, Stanford, California, 2000).
125. Ruhlen, M. *On the Origin of Languages: Studies in Linguistic Taxonomy* (Stanford Univ. Press, Stanford, California, 1994).
126. Eshel, I. & Cavalli-Sforza, L.L. Assortment of encounters and evolution of cooperativeness. *Proc. Natl. Acad. Sci. USA* **79**, 1331–1335 (1982).
127. Klein, R.G. *The Human Career* 2nd edn (Univ. of Chicago Press, Chicago, IL, 1999).
128. Ammerman, A.J. & Cavalli-Sforza, L.L. *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton, New Jersey, 1984).
129. King, R. & Underhill, P.A. Congruent distribution of Neolithic painted pottery and ceramic figurines with Y-chromosome lineages. *Antiquity* **76**, 707–714 (2002).
130. Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V. & Barbujani, G. Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc. Natl. Acad. Sci. USA* **95**, 9053–9058 (1998).
131. Chikhi, L., Nichols, R.A., Barbujani, G. & Beaumont, M.A. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. USA* **99**, 11008–11013 (2002).
132. Renfrew, C. Archaeology and Language: The Puzzle of Indo-European Origins (Jonathan Cape, London, 1987).
133. Bellwood, P.S. The colonization of the Pacific: some current hypotheses. In *The Colonization of the Pacific: A Genetic Trail* (eds. Hill, A.V.S. & Serjeantson, S.W.) 1–59 (Oxford Univ. Press, New York, 1989).
134. Cavalli-Sforza, L.L., Minch, E. & Mountain, J. Coevolution of genes and languages revisited. *Proc. Natl. Acad. Sci. USA* **89**, 5620–5624 (1992).
135. Cavalli-Sforza, L.L. & Feldman, M.W. *Cultural Transmission and Evolution: A Quantitative Approach.* (Princeton Univ. Press, Princeton, New Jersey, 1981).
136. Cavalli-Sforza, L.L. *Genes, Peoples and Languages* (North Point Press, New York, 2000).
137. Sugimoto, C. *et al.* Typing of urinary JC virus DNA offers a novel means of tracing human migrations. *Proc. Natl. Acad. Sci. USA* **94**, 9191–9196 (1997).
138. Covacci, A., Telford, J. L., Del Giudice, G., Parsonnet, J. & Rappuoli, R. *Helicobacter pylori* virulence and genetic geography. *Science* **284**, 1328–1333 (1999).
139. Mountain, J.L. *et al.* SNPSTRs: Empirically derived, rapidly typed, autosomal haplotypes for inference of population history and mutational processes. *Genome Res.* **12**, 1766–1772 (2002).
140. Hewlett, B.S., De Silvestri, A. & Guglielmino, C.R. Semes and genes in Africa. *Curr. Anthropol.* **43**, 313–321 (2002).
141. Latter, B.D.H. Genetic differences within and between populations of the major human subgroups. *Am. Nat.* **116**, 220–237 (1980).
142. Ryman, N., Chakraborty, R. & Nei, M. Differences in the relative distribution of human gene diversity between electrophoretic, and red and white cell antigen loci. *Hum. Hered.* **33**, 93–102 (1983).
143. Kivisild, T. *et al.* The genetic heritage of earliest settlers persist in both the Indian tribal and caste populations. *Am. J. Hum. Genet.* (in the press).
144. Thangaraj, K. *et al.* Genetic affinities of the Andaman islanders, a vanishing human population. *Curr. Biol.* (in the press).
145. Semino, O., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L.L. & Underhill, P. A. Ethopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**, 265–268 (2002).
146. Cruciani, F. *et al.* An Asia to sub-Saharan Africa back migration is supported by high-resolution analysis of human Y chromosome haplotypes. *Am. J. Hum. Genet.* **70**, 1197–1214 (2002).