Annu. Rev. Genomics Hum. Genet. 2004. 5:119–50
doi: 10.1146/annurev.genom.5.061903.180021

AFRICANS AND ASIANS ABROAD: Genetic Diversity in Europe

Guido Barbujani[1] and David B. Goldstein[2]
[1]*Department of Biology, University of Ferrara, 44100 Ferrara, Italy; email: g.barbujani@unife.it*
[2]*Centre for Population Genetics and Human Health, University College of London, London WC1E 6BT, United Kingdom; email: d.goldstein@ucl.ac.uk*

Key Words population structure, population history, Neolithic, Paleolithic, genetic drift, migration, selection, association studies

■ **Abstract** Besides its obvious intrinsic value, knowledge of population history, and of the demographic and evolutionary changes that accompany it, has proven fundamental to address applied research in human genetics. In this review we place current European genetic diversity in the context of the global human genome diversity and review the evidence supporting a recent African origin of the Europeans. We then discuss the results and the interpretation of genetic studies attempting to quantify the relative importance of various gene flow processes, both within Europe and from Asia into Europe, focusing especially on the initial, Paleolithic colonization of the continent, and on later, Paleolithic postglacial and Neolithic dispersals. Finally, we discuss how knowledge of the patterns of genetic diversity in Europe, and of their inferred generating processes, can be extremely useful in planning health care and in biomedical research.

INTRODUCTION

Modern Europeans are African immigrants. Bipedalism, thick dental enamel, large flat molars, and small canine teeth—in brief, the skeletal features typical of humans—appear in African fossils about 3.6 million years old. Fossils also show that for some 2.5 million years all human ancestors were restricted to Africa (173). Aside from an archaic human form dated at 1.7 million years old, which was recently discovered in the Caucasus (188), human presence in Eurasia is documented only in the last million years. Genetic data on current human diversity do not contain much information about such distant time periods. However, calculations based on comparisons of orthologous genes suggest that the human ancestral lineage became independent from that leading to chimpanzee 4.5 million years ago (mya) (177).

As for what happened in more recent times, and how Europeans came to be the way they are, opinions differ, although levels of disagreement are lower now than ten years ago. Along with paleoanthropological and archaeological data, genetic data have proven crucial for addressing this and related controversies. In what follows, we first place European genetic diversity in the context of global diversity of the human species. In the second section, we see how information on current human variation, along with data on ancient DNA, fossil, and archaeological evidence, gives us a fairly precise picture of the early stages of the peopling of Europe, namely the transition between anatomically different human forms. In the third section, we use current genetic diversity to reconstruct aspects of the European prehistory and early history, and in the fourth section describe how genetic data have been used to understand how the prehistoric European population increased in numbers to reach its current size. Finally, in the fifth section, we stress that population genetic concepts and data are crucial to address applied research, in particular to understand the distribution of genetic disease, and to refine tools for uncovering genetic differences among people that influence disease and its treatment.

GLOBAL HUMAN GENETIC DIVERSITY

In the animal kingdom, some species, including Asian lion, puma, and cheetah, show very little genetic diversity (51). However, most organisms, including humans, have a considerable amount of genetic variation, although humans are by no means the most variable of species (105).

In terms of how the variation is structured geographically, there are a continuum of possibilities. At one end one has species in which there are sharp regional or continental discontinuities, such that most of the variation is due to differences between groups. At the other end of the continuum, one has a geographically undifferentiated species, in which most of the variation is due to differences between individuals, wherever they are found. Species with limited dispersal, such as land snails (77), bats (113), and giant tortoises (36), tend to be nearer the former pole, whereas more mobile species, such as gray wolf, coyote (189), and the Antarctic krill *Euphausia superba* (64), tend to be closer to the latter, although the full range is represented.

Where does our species belong in that range of possibilities? For several centuries, the idea that all humans naturally fall into one of a few biologically different races was unchallenged. However, it was also difficult to reach any consensus on the number and the definition of such races (37). Now we know why. Research in the past 30 years shows that variation in humans does not apportion neatly into a small number of discrete racial or population groups. As Table 1 illustrates, independent studies demonstrate that a large fraction of the global human diversity is contained within populations, and that differences among continental human groups are comparatively small.

TABLE 1 Estimated fractions of the global human diversity at three hierarchical levels of population subdivision

Polymorphism	Reference	N of loci	Within population	Between populations, within continent	Between continents
Proteins	(103)	17	85.4	8.3	6.3
Proteins	(102)	18	85.5	5.5	9.0
MtDNA	(62)	(HV-I)[a]	75.4	3.5	21.1
Autosomal DNA	(10)	109	84.4	4.7	10.8
MtDNA	(156)	(HV-I and -II)	81.4	6.1	12.5
Y chromosome	(156)	10	35.5	11.8	52.7
Autosomal DNA	(92)	90	84.8	1.6	13.6
MtDNA	(92)	(HV-I)	71.5	6.1	23.4
Y chromosome	(92)	10	83.3	18.5	−1.8
Alu insertions	(147)	21	82.9	8.2	8.9
Y chromosome	(147)	14	42.6	17.3	40.1
Beta-globin	(147)	1	79.4	2.8	17.8
Median, all loci			82.2	6.1	13.0
Normalized[b] median, all loci			81.1	6.0	12.9
Median, autosomal			84.6	5.1	9.9
Normalized[c] median, autosomal			84.9	5.1	10.0

[a]HV-I and HV-II are the hypervariable regions I and II, respectively, of the mitochondrial genome control region.

[b]Normalization obtained by dividing by 1.013.

[c]Normalization obtained by dividing by 0.996.

The data underlying these studies are the genetic variances at a number of loci, each ideally representing independent genomic regions. The actual value of the global variance is of little interest. More important is the way that variance is distributed at three hierarchical levels, namely among individuals of the same population, among populations of the same continent, and among different continents. The methodology introduced by Lewontin (103), and currently implemented in the AMOVA algorithm (62), allows one to summarize by a simple statistic, Wright's F (196a), or by related statistics, patterns of differentiation that may be complex, and can thus be related with parameters such as effective population size and migration rates. The overall genetic diversity is broken down at three levels, by comparing individual genotypes between members of different continents (for the sake of simplicity, we refer to their average difference as d_C), of different populations in

the same continent (average: d_B), and of the same population (average: d_A). The actual calculations depend on whether only allele frequencies are available, or also measures of molecular distance between alleles. In all cases, the variance component representing within-population diversity is proportional to d_A, the effect of belonging to different populations of the same continent is a function of ($d_B - d_A$), and the variance component representing the effect of belonging to different continents is a function of ($d_C - d_B$) (62).

Because two variances are obtained by subtraction, it may occasionally happen that one of them takes a negative value. One example is in a study of microsatellites (92). The negative variance observed (which, however, is not significantly different from 0) means that, on average, members of different continents do not differ for those markers more than members of the same continent. Similarly, low variances within a continent may sometimes reflect the fact that some continents were represented by one population only.

Some inconcistencies are evident, most notably for the Y chromosome, although there are explanations for them (147). To summarize the results, the median can provide a more reliable guide than the mean, and after correcting for the fact that the sum of medians is not necessarily equal to 100%, we see that members of different populations of the same continent are 6% more distant genetically than two (unrelated) members of the same population, and that coming from different continents adds a further 12.9% to the expected genetic differentiation. These ballpark figures do not imply that any clustering of individuals based on their genotypes is impossible. However, they indicate that clustering is based on small genetic differences; in fact, smaller than their error in many cases. The question, then, becomes whether those differences, albeit small, are large enough to place individuals within their geographic area of origin (e.g., their continent) with good confidence.

The simplest methods to address this question do not use explicit assumptions and cluster genotypes based on their degree of similarity. Graphic representations of these clusters are given by trees and networks, or by the two-dimensional plots that can be obtained through various forms of multivariate analysis, such as multidimensional scaling. In some cases, these methods reveal clear substructure (20), but they do not provide any natural scheme for statistically assessing the resulting groupings.

One way to increase the power of the numerical methods is to analyze the data under models making assumptions, either on the existence of groups in the data, or on the Hardy-Weinberg equilibrium in each population. In the former case, the groups are predefined, and discriminant analysis or similar numerical techniques assign individual genotypes to the most likely of them (132, 142). In the latter case, there is no restriction to the potential number of groups one can recognize, nor does any geographic factor restrict the possibility that an individual genotype is assigned to one such group. A likelihood approach that has gained wide acceptance, Structure, (128) seeks to infer the substructure directly and to estimate the allele frequencies of each identified cluster. For a given number of

subgroups the method provides a statistically robust estimate of the substructure, but the method for identifying the most likely number of subgroups (denoted *K*) is, by the authors' admission, ad hoc.

Despite the differences among methods, there are similar answers. There is sufficient variation to allow one to place most genotypes on the right continent by discriminant analysis (147), and when analyzed by structure, clusters emerge that include mostly individuals of similar geographic origin (147, 148, 196). However, at least 27% of the genotypes end up in the wrong continent by discriminant analysis, and the clusterings obtained using different sets of loci do not overlap. Predictably, at the subcontinental level, the assignments are less accurate (147).

On one hand, these results confirm the common observation that the physical aspect of an individual tells us something about the individual's ancestry (20, 21, 60). On the other hand, they show why it has thus far proven impossible to agree on the major subdivisions of humankind. Different loci suggest different subdivisions and/or different boundaries among them. We do not know yet if that simply means that the samples are inadequate, and that when one uses many more markers stable groups will emerge. Alternatively, human diversity may be such that no clear boundaries, but only clines, exist, as a recent reanalysis of data suggests (Serre & Paabo, personal communication). For now, it is safe to say that analyzing discrete samples (such as different ethnic groups in the United States) has sometimes led to the identification of boundaries at the regional level (46), but in worldwide analyses nobody has recognized the sharp genetic boundaries that could define objective and stable clusters of human genotypes that one could equate to the races of other species. Thus, we believe that available data clearly reject the idea that a few distinct groups can fully explain the geographic component of human genetic variation (6, 179), but it remains unclear precisely how significant is the component of the total human genetic variation that can be explained by a simple scheme such as continent of origin.

The absence of clear-cut boundaries between groups does not mean that all human populations are identical. On the contrary, given a sufficient number of loci, almost any population of the world can be shown to differ significantly from any other population, including its neighbors (10). Especially in zones where gene flow is restricted, genetic differences do exist (12, 22, 168). Studies of human global diversity show for randomly chosen loci, and especially for autosomal loci, that average genetic differences among populations and continents are relatively small, and their degree of correlation across loci varies depending on the populations considered. Therefore, finding that two populations differ for gene A does not allow one to confidently predict whether they will also be different for gene B.

This finding has considerable importance in biomedical research. Some argue that the self-assessed racial or ethnic origin is a potentially misleading substitute for the actual description of an individual's genotype (40). Others maintain that racial labels help to predict patient disease risk and response to pharmacological treatment, and that self-identified ethnicities correspond well with continent of origin (76, 143), so that membership in one of just five distinct groups can largely

account for the geographical component of the global variance. The issue is complex, but it is worth emphasizing that fixed differences, even among continents, are rare among humans. For instance, European-specific haplotype blocks seem to represent only 2% of the total blocks studied (68). Accordingly, very few alleles are both confined to one geographical area and common enough in that area to dictate medical management for that area's population (95). Nevertheless, the genome is a big place, and although rare, this does happen (32), and could be medically significant in some circumstances.

GENETIC DATA ON THE EARLY STAGES OF THE EUROPEAN POPULATION HISTORY

Although there is a general consensus that the genus *Homo* originated in Africa in the Pliocene, not all scientists agree as to whether Africa is also the only place where our species, *Homo sapiens*, evolved in the Pleistocene. One view is that anatomically archaic people from Asia and Europe transmitted their genes to contemporary people, and hence there has been only one human species across the Pleistocene, from 1.8 mya to the present time (multiregional theory: 84, 182). However, other investigators recognized up to eight human species in the same time period (178) and view current humans as the result of one African species' expansion that replaced all archaic human forms in Eurasia (replacement theory: 66). Between these two views there are intermediate possibilites (134). In Europe, the question is whether or not anatomically archaic forms, and in particular the Neandertal people who are documented between 300,000 and 30,000 years ago, transmitted some of their genes to contemporary people.

Morphological studies suggest that Neandertals represent a separate lineage, and perhaps a different species, than modern humans (66, 99), hence supporting the replacement theory. Genetic analyses of contemporary humans are largely regarded as consistent with this interpretation (for different views see 81, 180). When it is possible to place their root, human phylogenetic trees consistently show an African origin (67, 89, 116, 187, 191). African populations are both the most divergent from the others and the most internally diverse at many loci (2, 92, 171, 184, 198, 202). They contain a broad range of haplotypes, only subsets of which, conversely, are found in the other continents (2, 68, 187), and it is in Africa that we observe the lowest levels of linkage disequilibrium (LD) worldwide (183). Data from the Y chromosome (42, 53) and from the Dystrophin gene (201) suggest that descendants of African emigrants occasionally reentered Africa, bringing back portions of genetic diversity that had evolved in Asia. All these data support the view that anatomically modern humans were an African taxon up to a recent past, perhaps less than 100,000 years ago (59, 151). However, an explicit test of the presence of Neandertal genes in contemporary Europeans requires some knowledge of the Neandertals' genes, which was simply out of reach until recently.

The first mitochondrial sequence of a Neandertal specimen, from the Neander valley in Germany, was published in 1997 (98). It falls outside the range of variation

of contemporay Europeans, and of contemporary humans. The age of the common mitochondrial ancestor of Neandertals and contemporary humans was 550,000 to 690,000 years ago, whereas figures between 120,000 and 150,000 are estimated for contemporary humans from the same mitochondrial data (98).

Three further Neandertal sequences, from the Caucasus (121), Croatia (97), and again from the Neander valley (152), confirm that Neandertals differ sharply from modern humans (17 substitutions from the closest modern human in the database, versus a maximum of 4 nucleotide differences between Neandertals). Neandertals show no special resemblance to modern Europeans, as opposed to people from other continents. When considering both hypervariable mitochondrial regions, their average differences from Europeans and Africans are 35.3 ± 2.1 and 33.9 ± 2.1 substitutions, respectively (97).

These mitochondrial differences, albeit large, may not be totally inconsistent with a multiregional model (78, 84, 118). For instance, Nordborg (118) showed that the differences between Neandertal's and modern mitochondrial sequences are sufficient to rule out random mating, but not more complicated models of interbreeding. In genetic terms, the question is whether anatomically modern humans who coexisted in Europe with Neandertals for millennia (45), sometimes referred to as Cro-Magnons or Cro-Magnoids, can be regarded as transition forms.

The recently published sequences of the hypervariable region I of two early anatomically modern Europeans, dated at 23,000 and 25,000 years ago, demonstrate a sharp genetic discontinuity also between Neandertals and Cro-Magnoids (8, 23). The Cro-Magnoid sequences fall within the range of contemporary European sequences, and differ sharply from the temporally much closer Neandertals. These results are hard to reconcile with the hypothesis of some genetic continuity among Neandertals and Cro-Magnoids, or at least among Neandertal and Cro-Magnoid women. No genetic study can ever disprove that some interbreeding may have occasionally occurred. In addition, ancient DNA studies are currently limited to the mitochondrial genome (87), and hence some contribution of Neandertal males to the current gene pool may not be ruled out. Still, several aspects of the available genetic evidence are at odds with the prediction of the multiregional theory in its classic version (182). Conversely, all ancient DNA data fit well into the predictions of a model in which the anatomically archaic population of Europe became extinct, between 20,000 and 30,000 years ago. Presumably, many factors concurred to cause that extinction, but two may have been very important, namely climatic instability (111) and competition with anatomically modern humans who arrived from Africa via Asia 35,000 to 40,000 years ago (174, 203).

GENETIC DATA ON RECENT STAGES OF THE EUROPEAN POPULATION HISTORY

The uncertainty of the fate of Neandertals concerns whether they gave an extremely small contribution to the modern European gene pool or none at all. Even if the former is true, for all practical purpose it is safe to consider that the

history of the Europeans' genes began when the first anatomically modern humans entered Europe from the Near East 35,000 to 40,000 years ago (120). We cannot reconstruct that history in detail, but information on important aspects of it can be extracted from the current patterns of genetic variation (190), and from comparison of biological and nonbiological evidence.

Starting from Haldane's (79) pioneering study on the AB0 blood group, most investigators have found that patterns of genetic diversity in and around Europe are simple and can be described as clinal. The allele frequencies of many protein markers have a maximum (or a minimum) in the southeastern corner of the map, in the Near East, or in Anatolia, and decrease (or increase) smoothly as one proceeds west and north (7, 11, 28, 112, 165, 166). Such a high degree of order cannot possibly be the result of isolation by distance, that is, of a combination of genetic drift and short-range gene flow (13). Therefore, these gradients must have been generated by a process affecting much of the continent and originating from its southeastern Asian border. Fix (65) proposed that that process might be an adaptation to variable environmental pressures. That may be true for single loci, such as those for lactase persistence in milk-drinking cultures (16), or for nonrecombining regions of the genome (96, 114). However, the parallel gradients of many independent loci are evidence that a common factor, hence not selection (106), shaped variation over the whole genome of the Europeans.

Figure 1 summarizes the main large-scale demographic transitions in prehistoric Europe, as inferred from radiocarbon dates (124). A first wave of Paleolithic settlers dispersed from the Near East (110, 120). These were bands of hunters and gatherers associated with the Gravettian culture who came to occupy a large part of the territory at low densities. Another group of hunter-gatherers, associated with the Solutrean culture, may have arrived from North Africa, but occupied only a small part of southwestern Europe (120). The second major continental shift in the archaeological record is the diffusion of artifacts, starting 10,000 years ago, demonstrating farming activities and animal breeding, which define the Neolithic period. These artifacts, initially concentrated in the Levant and in Anatolia, spread north and west throughout Europe, as farming-based societies came to occupy much of the continent in approximately 5000 years (1). The first Paleolithic settlements, and later Neolithic communities, expanded roughly along the same routes, so the geographic patterns generated by the two processes may not be very different.

Many things happened between these two expansions, and after the Neolithic period. However, to understand the origin of the continental clines, post-Neolithic migration is probably not important. None of the migration processes documented in the archaeological and historical records seems to have had a sufficient geographical scope to determine spatial structuring of genes across all Europe (169). Also, historical migrations occurred in all directions (169), and hence they are not expected to generate a continent-wide pattern. Therefore, the clinal patterns are there not because of, but despite, recent gene flow.

On the contrary, the profound climate change (44) that occurred at the end of the Paleolithic may have affected European genetic diversity. After anatomically modern people settled in Europe, the average temperatures decreased steadily, leading to a glacial maximum around 18,000 years before present (BP). In that period, many species went extinct, whereas others, probably including humans, retreated into southern areas of milder climate, the glacial refugia, from which they re-expanded as the climate improved (176). Based on paleobiological evidence, three refugia have been recognized in Iberia, in Italy, and south of the Balkanian range (194).

The question is whether the current European genetic gradients were established during the Paleolithic or the Neolithic period. Post-Neolithic demographic phenomena had an impact, sometimes a strong one, at the local scale, but cannot account for the fact that many allele frequencies form regular gradients over much of the continent.

In a principal component (PC) analysis of 120 allele frequencies, Menozzi et al. (112) identified a multilocus cline centered in the Levant, and extending west to Iberia, and north to the British Isles, which overlapped with the radiocarbon dates of the first Neolithic industries (see also 136, 170). A recent analysis of prehistoric population densities inferred from archaeological data also showed that the likely demographic impact of Neolithic dispersal was greatest in southeastern Europe, and decreased toward Iberia and in the north (100). These observations led to the proposal of a model that was termed Neolithic demic diffusion (NDD) (1, 112). According to the NDD model, four factors generated genetic gradients in Europe, namely: (*a*) the existence of gene-frequency differences between early farmers of the Near East and European hunters and gatherers; (*b*) an increase of the farmers' population sizes, due to the increased food made available by their new subsistence technologies; (*c*) a farmers' range expansion from the southeast; and (*d*) limited immediate admixture with the communities of hunters and gatherers encountered in the expansion. Factor (*d*) is important because if the two groups did not merge immediately after their contact, in each area the farmers would have kept increasing in numbers, whereas hunters and gatherers would have stayed at their limited population sizes. Theoretical calculations (2) and computer simulations (13, 135) show that the final result of such a process is the establishment of broad clines, at all loci to which condition (*a*) applied, over the area affected by the Neolithic expansion. Gradients resembling those observed in empirical studies are observed in computer simulations when two thirds or more of the modern population descended from Neolithic immigrants (13, 135). Conversely, the minimum proportion of Neolithic immigrants necessary to account for the observed clines has not yet been definitively determined.

An alternative model has emerged from the analysis of mitochondrial diversity. The first studies of the mitochondrial control region found little geographical structuring, with most European populations showing similar mtDNA haplotypes (137, 138). The inferred evolutionary networks showed that mtDNA haplotypes form clusters, which were called haplogroups. Richards et al. (137–139) defined these

haplogroups, calculated their ages based on reasonable estimates of mitochondrial mutation rates, and observed that for most haplogroups these ages indicated common molecular ancestors in Paleolithic times. The 95% confidence intervals for haplogroup ages proposed by Richards et al. (139) ranged from 53,600–58,900 for haplogroup U, to 6100–12,800 for T1, a subcluster of T. These figures were regarded as evidence that most European mitochondrial sequences descended from local Upper Paleolithic ancestors, and that only a small fraction of the European mitochondrial gene pool, less than 25% (139), entered Europe in the Neolithic. The distribution of haplogroup V, showing similarities between Iberia and the Saami (Lapps) of northern Finland, was interpreted as parallel to the second PC of Menozzi et al. (28, 29, 112) and proposed as evidence of the genetic effects of postglacial re-expansions from the Iberian refugium (185). However, ancient DNA data (90), and the analysis of a broader data set, led Torroni et al. to conclude that the similarities between Iberia and northeastern Europe may reflect recent, probably post-Neolithic, contacts (186).

What we call the postglacial expansion (PGE) model can be summarized as follows: (*a*) during the last glacial maximum the first Paleolithic settlers withdrew south and survived in small numbers in the glacial refugia; (*b*) after the ice sheet retired, the refugia populations increased in numbers and expanded north, much like other plant and animal populations (176); (*c*) Neolithic immigration from the Levant was comparatively small (204), so that the current patterns of genetic diversity in Europe reflect the founder effects occurring through the north-south glacial contractions and postglacial expansions of the Paleolithic people. Under the PGE model, the spread of farming is largely due to cultural contacts, with limited input of genes from the Near East (5, 138). Conversely, under the NDD model, farming spread because the farmers did (1, 8) (Figure 2).

Despite the schematism of these syntheses, the NDD and PGE models allow one to make some testable predictions, listed in Table 2. There is a third possibility, namely that the most significant episode in the European demographic history was the first Paleolithic colonization (an implication of 197). Under the PGE model, southeast-northwest clines are regarded as the exception, not the rule, over Europe. More important for the PGE model are the founder effects resulting from population contractions, and the successive northward re-expansions in response

TABLE 2 Schematic predictions of three models of the European demographic history

	European gene pool largely established in the		
	Upper Paleolithic period	**Late (postglacial) Paleolithic period**	**Neolithic period**
SE-NW Clines	Common	Uncommon	Common
N-S Clines	Uncommon	Common	Uncommon
Proportion of "Paleolithic" Alleles	Large	Large	Small

to climate changes. On the contrary, southeast-northwest gradients are expected both under the NDD model (because of the farmers' expansion from the Levant) and under the model whereby the European population structure was established during the first Upper Paleolithic settlement (because of repeated founder effects in the course of the dispersal from the Levant of bands of hunters and gatherers).

For several years mitochondrial DNA was the only DNA marker investigated on a sufficiently large geographic scale, hence the idea that its patterns could be generalized to the entire genome. However, later studies showed that broad southeast-northwest clines are common at the DNA level too, for autosomal (34, 35) and Y-chromosome polymorphisms (25, 130, 149, 157). In addition, a gradient of mitochondrial diversity radiating from the Near East was also significant, but only in southern Europe, whereas there was no clear mitochondrial pattern in the north (39, 161). A later study based on principal component analysis (141) confirmed this finding. Although not all authors attribute the origin of those clines to Neolithic dispersal (107, 141), it is clear that geographic patterns of nuclear DNA polymorphisms resemble those described at the protein level.

There are various attempts to directly relate the representation of mtDNA or Y-chromosome lineages to the proportion of Neolithic immigrants. For instance, Semino et al. (158) identified southeast-northwest clines for four Y-chromosome haplogroups, and proposed that their overall frequency, 25%, represents the fraction of Neolithic immigrants among Europeans. Another approach was to equate the proportion of Neolithic immigrants to the explained variance of the first PC axis for European allele frequencies (30, 140). However, these inferences were made without benefit of formal inferential models, and have subsequently come under strong criticism on numerous grounds.

Neither clines of particular lineages, nor relatively deep braching lineages in Europe, provide direct estimates of the amount of Neolithic immigration (7, 33, 72). Unfortunately, there are no other completely satisfactory approaches, although recent efforts to model population admixture provide an important direction. Although it is sometimes useful to schematically oppose the alternative hypotheses, the spread of agriculture was not likely caused by a complete population replacement, or solely by cultural transmission. Rather, different processes, both demographic and cultural, were probably important in different regions and at different moments (204). Consequently, one way to compare the NDD and the PGE models is to quantify the relative contribution of immigrating Neolithic farmers, and of previously settled hunter-gatherers, in various areas of Europe (in other words, to estimate Neolithic admixture). In those studies (17, 31), current populations are regarded as hybrids among two or more populations that mixed in the past. If the ancestral populations were available for study, it would be a straightforward problem to gain a statistical estimate of the derivation of current populations. Of course, this is not the case.

Semino et al.'s (158) calculations do not represent the appropriate way to quantify admixture. Unless admixture occurred between populations where the alternative alleles were fixed (an unlikely occurrence in Europe, as noted above), an

average allele frequency differs from an admixture rate. Suppose that a hybrid population H is formed by admixture between two parental populations, A and B, whose allele frequencies at a biallelic locus at the moment of hybridization were, respectively, p_H, p_A, and p_B. If p_A is 0.10, p_B is 0, and A and B contributed equal numbers of individuals to H (so that the real admixture rate is 50%), p_H would be 0.05 and, under the estimation procedure followed in Reference 158, that would also be the admixture estimate, i.e., one tenth of the real value.

As an alternative, Chikhi et al. (33) applied a maximum-likelihood method to the data of Semino et al. (158) to estimate the Near Eastern, and hence presumably Neolithic, Y-chromosome admixture in seventeen European populations. Their model assumes that admixture occurred at a given moment in the past, and attempts to use the descendendant populations, both parental and admixed, to estimate the relative contributions to the ancestral. All populations are allowed to have changed in allele frequency since the time of admixture because of genetic drift. The model then provides a likelihood-based estimate of the amount of drift, and the relative contributions of each parental population, based on simultaneous analysis of all haplotypes in the populations (Figure 3).

Each estimate had a large standard error, confirming that estimates based on a single locus are likely not precise. Despite that, a clear geographical pattern was evident, with Neolithic contributions representing more than 80% of the total gene pool in the Balkans, and lower values in central and western regions, down to a minimum of 15% to 30% in France and Germany. Because each region did not evolve independently from the others, but was presumably connected to its neighbors by genetic exchanges, there is no simple way to combine the results obtained for each region in a summary statistic. Regression analysis shows that the Neolithic admixture rate, 22%, proposed by Semino et al. (158) for the whole continent falls below the 0.999 confidence limit of the maximum-likelihood values estimated from the same data (33).

Recent developments are now making it possible to consider more complex scenarios, in which more than two parental populations contribute genes to the hybrid populations (52). In their study of Y-chromosome single nucleotide polymorphisms (SNPs), Rosser et al. (149) confirmed the presence of southeast-northwest clines, which they associated with the Neolithic demic diffusion. They also found evidence of other geographical trends that apparently have little to do with either the Paleolithic or the Neolithic expansions. These and similar results suggested a contribution of northern Asian (101, 192, 199) and northern African (131, 155) ancestors to the European gene pool. An admixture study based on eight autosomal, mitochondrial, and Y-chromosome markers found that only a small fraction of the current European genes seems to come from North Africa, whereas a component reflecting gene flow from northern Asia is nonzero, but largely restricted to the northeast region of the continent. The estimated Near Eastern contribution is, on average, large and decreases as one moves from east to west, with maxima around 80% in the Balkans and minima of 20 and 34% in the British Isles and Iberia, respectively (53). Therefore, admixture rates seem in agreement with the predictions of a model in which Neolithic immigrants from the Near East contributed a

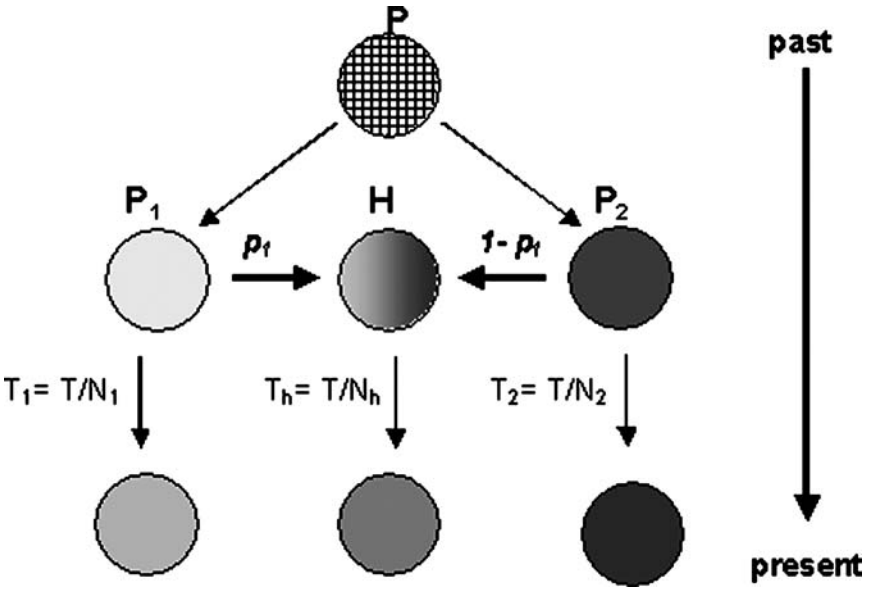

Figure 3 A model to estimate admixture. P is a founder population that splits into two populations, P_1 and P_2, at an unspecified moment in the past. A hybrid population, H, is formed by members of the P_1 and P_2 parental populations, whose proportions in the hybrid are p_1 and $(1 - p_1)$, respectively. The impact of genetic drift on allele frequencies, in the parental and in the hybrid population, depends on both the time elapsed (T), which is the same for all populations, and on the effective population sizes (N_1, N_2, N_h). A Monte Carlo–Markov Chain method is used to estimate posterior probability distributions for the parameters p_1, T_1, T_2, and T_h.

large share of the alleles in the genome of current Europeans. However, it is still unclear how misidentification of the ancestral populations may bias the results, or in other words to what extent current Near Eastern and Basque samples represent an acceptable approximation for the Neolithic and Paleolithic populations, respectively, that actually mixed.

Models of straightforward population divergence are even less applicable than any of the admixture models considered, but they nonetheless support a recent separation of the European gene pools. By analyzing a data set of seven autosomal loci by a measure of genetic distance for microsatellites that is largely linear with times since separation (74, 75), Chikhi et al. (34) found that the most likely separation times between European populations are, with one exception (the Saami or Lapps), in the Neolithic period. Similarly, occurrence of identical mitochondrial haplotypes between populations was investigated using coalescent-based simulations. Most comparisons gave levels of allele sharing compatible with a Neolithic separation of the gene pools, and once again the comparisons suggesting an older, Paleolithic separation involve Near Eastern or Saami samples (7). These

calculations are approximate and depend on several assumptions that still need to be validated. Nevertheless, they consistently suggest that the European gene pools were largely separated in Neolithic times, although it is difficult to quantify this statement (27).

Although post-Neolithic gene flow does not likely account for the main European pattern of genetic diversity, the consequences of relatively recent, short-range migration are evident at a smaller geographical scale (164). The genetic distances between populations correlate well with their expectations based on an extensive set of historical data from 2000 BC up to present time (164, 169), and the same ethnohistorical data better predict patterns of cancer mortality than genetic data (167). Passing from the continental to the regional level of analysis, the patterns of genetic diversity become more complicated (49), and there are a large number of factors that can be invoked to explain them, including isolation determined by both geographic and cultural barriers (leading populations to diverge: 12, 19, 38, 55, 88, 108, 150, 162, 163, 168, 172, 199), and various historical contacts and exchanges (leading populations to become genetically more similar: 18, 28, 29, 109, 126, 200). However, because the differences between populations are generally small, predictions based on one or a few genes may be inaccurate. For instance, Basques and Sardinians are notorious outliers for many genetic polymorphisms (29), but are very similar to all other Europeans for allele frequencies at loci of pharmacogenetic relevance (NAT2, CYP1A1, and GSTM1) (69).

Should one conclude that inferences based on allele frequency analysis and on genealogical analysis are inconsistent, with the former supporting the NDD model, and the latter supporting the PGE model? We believe that the answer is no, and that these differences depend mostly on the method used. The Y-chromosome and mtDNA data have usually been analyzed in a phylogeographic framework, seeking to make direct connections between lineages and ages of population processes. When more formal methods are used, the different data types do not disagree. The crucial point is that whether the current Europeans are mainly descended from Neolithic or Paleolithic ancestors is a question about population history. As such, it can only be answered at the population level. Data on DNA in ancient populations are too scant (48, 80) and are not going to increase dramatically any time soon. Therefore, the most direct way to answer the question is to consider data on modern populations, in the light of evolutionary models. One must keep in mind that ages of particular genealogical lineages present in defined geographical areas do not correspond directly with the arrival of the population in that area, or with any other phenomenon at the population level. Instead, the distributions of lineages in modern populations reflect the combined and complex effects of processes affecting both the DNA molecules within the fertile cells, and the populations of individuals who reproduce by means of those cells. A clear example comes from the study of the major histocompatibility complex, MHC, whose alleles coalesce 10 to 17 mya, but owe their current distribution to recent changes in selection pressures (125). Another example is the (C/T-13,910) regulatory variant of the gene coding for lactase-phlorizin hydrolase, which determines lactose tolerance.

This variant is very old from the molecular standpoint, but it reached its current high frequency in Europe only after dairy cultures were introduced, i.e., in the last 10,000 years (58).

To better illustrate this concept, note that the age of some European mitochondrial haplogroups (lower confidence limit >53,000 years for haplogroup U) (139), and therefore of the entire mitochondrial genealogy, is greater than the age of the anatomically modern European population (around 40,000 years). Because these lineages have nothing to do with Neandertal lineages (152), their common ancestors must have lived somewhere out of Europe, and they must have entered Europe at a later time. In addition, genealogies of different genes in the same individuals coalesce to common molecular ancestors who lived in different time periods, and hence the ages of these ancestral molecules have nothing to do with the ages of the ancestral individuals from whom the population is descended (56, 117). Unless a drastic founder effect occurred, the age of an allele predates, sometimes by a lot, the age of the population where that allele is found (9, 175). Despite past disagreement, most authors now acknowledge that there is no correlation between the timing of migrations and the age of mitochondrial or Y-chromosome clades (141), although the steps necessary to infer population processes from phylogeographic data are far from agreed.

In summary, no one can quantify with absolute confidence the contribution of the first Paleolithic settlers, or of later Neolithic immigrants, to the Europeans' genome. However, whereas a pre-Neolithic age of many gene genealogies is by no means incompatible with the NDD model, the broad southeast-northwest clines and the estimated high Neolithic admixture rates do not appear compatible with the PGE model. A prudent conclusion, until greater numbers of loci are studied, is that the PGE model is currently at odds with various aspects of European genetic diversity. That does not mean that populations did not shrink and re-expand in response to climate changes, but that these demographic changes, if they occurred, had a limited influence on current genome diversity. The direction of the main European clines is expected under the NDD model, but also under a model (197), compatible with computer simulation results (13), whereby the initial peopling of Europe left the main genetic mark at the continental scale, so that we now observe gradients established long before the last glacial maximum. The observation that mitochondrial clines are restricted to southern Europe for mtDNA may point to different patterns of gene flow in different regions, which is a possibility that should be explored. Other explanations for that include a different female and male mobility (156), or the effects of selective pressures on the mitochondrial genome (114).

GENETIC DATA ON PAST DEMOGRAPHIC CHANGES

The genes of contemporary people also contain information on the processes through which a population of a few thousand individuals reached the current size of several hundred million because changes of population size affect the shape of

gene genealogies. Expansions result in star-like genealogies, and most mutations occurring on those genealogies do not tend to be shared among lineages (50). The resulting plots of sequence differences between pairs of individuals, or mismatch distributions, are smooth and unimodal; their mode depends on the time passed since the expansion (82, 145). However, in stationary populations, the genealogical tree has a structure such that mutation sharing among lineages tends to be higher (146), and so the mismatch distributions show multiple peaks. Consequently, although the actual genealogies are impossible to reconstruct, comparisons of sequences in a population sample allow inferences about some properties of those genealogies, and in particular about past changes in population size.

At the world scale, most mitochondrial mismatch distributions show a single peak, in agreement with expansion expectations. The exceptions correspond to populations that do not produce food, and have been explained as a result of demographic crises in hunting-gathering communities after contacts with farming communities (61). The oldest expansions appear to be in Africa, then in Asia, then in the other continents. Estimated dates for Europe range from 56,000–58,000 years ago for some language minorities in the Alps, to less than 20,000 years ago for the Basques and Cornish (61).

The results are not equally straightforward for the Y chromosome. Two studies of microsatellites (127) and sequences (159) concluded that Y-chromosome diversity is more compatible with exponentially increasing, rather than with constant, population sizes, but they suggest a more recent population growth than mitochondrial data. Conversely, studies of SNPs found no evidence of growth because all distributions had multiple peaks, and statistics sensitive to demographic changes were insignificant (54, 123).

Those incongruences are partly due to methodological problems. Mismatch distributions and related statistics must be estimated for single populations. When individuals of populations that evolved separately are clumped, few mutations will likely be shared, and this generates results that suggest an expansion, even if no single population has expanded. Shen et al. (159) jointly analyzed individuals of different continents, so it is impossible to tell whether the expansion inferred in that study actually occurred or is a statistical artifact. Computer simulations show that expansions may escape detection if only highly variable SNP sites are typed, especially when these expansions are recent (54). This suggests that the effective population sizes of European males (N_m) remained low until recently, and then increased in a time sufficient for a significant accumulation of mutations at the microsatellite, but not the SNP, level. Because female population sizes (N_f) increased sharply in prehistoric times, this implies that Europeans, or their ancestors out of Europe, were essentially polygynous during much of their history (54).

Different effective population sizes for European males and females is the simplest, but not the only, explanation of the observed data. What one estimates in these studies is the product of the effective population size (N) times the migration rate (m). Therefore, another possibility is that the two sexes differed in their relative mobility (156), with females showing a greater tendency to migrate than males.

Polygyny and patrilocality are not mutually exclusive; either one might be true, for different areas and/or time periods. But it is clear that the mitochondrial genome and the Y chromosome, or perhaps all nuclear genes (83), show different patterns in all comparisons carried out so far (85, 93, 122, 154, 195). Unless deep differences in the mutational mechanisms or in generation times (86) account for these differences, one has to conclude that patrilocality, or polygyny, or any demographic factor causing departures from an equal variance of the reproductive success of males and females (52), has been common in Europe for long evolutionary times.

USING POPULATION-GENETIC KNOWLEDGE IN APPLIED RESEARCH

The Design of Genetic Association Studies

There is a contradiction in the research literature concerning the extent of genetic differentiation among populations. On one hand, the literature repeatedly emphasizes that different groups of humans, defined either racially or in terms of geographic ancestry, differ very little from one another. Feldman et al. (63) wrote recently ". . .that most genetic diversity occurs within groups and that very little is found between them." Similary, Schwartz (153) criticized racial profiling in biomedical research, arguing that differences among racial groups are largely or entirely cosmetic and do not reflect the underlying genetic makeup of individuals.

On the other hand, in many areas of biomedical research, failure to properly account for ethnic orgin of study subjects would preclude publication precisely because of the assumption that the groups would have meaningful differences in genetic makeup, and that these differences would compromise analyses of the relationship between genotypes and phenotypes. For example, association studies are broadly viewed as one of the most important contemporary tool for studying the genetic control of common diseases and variable drug response. The basic structure of such studies involves comparing the genetic makeup of individuals with different phenotypes, such as in the common case control design in which individuals with a certain disease would be compared to individuals that do not have that disease, or individuals that respond well to a drug are compared with those that do not. The idea is that if a gene variant predisposes to disease or response, it will be enriched in the population of individuals with disease, or with that tendency to respond to treatment.

However, the problem is that genetic differences among groups of people can create spurious associations by an effect called stratification. Imagine a case control study in which both patients and controls are drawn from two subpopulations and that the investigator fails to take account of this substructure. Now assume further that the disease has higher incidence in group A compared with group B. In this situation, any polymorphism that differs in frequency between group A and group B will be associated with the disease, and will thus appear to cause the disease.

But the polymorphisms are only correlated because of the underlying population structure.

The research community has quietly invested tremendous energy in overcoming this difficulty. Early efforts focused on alternative study designs in which family members are used as controls to ensure that chromosomes from the same subpopulations are compared. Recently, attention returned to the case control design, but using genetic methods to detect stratification and correct for it (129, 133). The concerns are greatest when the allele frequency differences are greatest. If there was stratification involving individuals of African and European ancestry, for example, most polymorphisms in the genome could show significant differences for any disease that differs in incidence between Africa and Europe (47).

Even in Europe, where allele frequency differences are slight, the same concerns remain. There are probably no leading journals that would accept a genetic association study carried out on Europeans without other indications of their provenance. And even within specific parts of Europe (for example, northern Europe), it is clear that there are gradations of allele frequency that could lead to spurious association. The main question is the magnitude of the effect. In this context, it would be incorrect to say that geographic ancestry is irrelevant to study design. So what level of ethnic matching of individuals with different geographic ancestry is required in Europe? The current convention in the literature is to match by political boundaries, which makes little sense. But there is no way to objectively decide what level of matching is necessary. On average, allele frequency differences are slight, but significant. Many markers show clear evidence of clinality, as argued above. But the differences, even from, say, south Italy to West Britain, are modest for typical markers. For example, looking at only 18 markers from south Italy and west Britain, the average allele frequency difference was under 5%, and there was little evidence for differentiation between the populations (D. Goldstein, unpublished data).

This means that the level of stratification possible in European populations could not drive large allele frequency differences for typical loci between cases and controls. One of the most troublesome features of stratification is that its importance increases with increases in sample size. Thus, if one looks for variants that have only modest effects on risk then it is necessary to use very large sample sizes. In this context, even the modest allele frequency differences in Europe could interefere with the detection of the real signal, and could create spurious associations.

Therefore, to estimate the extent of the problem in Europe, it will be necessary to assess patterns of geographic variation in very large samples and for large numbers of markers. The patterns observed in controls could be used to estimate expected levels of stratification as a function of prevalence differences, or assessments could be made more directly by looking at unlinked (and nonfunctional) markers in large case control comparisons.

This will indicate what sorts of approaches are most suitable for dealing with stratification. One approach would be to use genetic methods of the kind described

above to assess associations within each of the groups, e.g., structured association (129) or the simpler approach of genomic control (47, 133). Genomic control would entail a significant loss of power in the face of modest levels of stratification and variants of low to moderate effect. In this case, the allele frequency differences expected for the causal variant and for unlinked variants influenced by stratification are similar, and more sophisticated methods such as structured association are required to pick up the real signal. But if stratification effects are more modest, or the variants under study have a large effect, genome control will likely be sufficient.

Founder Populations in Europe

Another way to see that Europe is not a single homogenous genetic population is in terms of the distribution of rare genetic diseases. Many European groups have distinctive constellations of mutations causing Mendelian disease, most famously Ashkenazi Jews (119) and the Finns (94). Ostrer (119) lists 40 Mendelian diseases in Ashkenazi Jews that are distinctive in having frequencies that differ from other European populations, and Kere (94) lists a different constellation of 35 Mendelian-diseases that have a significantly greater prevalence in Finland than elsewhere.

These observations have led to the suggestion that the Finns and the Ashkenzi Jews are founder populations that recently grew from relatively small sizes. In this situation a rare deleterious allele may increase to unusually high frequency, as dating analyses have suggested for some, but not all, of the underlying mutations (73, 144). When the deleterious allele is young it is more likely to reside on an extended haplotypic background, which can facilitate population-based fine mapping of disease genes. Many of the genes discovered in both populations have been positionally cloned with the help of linkage disequilibrium (LD) mapping relying on these extended haplotypes.

Partially because of this experience with rare variation causing rare Mendelian diseases, there have been many suggestions that the same populations, and some other putative founder populations, are more generally different from other European populations, and thus would also be useful for mapping genetically complex traits such as common diseases or drug responses. The idea is that such founder populations will be charaterized by (*a*) less genetic diversity (i.e., genetically complex traits will be a little less complex), and (*b*) greater LD because of the elevated drift (i.e., few genetic markers would be necessary to represent the variation in the genome).

If rare variants were largely responsible for the genetic component of complex diseases, the first claim, of reduced heterogeneity, would seem supportable based on the experience of Mendelian disease. However, it is much less clear that common variants would be similary affected by whatever demographic events affected Finnish, Ashkenazi, and other founder populations. For example, this is clear in the observations that the so-called founder populations have similar levels of genetic diversity as other European populations using standard measures of diversity (3, 181).

Genetic association studies will also depend heavily on the pattern of association among variants, or the pattern of LD. Many have argued that founder populations in Europe will have significantly more long-range LD than other populations. One inplication of this expected LD level is that fewer markers would be required to represent variation in these populations than in nonfounder populations. For example, Shifman & Darvasi (160) studied markers more than 200 kb apart in Ashkenazi Jews and in a set of unrelated individuals of North European ancestry available from the Centre pour l'Etude des Polymorphismes Humaines (CEPH), and found a significant excess of LD in the Ashkenazi Jews.

In the context of common disease studies, Iceland is the founder population generating the most attention, thanks to the activities of the commercial firm deCODE. In various publications and in the popular press, scientists associated with deCODE have argued that Iceland has reduced genetic diversity and that it is therefore more useful for gene mapping studies than other European populations. However, this claim has met with serious criticism, including recent claims that some evidence for the homogeneity of Icelanders resulted from errors in the mtDNA database and that, using some measures of diversity, Icelanders actually rank among the most diverse European populations (3). In a rigourous and thoughtful response, Helgason et al. (86) used more sophisticated methods to demonstrate that Icelanders do show evidence of greater drift effects than most other European populations. One must appreciate, however, that evidence of a slightly different demographic history does not translate into important differences in terms of the properties of population for genetic mapping. Without knowing whether the relevant genetic variation that is sought is common or rare, it is impossible to know whether the detectable differences between Iceland and other populations are significant in the context of genetic mapping. To date, we are aware of no strong evidence that any European populations differ substantially in terms of the common variation they carry.

Simarily, detectable differences in LD do not mean that comprehensive genetic mapping would require significantly different numbers of genetic markers in so-called founder populations and in "ordinary" European populations. Methods for designing sets of genetic markers sufficient to represent common variation have received considerable attention lately, inspired in part by the demonstration of relatively large tracts of limited haplotype diversity, or "blocks" of linkage disequilibrium (43). Polymorphisms selected to represent other polymorphisms are often called tagging Single Nucleotide Polymorphisms (tSNPs). There are several different approaches available now for both selecting tSNPs and for evaluating how well they represent variation at other SNPs. This work is directly relevant to the question of how many SNPs are required to represent genome variation in a given population, and will ultimately reveal whether different European populations are significantly different from one another in this regard (71). However, the general impression from these studies to date is that there are important differences in patterns of LD between African and non African populations that translate into differences in the number of SNPs that would be required for comprehensive,

LD-based genetic association studies. The differences between so-called founder populations and other European populations has not yet been systematically studied in this framework. Nevertheless, the data that are available argue against any important differences emerging in Europe. Focusing both on the distribution of haplotypes (68) or on multimarker measures of association, the number of SNPs required is similar in non-African populations from both Europe and Asia. This strongly suggests that when these more comprehensive approaches are applied in Europe, the detectable differences in LD will not translate into important differences in the number of markers required.

FINAL REMARKS AND FUTURE PROSPECTS

Is there anything we can call a distinctive European gene pool, or is Europe better described as an assemblage of different gene pools? And, should the latter prove true, what are these groups and what are their boundaries? As noted above, there is some reason to consider the Europeans as a heterogeneous group formed by people of African and Asian descent. However, with the increasing availability of large data sets of genetic polymorphisms, it will be possible to more precisely define what "Europe" means in genetic terms. The obvious approach is to identify clusters in the genomic data to see whether they can be associated with a specific geographical area, and to test for the significance of their differences.

Before embarking into that exercise, a preliminary and time-honored problem should be addressed: the identification of outliers. The idea that unusually high or low F_{ST} values at certain loci point to specific selective pressures dates back to the sixties (26), although a method to separate locus-specific effects (mainly related to selection and mutation) from genome-wide effects (mainly due to drift and gene flow) has not yet been established (104, 115). However, the inclusion of loci reflecting adaptation leads to erroneous reconstructions of population histories (193). The already existing, efficient algorithms to predict levels of genetic variation under neutrality (e.g., 14, 15), and the foreseeable availability of adequate data sets in the near future (4), suggest that we shall soon have reliable ways to identify selection at the genomic level. However, specific studies at the single locus level are always necessary if one is to understand the details and the dynamics of the adaptation process.

Once the loci subjected to selection have been recognized, it will be possible to focus on the genome-wide effects that contain information about demographic history. Future European population genomic studies may better address a large number of questions that are currently unanswered. Many of them are listed above. Others regard the extent and nature of the evolutionary interactions between humans and domesticated species, leading to two classes of consequences: (*a*) the coevolution of certain traits (16), and (*b*) the parallel evolution of different traits in humans and domesticated animals due to their common migratory history (91).

Parallel analyses of gene diversity, and of gene expression diversity (57, 70), in humans and related species may pave the path for a clearer reconstruction of our evolutionary and demographic history, and eventually lead us to a deeper understanding of the genetic bases of disease susceptibility and response to drugs.

ACKNOWLEDGMENTS

We thank Svante Pääbo, David Serre, Peter Forster, and Ewa Zietkiewicz for providing us with unpublished results, or with papers that otherwise would have been difficult to obtain. In the development of this manuscript, G.B. was supported in part by the FISR fund of the Italian Ministry for Universities (MIUR). This paper is dedicated to the memory of an unforgettable friend, Ryk Ward (1943–2003).

The *Annual Review of Genomics and Human Genetics* is online at http://genom.annualreviews.org

LITERATURE CITED

1. Ammerman AJ, Cavalli-Sforza LL. 1984. *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton, NJ: Princeton Univ. Press
2. Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, et al. 1996. Minisatellite diversity supports a recent African origin for modern humans. *Nat. Genet.* 13:154–60
3. Arnason E. 2003. Genetic heterogeneity of Icelanders. *Ann. Hum. Genet.* 67:5–16
4. Bamshad M, Wooding SP. 2003. Signatures of natural selection in the human genome. *Nat. Rev. Genet.* 4:99–111
5. Bandelt H-J, Macaulay V, Richards M. 2003. What molecules can't tell us about the spread of languages and the Neolithic. In *Examining the Farming/Language Dispersal Hypothesis*, ed. P Bellwood, C Renfrew. pp 99–107. Cambridge, UK: McDonald Inst. Archaeol. Res.
6. Barbujani G. 2001. What genetics tells us about races. In *International Encyclopedia of the Social and Behavioral Sciences*, ed. NJ Smelser, PB Baltes, 19:12694–700. Oxford: Elsevier
7. Barbujani G, Bertorelle G. 2001. Genetics and the population history of Europe. *Proc. Natl. Acad. Sci. USA* 98:22–25
8. Barbujani G, Bertorelle G. 2003. Were Cro-Magnons too like us for DNA to tell? *Nature* 424:127
9. Barbujani G, Bertorelle G, Chikhi L. 1998. Evidence for Paleolithic and Neolithic gene flow in Europe. *Am. J. Hum. Genet.* 62:488–92
10. Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL. 1997. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* 94:4516–19
11. Barbujani G, Pilastro A, De Domenico S, Renfrew C. 1994. Genetic variation in North Africa and Eurasia: Neolithic demic diffusion versus Paleolithic colonisation. *Am. J. Phys. Anthropol.* 95:137–54
12. Barbujani G, Sokal RR. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. USA* 87:1816–19
13. Barbujani G, Sokal RR, Oden NL. 1995. Indo-European origins: a computer simulation test of five hypotheses. *Am. J. Phys. Anthropol.* 96:109–32
14. Beaumont MA, Zhang W, Balding DJ.

2002. Approximate Bayesian computation in population genetics. *Genetics* 162:2025–35
15. Beaumont MS, Nichols RA. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. London B* 263:1619–26
16. Beja-Pereira A, Luikart G, England PR, Bradley DG, Jann OC, et al. 2003. Gene-culture coevolution between cattle milk protein genes and human lactase genes. *Nat. Genet.* 35:311–13
17. Bertorelle G, Excoffier L. 1998. Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15:1298–311
18. Bertranpetit J. 2000. Genome, diversity, and origins: the Y chromosome as a storyteller. *Proc. Natl. Acad. Sci. USA* 97:6927–29
19. Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, et al. 2001. High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. *Am. J. Hum. Genet.* 68:1019–29
20. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, et al. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–57
21. Calafell F. 2003. Classifying humans. *Nat. Genet.* 33:435–36
22. Calafell F, Bertranpetit J. 1994. Principal component analysis of gene frequencies and the origin of Basques. *Am. J. Phys. Anthropol.* 93:201–15
23. Caramelli D, Lalueza-Fox C, Vernesi C, Lari M, Casoli A, et al. 2003. Evidence for a genetic discontinuity between Neandertals and 24,000-years-old anatomically modern Europeans. *Proc. Natl. Acad. Sci. USA* 100:6593–97
24. Carlson CS, Eberle MA, Rieder MJ, Smith JD, Kruglyak L, et al. 2003 Additional SNPs and linkage-disequilibrium analyses are necessary for whole-genome association studies in humans. *Nat. Genet.* 33:518–21
25. Casalotti R, Simoni L, Belledi M, Barbujani G. 1999. Y-chromosome polymorphism and the origins of the European gene pool. *Proc. R. Soc. London B* 266:1959–65
26. Cavalli-Sforza LL. 1966. Population structure and human evolution. *Proc. R. Soc. London B* 164:362–79
27. Cavalli-Sforza LL, Feldman MW. 2003. The application of molecular genetic approaches to the study of human evolution. *Nat. Genet.* 33(Suppl):266–75
28. Cavalli-Sforza LL, Menozzi P, Piazza A. 1993. Demic expansions and human evolution. *Science* 259:639–46
29. Cavalli-Sforza LL, Menozzi P, Piazza A. 1994. *The History and Geography of Human Genes.* Princeton, NJ: Princeton Univ. Press
30. Cavalli-Sforza LL, Minch E. 1997. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 61:247–51
31. Chakraborty R. 1986. Gene admixture in human populations: models and predictions. *Yearb. Phys. Anthropol.* 29:1–43
32. Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. 1995. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am. J. Hum. Genet.* 57:133–49
33. Chikhi L, Nichols RA, Barbujani G, Beaumont MA. 2002. Y genetic data support the Neolithic demic diffusion model. *Proc. Natl. Acad. Sci. USA* 99: 11008–13
34. Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G. 1998. Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc. Natl. Acad. Sci. USA* 95:9053–58
35. Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, et al. 1998. Clinal

variation in the nuclear DNA of Europeans. *Hum. Biol.* 70:643–57
36. Ciofi C, Milinkovitch MC, Gibbs JP, Caccone A, Powell JR. 2002. Microsatellite analysis of genetic divergence among populations of giant Galapagos tortoises. *Mol. Ecol.* 11:2265–83
37. Cohen C. 1991. Les races humaines en histoire des sciences. In *Aux Origines d'Homo sapiens*, ed. JJ Hublin, AM Tillier, pp. 9–56. Paris: Presses Universitaires de France
38. Comas D, Calafell F, Bemchemsi N, Helal A, Lefranc G, et al. 2000. Alu insertion polymorphisms in NW Africa and the Iberian peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum. Genet.* 107:312–19
39. Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, et al. 1997. Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.* 99:443–49
40. Cooper RS, Kaufman JS, Ward R. 2003. Race and genomics. *New Engl. J. Med.* 348:1166–70
41. Crow JF. 1958. Some possibilities for measuring selection intensities in man. *Hum. Biol.* 30:1–13
42. Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, et al. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* 70:1197–214
43. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. 2001. High-resolution haplotypes structure in the human genome. *Nat. Genet.* 29:229–32
44. Dansgaard W, Johnsen SJ, Clausen HB, Dahl-Jensen D, Gundestrup NS, et al. 1993. Evidence for general instability of past climate from a 250-kyr ice-core record. *Nature* 364:218–20
45. D'Errico F, Sanchez-Goñi MF. 2003. Neandertal extinction and the millennial scale climatic variability of OIS 3. *Q. Sci. Rev.* 22:769–88
46. Destro-Bisol G, Maviglia R, Caglia A, Boschi I, Spedini G, et al. 1999. Estimating European admixture in African Americans by using microsatellites and a microsatellite haplotype (CD4/Alu). *Hum. Genet.* 104:149–57
47. Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* 55:997–1004
48. Di Benedetto G, Nasidze IS, Stenico M, Nigro L, Krings M, et al. 2000. Mitochondrial DNA sequences in prehistoric human remains from the Alps. *Eur. J. Hum. Genet.* 8:669–77
49. Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Corbo RM, et al. 2003. Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Mol. Phylogenet. Evol.* 28:387–95
50. Donnelly P. 1996. Interpreting genetic variability: the effects of shared evolutionary history. In *Variation in the Human Genome*, ed. K Weiss, pp 25–50. Chichester, UK: John Wiley
51. Driscoll CA, Menotti-Raymond M, Nelson G, Goldstein D, O'Brien SJ. 2002. Genomic microsatellites as evolutionary chronometers: a test in wild cats. *Genome Res.* 12:414–23
52. Dupanloup I, Bertorelle G. 2001. Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* 18:672–75
53. Dupanloup I, Chikhi G, Bertorelle G, Barbujani G. 2004. Estimating the impact of prehistoric admixture on the Europeans' genome. *Mol. Biol. Evol.* 21:In press
54. Dupanloup I, Pereira L, Bertorelle G, Calafell F, Prata MJ, et al. 2003. A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J. Mol. Evol.* 57:85–97

55. Dupanloup I, Schneider S, Langaney A, Excoffier L. 2000. Inferring the impact of linguistic boundaries on population differentiation: application to the Afro-Asiatic-Indo-European case. *Eur. J. Hum. Genet.* 8:750–56
56. Edwards SV, Beerli P. 2000. Gene divergence, population divergence, and the variance in coalescence times in phylogeographic studies. *Evolution* 54:1839–54
57. Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, et al. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* 296:340–43
58. Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, et al. 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30:233–37
59. Excoffier L. 2002. Human demographic history: refining the recent African origin model. *Curr. Opin. Genet. Dev.* 12:675–82
60. Excoffier L. 2003. Human diversity: our genes tell where we live. *Curr. Biol.* 13:R134–36
61. Excoffier L, Schneider S. 1999. Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc. Natl. Acad. Sci. USA* 96:10597–60
62. Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–91
63. Feldman MW, Lewontin RC, King MC. 2003. Race: a genetic melting pot. *Nature* 424:374
64. Fevolden SE, Scheppenheim S. 1989. Genetic homogeneity of krill (*Euphausia superba* Dana) in the Southern Ocean. *Polar Biol.* 9:533–39
65. Fix AG. 1996. Gene frequency clines in Europe: demic diffusion or natural selection *? J. Royal Anthropol. Inst.* 2:625–43
66. Foley R. 1998. The context of human genetic evolution. *Genome Res.* 8:339–47
67. Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, et al. 2000. A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* 67:182–96
68. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, et al. 2002. The structure of haplotype blocks in the human genome. *Science* 296:2225–29
69. Garte S, Gaspari L, Alexandrie AK, Ambrosone C, Autrup H, et al. 2001. Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol. Biomark. Prev.* 10:1239–48
70. Gilad Y, Bustamante CD, Lancet D, Paabo S. 2003. Natural selection on the olfactory receptor gene family in humans and chimpanzees. *Am. J. Hum. Genet.* 73:489–501
71. Goldstein DB, Ahmadi KR, Weale ME, Wood NW. 2003. Genome scans and candidate gene approaches in the study of common diseases and variable drug responses. *Trends Genet.* 19:615–22
72. Goldstein DB, Chikhi L. 2002. Human migration and population structure: what we know and why it matters. *Annu. Rev. Genomics Hum. Genet.* 3:129–52
73. Goldstein DB, Reich DE, Bradman N, Ulsher S, Seligsohn U, Peretz H. 1999. Age estimates for two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. *Am. J. Hum. Genet.* 64: 1071–75
74. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. *Proc. Natl. Acad. Sci. USA* 92:6723–27
75. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139:463–71
76. Gonzalez-Burchard E, Ziv E, Coyle N,

Gomez SL, Tang H, et al. 2003. The importance of race and ethnic background in biomedical research and clinical practice. *New Engl. J. Med.* 348:1170–75
77. Goodacre SL. 2002. Population structure, history and gene flow in a group of closely related land snails: genetic variation in Partula from the Society Islands of the Pacific. *Mol. Ecol.* 11:55–68
78. Gutierrez G, Sanchez D, Marin A. 2002. A reanalysis of the ancient mitochondrial DNA sequences recovered from Neandertal bones. *Mol. Biol. Evol.* 19:1359–66
79. Haldane JBS. 1940. The blood-group frequencies of European peoples and racial origins. *Hum. Biol.* 12:457–80
80. Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, et al. 1994. Molecular genetic analyses of the Tyrolean Ice Man. *Science* 264:1775–78
81. Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, et al. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* 60:772–89
82. Harpending HC, Batzer MA, Gurven M, Jorde JB, Rogers AR, et al. 1998. Genetic traces of ancient demography. *Proc. Natl. Acad. Sci. USA* 95:1961–67
83. Harris EE, Hey J. 1999. Human demography in the Pleistocene: Do mitochondrial and nuclear genes tell the same story? *Evol. Anthropol.* 8:81–86
84. Hawks J, Hunley K, Lee S, Wolpoff M. 2000. Population bottlenecks and Pleistocene human evolution. *Mol. Biol. Evol.* 17:2–22
85. Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefansson K. 2003. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am. J. Hum. Genet.* 72:1370–88
86. Helgason A, Nicholson G, Stefansson K, Donnelly P. 2003. A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann. Hum. Genet.* 67:281–97
87. Hofreiter H, Serre D, Poinar HN, Kuch M, Pääbo S. 2001. Ancient DNA. *Nat. Rev. Genet.* 2:353–59
88. Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, et al. 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia suggested by analysis of a Y-chromosomal polymorphism. *Am. J. Hum. Genet.* 65:1437–48
89. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. *Nature* 408:708–13
90. Izagirre N, de la Rua C. 1999. An mtDNA analysis in ancient Basque population: implications for haplogroup V as a marker for a major Paleolithic expansion from Southwestern Europe. *Am. J. Hum. Genet.* 65:199–207
91. Jansen T, Forster P, Levine MA, Oelke H, Hurles M, et al. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc. Natl. Acad. Sci. USA* 99:10905–10
92. Jorde, LB, Watkins WS, Bamshad MJ, Dixon ME, Ricker CE, et al. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal and Y-chromosome data. *Am. J. Hum. Genet.* 66:979–88
93. Kalaydjieva L, Calafell F, Jobling MA, Angelicheva D, de Knijff P, et al. 2001. Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur. J. Hum. Genet.* 9:97–104
94. Kere J. 2001. Human population genetics: lessons from Finland. *Annu. Rev. Genomics Hum. Genet.* 2:103–28
95. King MC, Motulsky AG. 2002. Human genetics. Mapping human history. *Science* 298:2342–43
96. Krausz C, Quintana-Murci L, Rajpert-De Meyts E, Jorgensen N, Jobling MA,

et al. 2001. Identification of a Y chromosome haplogroup associated with reduced sperm counts. *Hum. Mol. Genet.* 10:1873–77

97. Krings M, Capelli C, Tschentscher F, Geisert H, Meyer S, et al. 2000. A view of Neandertal genetic diversity. *Nat. Genet.* 26:144–46
98. Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, et al. 1997. Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19–30
99. Lahr MM. 1996. *The Evolution of Human Diversity.* Cambridge, UK: Cambridge Univ. Press
100. Lahr MM, Foley RA, Pinhasi R. 2000. Expected regional patterns of Mesolithic-Neolithic human population admixture in Europe based on archaeological evidence. In *Archaeogenetics: DNA and the Population Prehistory of Europe*, ed. C Renfrew, K Boyle, pp. 81–88. Cambridge, UK: McDonald Inst. Monogr.
101. Larsen LA, Vuust J, Nystad M, Evseeva I, VanGhelue M, et al. 2001. Analysis of FMR1 (CGG)(n) alleles and DXS548-FRAXAC1 haplotypes in the European circumpolar populations: traces of genetic relationship with Asia. *Eur. J. Hum. Genet.* 9:724–27
102. Latter BDH. 1980. Genetic differences within and between populations of the major human subgroups. *Am. Nat.* 116:220–37
103. Lewontin RC. 1972. The apportionment of human diversity. *Evol. Biol.* 6:381–98
104. Lewontin RC, Krakauer J. 1973. Distribution of gene frequency as a test of the theory of selective neutrality of polymorphisms. *Genetics* 74:175–95
105. Li WH, Sadler LA. 1991. Low nucleotide diversity in man. *Genetics* 129:513–23
106. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4:981–94
107. Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, et al. 1998. Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. *Am. J. Hum. Genet.* 63:847–60
108. Manni F, Leonardi P, Barakat A, Rouba H, Heyer E, et al. 2002. Y-chromosome analysis in Egypt suggests a genetic regional continuity in Northeastern Africa. *Hum. Biol.* 74:645–58
109. Meinila M, Finnila S, Majamaa K. 2001. Evidence for mtDNA Admixture between the Finns and the Saami. *Hum. Hered.* 52:160–70
110. Mellars PA. 1992. Archaeology and the population dispersal hypothesis of modern human origins in Europe. *Phil. Trans. R. Soc. London B* 337:225–34
111. Mellars PA. 1998. The impact of climatic changes on the demography of late Neandertal and early anatomically modern populations in Europe. In *Neandertals and Modern Humans in Western Asia*, ed. T Akazawa, K Aoki, O Bar-Yosef, pp. 493–507. New York: Plenum Press
112. Menozzi P, Piazza A, Cavalli-Sforza L. 1978. Synthetic maps of human gene frequencies in Europeans. *Science* 201: 786–92
113. Miller-Butterworth CM, Jacobs DS, Harley EH. 2003. Strong population substructure is correlated with morphology and ecology in a migratory bat. *Nature* 424:187–91
114. Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, et al. 2003. Natural selection shaped regional mtDNA variation in humans. *Proc. Natl. Acad. Sci. USA* 100:171–76
115. Nei M, Chakravarti A. 1977. Drift variances of Fst and Gst statistics obtained from a finite number of isolated populations. *Theor. Popul. Biol.* 11:307–25
116. Nei M, Takezaki N. 1996. The root of the

phylogenetic tree of human populations. *Mol. Biol. Evol.* 13:170–77
117. Nichols R 2001. Gene trees and species trees are not the same. *Trends Ecol. Evol.* 16:358–64
118. Nordborg M. 1998. On the probability of Neanderthal ancestry. *Am. J. Hum. Genet.* 63:1237–40
119. Ostrer H. 2001. A genetic profile of contemporary Jewish populations. *Nat. Rev. Genet.* 2:891–98
120. Otte M. 2000. The history of European populations as seen by archaeology. In *Archaeogenetics: DNA and the population prehistory of Europe*, ed. C Renfrew, K Boyle, pp. 41–44. Cambridge, UK: McDonald Inst. Monogr.
121. Ovchinnikov IV, Gotherstrom A, Romanova GP, Kharitonov VM, Liden K, et al. 2000. Molecular analysis of Neanderthal DNA from the northern Caucasus. *Nature* 404:490–93
122. Passarino G, Cavalleri GL, Lin AA, Cavalli-Sforza LL, Borresen-Dale AL, et al. 2002. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *Eur. J. Hum. Genet.* 10:521–29
123. Pereira L, Dupanloup I, Rosser ZH, Jobling MA, Barbujani G. 2001. Y-chromosome mismatch distributions in Europe. *Mol. Biol. Evol.* 18:1259–71
124. Pinhasi R, Foley RA, Mirazòn-Lahr M. 2000. Spatial and temporal patterns in the Mesolithic-Neolithic archaeological record of Europe. In *Archaeogenetics: DNA and the Population Prehistory of Europe*, ed. C Renfrew, K Boyle, pp. 45–56. Cambridge, UK: McDonald Inst. Monogr.
125. Piontkivska H, Nei M. 2003. Birth-and-death evolution in Primate MHC Class I genes: divergence time estimates. *Mol. Biol. Evol.* 20:601–9
126. Plaza S, Calafell F, Helal A, Bouzerna N, Lefranc G, et al. 2003. Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann. Hum. Genet.* 67: 312–28
127. Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW. 1999. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–98
128. Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59
129. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. 2000. Association mapping in structured populations. *Am. J. Hum. Genet.* 67:170–81
130. Quintana-Murci L, Veitia R, Fellous M, Semino O, Poloni ES. 2003. Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. *Am. J. Phys. Anthropol.* 121:157–71
131. Rando JC, Pinto F, Gonzalez AM, Hernandez M, Larruga JM, et al. 1998. Mitochondrial DNA analysis of northwest African populations reveals genetic exchanges with European, near-eastern and sub-Saharan populations. *Ann. Hum. Genet.* 62:531–50
132. Rannala B, Mountain JL. 1997. Detecting immigration by using multilocus genotypes. *Proc. Natl. Acad. Sci. USA* 94:9197–201
133. Reich DE, Goldstein DB. 2001. Detecting association in a case-control study while correcting for population stratification. *Genet. Epidemiol.* 20:4–16
134. Relethford J. 2001. Absence of regional affinities of Neandertal DNA with living humans does not reject multiregional evolution. *Am. J. Phys. Anthropol.* 115:95–98
135. Rendine S, Piazza A, Cavalli-Sforza LL. 1986. Simulation and separation by principal components of multiple demic expansions in Europe. *Am. Nat.* 128:681–706
136. Renfrew C. 1987. *Archaeology and*

Language: The Puzzle of Indo-European Origins. London: Jonathan Cape
137. Richards M, Corte-Real H, Forster P, Macaulay VA, Wilkinson-Herbots H, et al. 1996. Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 59:185–203
138. Richards M, Macaulay VA, Bandelt HJ, Sykes BC. 1998. Phylogeography of mitochondrial DNA in Western Europe. *Ann. Hum. Genet.* 62:241–60
139. Richards M, Macaulay VA, Hickey E, Vega E, Sykes BC, et al. 2000. Tracing European founder lineages in the Near East mtDNA pool. *Am. J. Hum. Genet.* 67:1251–76
140. Richards M, Macaulay VA, Sykes BC, Pettit P, Hedges R, et al. 1997. Reply to Cavalli-Sforza and Minch. *Am. J. Hum. Genet.* 61:251–54
141. Richards M, Macaulay VA, Torroni A, Bandelt HJ. 2002. In search of geographical patterns in European mitochondrial DNA. *Am. J. Hum. Genet.* 71:1168–74
142. Ripley BD. 1996. *Pattern Recognition and Neural Networks.* Cambridge, UK: Cambridge Univ. Press
143. Risch N, Burchard E, Ziv E, Tang H. 2002. Categorization of humans in biomedical research: genes, race and disease. *Genome Biol.* 3:1–11
144. Risch N, de Leon D, Ozelius L, Kramer P, Almasy L, et al. 1995. Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat. Genet.* 9:152–59
145. Rogers AR, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9:552–69
146. Rogers AR, Jorde LB. 1995. Genetic evidence on modern human origins. *Hum. Biol.* 67:1–36
147. Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, et al. 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Res.* 12:602–12
148. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, et al. 2002. Genetic structure of human populations. *Science* 298:2381–85
149. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, et al. 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* 67:1526–43
150. Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, et al. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res.* 5:42–52
151. Satta Y, Takahata N. 2002. Out of Africa with regional interbreeding? Modern human origins. *BioEssays* 24:871–75
152. Schmitz RW, Serre D, Bonani G, Feine S, Hillgruber F, et al. 2002. The Neandertal type site revisited: interdisciplinary investigations of skeletal remains from the Neander Valley, Germany. *Proc. Natl. Acad. Sci. USA* 99:13342–47
153. Schwartz RS. 2001. Racial profiling in medical research. *N. Engl. J. Med.* 344:1392–93
154. Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, et al. 1997. Differential structuring of human populations for homologous X and Y microsatellite loci. *Am. J. Hum. Genet.* 61:719–33
155. Scozzari R, Cruciani F, Pangrazio A, Santolamazza P, Vona G, et al. 2001. Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. *Hum. Immunol.* 62:871–84
156. Seielstad M, Minch E, T Cavalli-Sforza LL. 1997. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* 20:278–80
157. Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS. 1996. A

view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* 59:964–68
158. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, et al. 2000. The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–59
159. Shen P, Wang F, Underhill PA, Franco C, Yang WH, et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* 97:7354–59
160. Shifman S, Darvasi A. 2001. The value of isolated populations. *Nat. Genet.* 28: 309–10
161. Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G. 2000. Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* 66:262–78
162. Simoni L, Gueresi P, Pettener D, Barbujani G. 1999. Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum. Biol.* 71:399–415
163. Sokal RR. 1988. Genetic, geographic and linguistic distances in Europe . *Proc. Natl. Acad. Sci. USA* 85:1722–26
164. Sokal RR. 1991. Ancient movement patterns determine modern genetic variances in Europe. *Hum. Biol.* 63:589–606
165. Sokal RR, Harding RM, Oden NL. 1989. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthropol.* 80:267–94
166. Sokal RR, Menozzi P. 1982. Spatial autocorrelation of HLA frequencies in Europe supports demic diffusion of early farmers. *Am. Nat.* 119:1–17
167. Sokal RR, Oden NL, Rosenberg MS, Thomson BA. 2000. Cancer incidences in Europe related to mortalities, and ethnohistoric, genetic, and geographic distances. *Proc. Natl. Acad. Sci. USA* 97: 6067–72
168. Sokal RR, Oden NL, Thomson BA. 1988. Genetic change across language boundaries in Europe *Am. J. Phys. Anthropol.* 76:337–61
169. Sokal RR, Oden NL, Walker J, Di Giovanni D, Thomson BA. 1996. Historical population movements in Europe influence genetic relationships in modern samples. *Hum. Biol.* 68:873–98
170. Sokal RR, Oden NL, Wilson C. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143–45
171. Stead JD, Jeffreys AJ. 2002. Structural analysis of insulin minisatellite alleles reveals unusually large differences in diversity between Africans and non-Africans. *Am. J. Hum. Genet.* 71:1273–84
172. Stefan M, Stefanescu G, Gavrila L, Terrenato L, Jobling MA, et al. 2001. Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region. *Eur. J. Hum. Genet.* 9:27–33
173. Stringer C. 2002. Modern human origins: progress and prospects . *Philos. Trans. R. Soc. London B* 357:563–79
174. Stringer C, Davies W. 2001. Those elusive Neanderthals. *Nature* 413:791–92
175. Stumpf MP, Goldstein DB. 2001. Genealogical and evolutionary inference with the human Y chromosome. *Science* 291:1738–42
176. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF. 1998. Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.* 7:453–64
177. Takahata N, Satta Y. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* 94:4811–15
178. Tattersall I, Schwartz J. 2000. *Extinct Humans*. Boulder, CO: Westview Press
179. Templeton A 1999. Human races: a genetic and evolutionary perspective. *Am. Anthropol.* 100:632–50
180. Templeton A. 2002. Out of Africa again and again. *Nature* 416:45–51

181. Thomas MG, Weale ME, Jones AL, Richards M, Smith A, et al. 2002. Founding mothers of Jewish communities: Geographically separated Jewish groups were independently founded by very few female ancestors. *Am. J. Hum. Genet.* 70:1411–20
182. Thorne A, Wolpoff M. 1992. The multiregional evolution of modern humans. *Sci. Am.* 266:76–83
183. Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, et al. 1996. Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. *Science* 271:1380–87
184. Tishkoff SA, Verrelli BC. 2003. Patterns of human genetic diversity: implications for human evolutionary history and disease. *Annu. Rev. Genomics Hum. Genet.* 4:293–340
185. Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, et al. 1998. MtDNA analysis reveals a major Late Paleolithic population expansion from Southwestern to Northeastern Europe. *Am. J. Hum. Genet.* 62:1137–52
186. Torroni A, Bandelt HJ, Macaulay V, Richards M, Cruciani F, et al. 2001. A signal, from human mtDNA, of postglacial recolonization in Europe. *Am. J. Hum. Genet.* 69:844–52
187. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, et al. 2000. Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26:358–61
188. Vekua A, Lordkipanidze D, Rightmire GP, Agusti J, Ferring R, et al. 2002. A new skull of early Homo from Dmanisi, Georgia. *Science* 297:85–89
189. Vila C, Amorim IR, Leonard JA, Posada D, Castroviejo J, et al. 1999. Mitochondrial DNA phylogeography and population history of the grey wolf *canis lupus*. *Mol. Ecol.* 8:2089–103
190. von Haeseler A, Sajantila A, Paabo S. 1996. The genetical archaeology of the human genome. *Nat. Genet.* 14:135–40
191. Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, et al. 2003. Genetic variation among world populations: inferences from 100 Alu insertion polymorphisms. *Genome Res.* 13:1607–18
192. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, et al. 2001. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl. Acad. Sci. USA* 98:10244–49
193. Wilding CS, Butlin RK, Grahame J. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatalis* detected using AFLP markers. *J. Evol. Biol.* 14:611–19
194. Willis KJ, Whittaker RJ. 2000. Paleoecology. The refugial debate. *Science* 287:1406–7
195. Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, et al. 2001. Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci. USA* 98:5078–83
196. Wilson JF, Weale ME, Smith AC, Gratrix F, Fletcher B, et al. 2001. Population genetic structure of variable drug response. *Nat. Genet.* 29:265–69
196a. Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420
197. Xiao FX, Yotova V, Zietkiewicz E, Lovell A, Gehl D, et al. 2004. Human X-chromosomal lineages in Europe reveal Middle Eastern and Asiatic contacts. *Eur. J. Hum. Genet.* In press
198. Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, et al. 2002. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161:269–74
199. Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, et al. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* 60:1174–83
200. Zietkiewicz E. 2001. Modern human

origins and prehistoric demography of Europe in light of the present-day genetic diversity. *J. Appl. Genet.* 42:509–30

201. Zietkiewicz E, Yotova V, Gehl D, Wambach T, Arrieta I, et al. 2003. Haplotypes in the dystrophin DNA segment point to a mosaic origin of modern human diversity. *Am. J. Hum. Genet.* 73: 994–1015
202. Zietkiewicz E, Yotova V, Jarnik M, Korab-Laskowska M, Kidd KK, et al. 1998. Genetic structure of the ancestral population of modern humans. *J. Mol. Evol.* 47:146–55
203. Zilhão J, d'Errico F. 1999. The chronology and taphonomy of the earliest Aurignacian and ist implications for the understanding of Neandertal extinction. *J. World Prehist.* 13:1–68
204. Zvelebil M. 2000. The social context of the agricultural transition in Europe. In *Archaeogenetics: DNA and the Population Prehistory of Europe*, ed. C Renfrew, K Boyle, pp. 57–79. Cambridge, UK: McDonald Inst. Monogr.

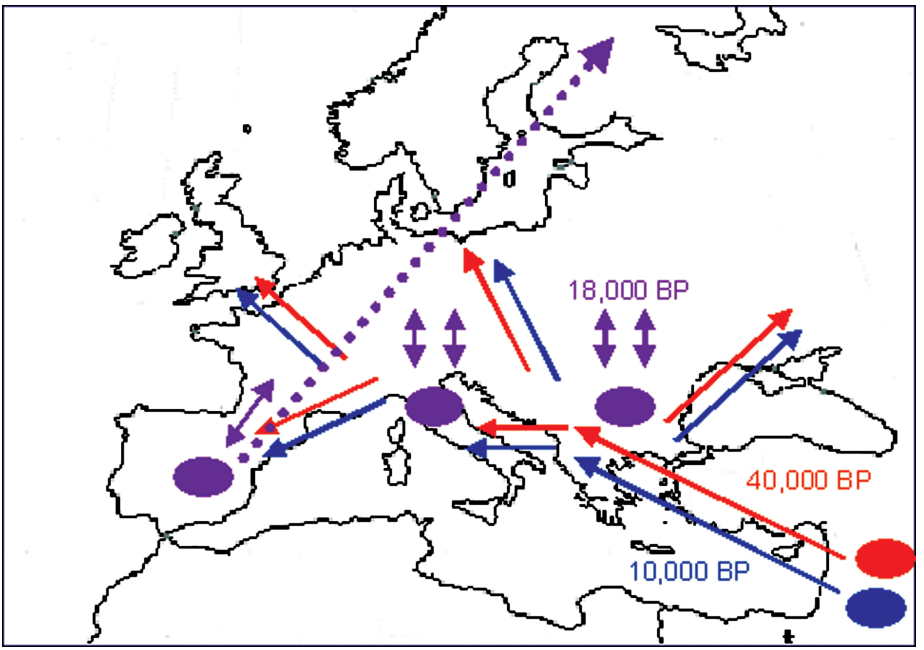

Figure 1 A scheme of possible population movements in Upper Paleolithic (*orange*), late Paleolithic (*purple*), and Neolithic (*blue*) times. Arrows represent dispersal processes inferred from the archaeological record with the approximate date of their beginning. The purple double arrows indicate a southward population contraction at the latest glacial maximum, followed by a northward expansion when temperatures increased; purple ovals represent the approximate areas where these populations concentrated at the last glacial maximum (refugia). A dotted arrow represents the postglacial expansion from Iberia into northeastern Europe, proposed on genetic grounds by Torroni et al. (185).

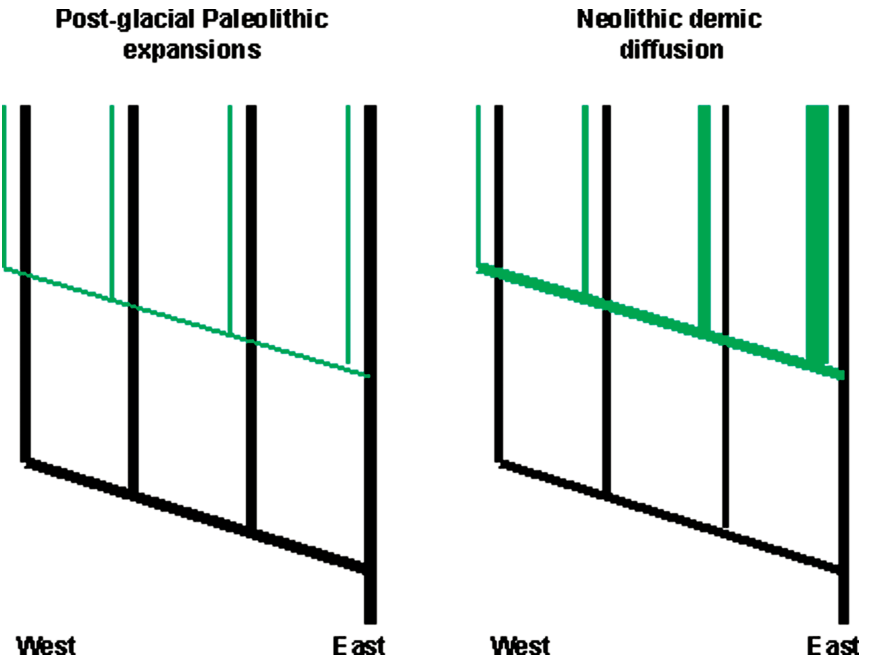

Figure 2 A schematic representation of the two main models of the European population history. Past is at the bottom, present at the top. The thickness of the rods is proportional to the relative demographic impact of Paleolithic (*black*) and Neolithic (*green*) immigration, under the PGE (*left*) and NDD (*right*) models, in four areas of Europe, along a west-east transect.