The DNA revolution in population genetics

LUCA L. CAVALLI-SFORZA (cavalli@volterra.stanford.edu)

Unprecedented clarity has come to our understanding of genetic variation by the analysis of DNA sequences. It is not surprising that the new DNA technologies are leading to a resurgence of interest in population genetics. In this review, I discuss recent progress and future directions towards reconstructing the history of human populations. There is increasing consensus on a recent 'Out of Africa' origin of modern humans, which explains why the greatest fraction of genetic diversity is found within populations, rather than between them. The comparison of Y chromosome and mitochondrial DNA data shows remarkable sex differences in geographic variation. The analysis of Neanderthal DNA has been a major breakthrough in the study of fossil DNA. Among major hopes for the future are applications to polygenic diseases.

Population genetics aims at understanding the forces that determine evolution. More than any other biological discipline, it uses a well developed and ever growing body of theoretical knowledge that allows quantitative predictions. The most useful information comes from 'genetic polymorphisms' – variants transmitted in strict mendelian fashion. Evolution of living humans is very recent, and rarely, if ever, has there been fixation of new alleles, so that variation in populations is almost always still at the polymorphic stage. Allele frequencies vary, however, among populations. The arsenal of available markers is the key to the analysis, and it is here that the molecular revolution, especially the more recent developments in the direct study of DNA, are generating great (potentially gigantic) progress.

Genetic markers

In order to understand the history and evolution of populations it is usually necessary to study a large number of polymorphisms. Until recent developments in molecular genetics, analysis of the genotype was usually indirect and limited to a small number of markers. These were based on the study of gene products, almost all protein polymorphisms, and only a few hundred[1] were known. They are usually referred to as 'classical polymorphisms' to distinguish them from those obtained by direct DNA analysis. The advantages of analyzing genetic polymorphisms at the level of DNA, rather than that of proteins or other gene products, are manifold. There is more information in DNA than in protein sequences; moreover, the techniques are, in principle, the same for any segment of DNA and there is a great number of genetic polymorphisms. Automation is easier, which is a great help, because we do need a greater number of markers than was ever studied or even available for most problems. There is also a considerable variety of types of polymorphisms now amenable to analysis, each of which has special merits. Most common are single nucleotide polymorphisms (SNPs or 'snips'), that is, replacement, loss or addition of one nucleotide. Indels (insertions/deletions) of longer segments are rarer, but also useful, being less likely to recur than SNPs. The great majority of these changes tend to remain biallelic. At the cost of some imprecision I refer to SNPs as BAs (biallelic polymorphisms). Oligoallelic polymorphisms occur in very rare cases in which there are more than two alleles.

There are probably several million BAs in the three billion nucleotide pairs forming the human genome, and until recently[2] it seemed impossible to analyze more than a small sample of them. Currently, we are bolder and, theoretically at least, we might become able to study almost the entire genome of an individual in one go. This is the promise of 'chip hybridization'[3], a combination of high-resolution nucleotide chemistry and analysis techniques. Current technology already permits the re-sequencing of DNA sequences as long as the human mitochondrial DNA (mtDNA) and of hundreds or thousands of BAs.

Until an adequate number of BAs are known and available, and chips and relevant machinery are more easily accessible, the most popular DNA markers will remain microsatellites. These widely employed multiallelic polymorphisms are repeats of very short nucleotide sequences. They have a high rate of addition or loss of repeat units so that the alleles are usually very numerous, and the number of repeats is an easily scored genetic marker. The analysis of microsatellites has also been automated to a reasonable extent. They have revolutionized the analysis of linkage and the building of genetic chromosome maps (e.g. Ref. 4). Repeats longer than microsatellites and less easy to study are often called minisatellites. Although minisatellites are less common than microsatellites, they have been used for population genetic studies (e.g. Ref. 5).

DNA analysis has also made it easier to study haplotypes, arrays of alleles at closely linked loci along a chromosome. These regions are short enough to show very little, or no, recombination and behave as blocks that might have ancient origins. Thus, they extend the power for evolutionary dating of the polymorphisms that they carry[6].

mtDNA

Being essentially haploid, and transmitted almost uniparentally, mtDNA and Y chromosomes have opened new evolutionary vistas. They complement each other nicely, giving separate information on the paternal and maternal contributions to evolution. Absence of recombination, due to haploidy and uniparental transmission, allows one to consider the array of loci along the chromosome as a single block or haplotype. There are limitations in the conclusions available today, but these studies have considerable potential.

mtDNA was propelled to celebrity by 'African Eve'. This biblical name, a notorious misnomer, is still clouding the minds of some by giving the impression that there was a time when there lived a single woman on Earth. There is no evidence that this was ever the case and, in fact, most estimates of the early number of humans called anatomically modern human, or *Homo sapiens sapiens*, are in the range of 10 000 to 100 000 (e.g. Ref. 7). The first major result of mtDNA analysis was the origin of anatomically modern humans at about 200 kya [1 kya

 0168-9525/98/$19.00
PII: S0168-9525(97)01327-9

(kiloyear) ago = 1000 years ago]. This estimate[8], based on a sample of restriction polymorphisms from the whole mtDNA, and confirmed by work on a fully sequenced shorter section of about 600 nucleotides (the 'D loop'), generated much criticism (e.g. Ref. 9). Recent research based on the full sequence of mtDNA of three humans and four most-closely related primates reduced this estimate to slightly less than 150 kya (Ref. 10). Knowledge of the sequences of many other individuals will be necessary to strengthen these estimates. Differences between these results and those using other methods of estimation are discussed in the next section.

Population genetic studies are in approximate agreement with archeological observations indicating that anatomically modern humans (i.e. similar, as far as bone morphology goes, to living humans) are found in the past 100 000 years exclusively in Africa, or very close to it (the Middle East)[11] and spread from it to the other continents. They are also in agreement with the earlier finding, using classical markers, that the first split in modern evolutionary history was between Africans and non-Africans[12]. Some paleoanthropologists support the alternative 'multiregional hypothesis': that *Homo sapiens* evolved in a nonfocused way throughout the region occupied by *Homo erectus*, an earlier human species that migrated from Africa to the Old World one million or more years ago[11].

Another interesting result, originally obtained with mtDNA, but extensible to some other markers, is the analysis of the rate of population growth in the past tens of thousands of years[13,14]. The summation of a considerable number of observations has shown conclusively that there have been important periods of exponential growth in fractions of the human population[15].

Arguably, the most important contribution of molecular genetics to the study of human evolution was published last year – the analysis of DNA from the right humerus of a famous fossil, Neanderthal. The original specimen, discovered in 1856 near Dusseldorf, was the first human fossil found, and was analyzed in Munich by Pääbo's team[16]. Earlier studies on a Bronze-Age man in the same laboratory[17] show that mtDNA can supply convincing information on fossils because its presence in high numbers in the cytoplasm makes it more likely that enough DNA fragments will survive for analysis compared with chromosomal genes. But DNA fragmentation requires repeated cloning and it is only the consensus of cloned fragments that generates credible results.

More than 380 nucleotides in region 1 of the D loop were studied. The average difference among pairs of modern humans is 8.0 (range 1–24), while the range of the difference between a modern human and Neanderthal is 22–36. These results put Neanderthal out of the modern human line, and confirm that it is most probably completely extinct. The common ancestors to Neanderthal and modern humans were dated between 550 kya and 690 kya.

Y-chromosome markers

Data from Y chromosome markers[18,19] have been used to generate estimates of the date of 'birth' of the first modern human and suggest values that confirm a recent African origin. However, the number of published polymorphisms of the Y chromosome is limited[20–23]. Recently, a systematic search for Y biallelic markers has begun[24,25] and a large number of markers is being detected[26].

A comparison indicated that variants of the Y chromosomes are more highly geographically clustered than the mtDNA (Refs 26, 27). This has been observed very clearly in Europe[28], were mtDNA shows almost no variation while markers of the Y chromosome show a major gradient. One major explanation must be that the migration of men is usually more limited than that of women. This may seem surprising, but anthropological observations confirm it; women are those who move more often to join husbands at marriage. The greater prevalence of polygamy in males also contributes to sex-differential mobility. Moreover, the high mutation rates of mtDNA tend to simulate a greater geographic spread and, hence, flatten the geographic distribution of mtDNA variants. It seems that, for these reasons, mtDNA is not, or not always, a good marker for population or migration studies.

It has been known for some time, thanks to observations on classical polymorphisms, that human genetic variation is largely between individuals within populations, rather than between populations or even continents[29,30]. DNA markers (RFLPs and microsatellites) have confirmed this, giving very nearly the same apportionment of genetic variation as classical markers – 85% among individuals within populations and 15% among populations[31]. In agreement with the observations of higher geographic clustering, the portion of genetic variation of Y chromosomes within populations is significantly smaller than that of mtDNA and autosomes (M.T. Seielstad *et al.*, unpublished).

Dating evolution by genetic methods – gene trees and population trees

There are various ways of using genetic information for dating evolution, and they lead to results that have different meanings. If one or more mutations of a gene are found in at least some individuals of one population, but not in those of another, evaluating the birth date of the mutation(s) is a help in predicting the date at which the two populations separated. But it is not the same date; to gain useful information, the mutation must have occurred before the population fission. Examples of dates derived from such 'gene trees' include all investigations published on mtDNA and HLA (Refs 7, 32), hemoglobin genes[33] or haplotypes[6]. But trees based on genetic distances calculated from gene frequencies of populations refer to dates of population fissions ('population trees'). The fundamental difference between a gene tree and a population tree seems to have been overlooked by a number of researchers (Ref. 34, p. 277; Ref. 35, p. 87).

It is the date of population splits, not the birth of individuals carrying the first mutation, that is of interest for comparison with archeological dates, which usually refer to the early settlement of new areas. The common use of gene trees in the molecular evolution of species does not raise problems, because the time interval between the origin of a mutation and the physical fission of two species is usually small, compared with the interval between two successive species fissions in a species tree. But the distinction between gene trees and population trees is very important in intraspecific evolution, because in the evolution of modern humans the delay between the appearance of a mutation in and the

subsequent split of populations can be very long. An estimate was generated for the average time of origin of mutations leading to RFLPs found in modern humans, which is about 700 kya, which is compatible with an estimate of the split of African and non-African populations at around 100 kya[36]. On the other hand, the dates one can hope to obtain from archeology (e.g. dates of first settlement in a geographic area) are much more closely related to those of population fissions, and might be much later than those of first origins of mutations. Even if archeological dates inevitably suffer from considerable approximation, and are subject to frequent revisions, they are our major, if not only, hope of independent control.

Even keeping in mind that gene trees and population trees answer different questions, we still need a calibration method for both. There are, in principle, two approaches. The first approach relies on paleontological events that have been dated and provide a calibration curve of genetic distance versus time elapsed. This is the method commonly employed in most studies of molecular evolution. Unfortunately, time reference standards are few and uncertain. The separation of chimpanzees and humans, usually set in the neighborhood of five million years, is often employed in studies of human evolution and is the basis of most evolutionary dates given in the literature. For population trees of modern humans, current estimates of first settlement of major continents were found to be reasonably proportional to genetic distances (Ref. 35), and seem compatible with the first split of African and non-African populations of around 100 kya.

The second approach to calibration is the use of an internal clock: mutation rates. The genetic distance between two individuals or two populations should be proportional, in general, to the product of the mutation rate and the time elapsed since their separation. Unfortunately, our knowledge of mutation rates is limited. It is very likely that mutation rates vary greatly from locus to locus, and also among sites of the same locus. This means that most evaluations in the literature, based on guesstimates of the rates of one or a few mutations, are subject to strong doubt.

Microsatellites have properties that make them very suitable for dating evolutionary events on the basis of mutation rates. It is easy to average data from many loci and the mutation rate per locus is so high that a reasonable number of mutational events can occur in the short time of modern evolution, which is the past 50–200 kya. For CA dinucleotide microsatellites the average estimate *in vivo* is about 1/5000 (Refs 4, 37) per locus per generation. Microsatellites mutate back and forth, and usually one repeat is added or deleted. The first application of microsatellites in human population studies[38] used 30 microsatellites[39] and gave a time of separation for Africans and non-Africans of 156 kya, with a large statistical error (95% fiducial interval, 95–290 kya). In later experiments (L. Jin *et al.*, unpublished) the number of microsatellites was increased to over 100. A current estimate, including needed correction factors, gives a value closer to 60 kya, with a standard error close to 20%. More refined formulae of genetic distance are being developed, also taking into account the fact that some mutations involve a change of number of repeats greater than one, and of peculiarities of the mutation process.

Trees for population history and evolutionary mechanisms

Trees have been used extensively since their first introduction as an attempt at reconstructing the history of populations. Naturally, there are many possible sources of errors (Ref. 35, pp. 30–39, 376) – trees are 'fallible friends'.

When different trees show different results, I suspect a major culprit is the number of markers, which is almost always too low in published trees. The best test of the validity of a tree uses a statistical method of re-sampling the genes tested, called the 'bootstrap'[40]. It usually shows that it would take many more markers than normally employed to generate trees highly reproducible in all details of their topology. It can also be applied to estimating the standard errors of branch lengths. This is rarely done and would give unpleasant surprises.

Even if it is true that adequate numbers of markers have been very rarely used, there are now several cases in which the study of the same or similar sets of individuals, or at least similar sets of populations, gives the same or a very similar topological tree using different types of markers. But a serious problem is that two popular sets of tree methods tend to give consistently different results and, in some cases, even gross differences in topology. Typical representatives of the first set are maximum likelihood or ML (and its usually faithful, numerically convenient short-cut called the Unweighted Pair Group Method using Arithmetic Averages, or UPGMA), and the second set is represented by Neighbor Joining (NJ)[35] methods. The first set ordinarily assumes independent evolution in all the branches of the tree, as would be expected for mutation, drift and certain models of variable selection. The second set assumes that the number of mutations has been the 'minimum'[41] possible, or, if the variable studied is a gene frequency, that it has followed the fastest path from the gene frequencies of one population to those in another. Unfortunately, statistics minimizing evolution do not make much sense from the point of evolutionary theory. Naturally, the principle of minimum evolution can supply reasonable approximations, but there is proof that it underestimates the number of mutations[42].

An example of discrepancy between the results obtained with the two sets of tree methods is shown in Fig. 1 (Ref. 43). NJ typically produces trees with segments of different length, and cannot suggest the position of the tree route, unlike ML or UPGMA. The difference in the length of segments might be due to differences in drift because of different population size. This is an advantage of NJ, but the disadvantage is that if there are admixtures of populations that had previously diverged, branches of admixed populations become short and move towards the center of the tree (Ref. 43). Although the results of tree reconstruction are largely independent of the class of markers used, mtDNA shows an unacceptable major discrepancy in the relative lengths of branches[44,45], compared with data obtained using chromosomal DNA polymorphisms (Fig. 1). With mtDNA, irrespective of the tree method used, African populations tend to diverge regularly one from the other, while all non-African populations have extremely short branches and form a small cluster in which it is difficult to recognize the standard branchings. A possible explanation is the segregation lag of new mutants due to heteroplasmy (genetic heterogeneity due to the large number of mtDNA chromosomes

usually present in the cell). The level of heteroplasmy – the numbers of mitochondria per cell – during the germinal cell cycle is not sufficiently well-known to test the validity of this explanation[46].

Genetic geography

Trees attempt to describe the evolution of a group of populations. Confidence in them is generated by their being reproducible with different sets of markers, provided that their numbers are sufficiently large. But there are limits to their usefulness; they are increasingly difficult to reconstruct as the number of populations grows, and it is difficult to obtain a sample of populations that is truly representative of a region, or of the world.

When it is possible to have data on an adequate number of populations covering a large area, it is more satisfactory to study their genetic geography. This approach has been useful for understanding the pattern of natural selection behind certain genetic diseases; the first and classical example being that of resistance to malaria of sickle cell anemia and thalassemia heterozygotes.

The history of mutation and natural selection is usually different from gene to gene, but the two other major evolutionary factors, migration and drift, influence all genes in parallel ways and can be studied efficiently only by summing the information gained from many genes. This is of major importance to the study of population trees. An important requirement is that one needs to average the effects of very many genes before reconstructed trees are reproducible. The same considerations apply to the geographic analysis of genes[47]. As for trees, a requirement for validity is the independence of conclusions that are drawn from the set of genes employed. The most informative geographic syntheses are obtained by the use of statistical techniques, such as principal components analysis (also principal coordinates and multidimensional scaling), which identify hidden patterns in the original multivariate data, usually the frequencies of many genes in many populations. The principal components correspond to the statistically independent hidden patterns identified by the method. When the original population data are sufficiently numerous to be displayed in a geographic map, the principal components can also be shown in the same way, and might immediately lead to some meaningful and testable interpretation (Fig. 2)[35,47–49].

Migration has a linear effect on gene frequencies and, thus, principal components are especially useful[47] for studying it. But the analysis of geography requires intensive data coverage of extended regions.

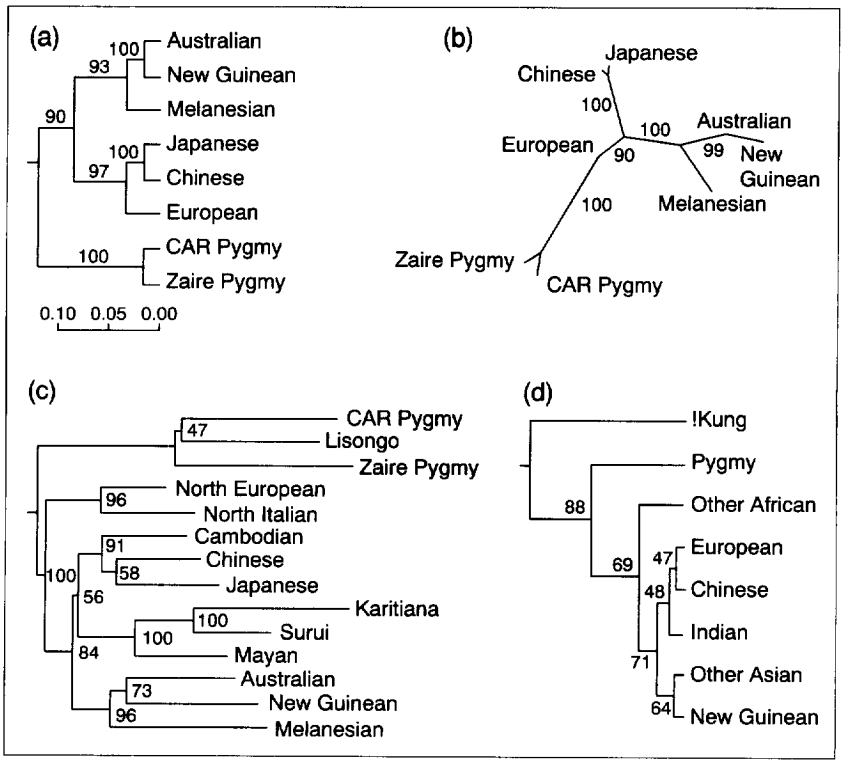
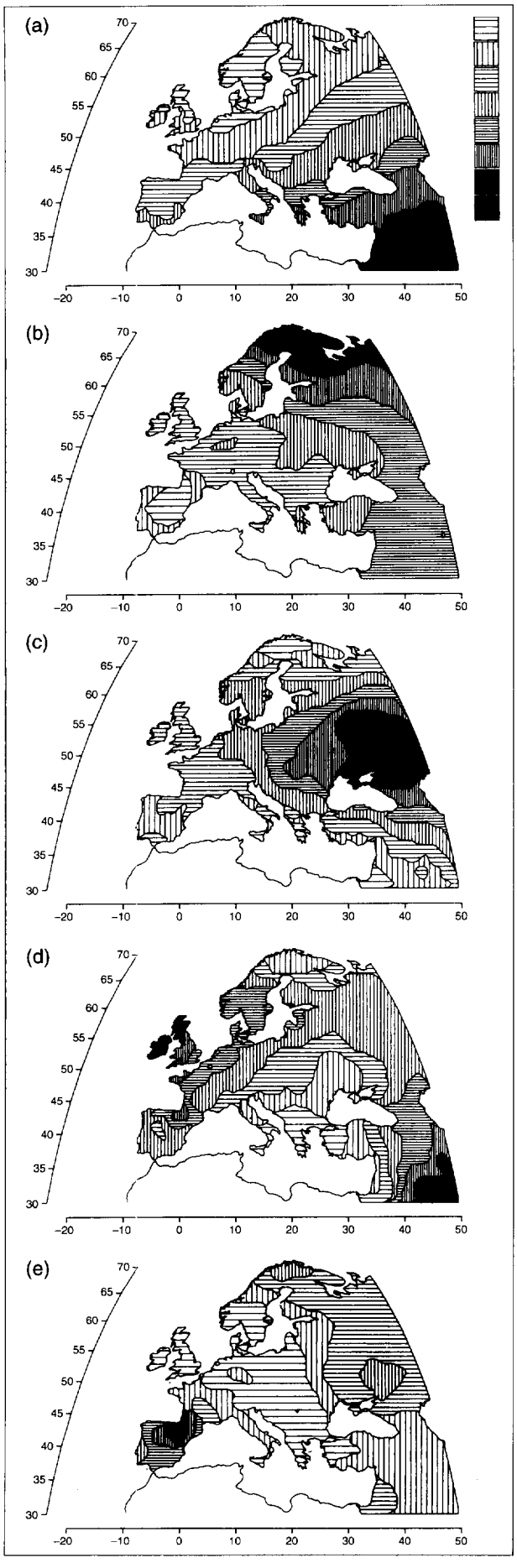


FIGURE 1. Comparison of population trees constructed (a) with UPGMA (Unweighted Pair-Group Method using Arithmetic Averages – a method assuming constant evolutionary rates) and (b) with NJ (neighbor joining). The trees were obtained with 87 RFLPs (Ref. 36). Numbers at the nodes are bootstrap values from 100 bootstraps. Note the difference in the position of the European branch, which is closest to Africa using the NJ method, and its extreme shortness. In part, this is due to the biased selection of RFLPs that were detected using a small panel made exclusively of individuals of European origin, and also to the admixture of Europeans because of their proximity to Africa and Asia (as confirmed by simulations). A similar situation is observed with classical polymorphisms and with microsatellites. (c) Microsatellite polymorphisms analyzed with NJ (Ref. 39), rooted by using the midpoint among the two most-distant polymorphisms. In this case, the root corresponds to that obtained by using chimpanzees as an outgroup. (d) A UPGMA mtDNA tree[45]. Note the extraordinary shortening of the branches leading to non-African populations. (Adapted, with permission, from Ref. 35.)

One conclusion, reached from genetic geography of classical markers, is that migrations and expansions stimulated by major technological innovations have played a major role in the evolution of modern humans. Unfortunately, DNA polymorphisms have not yet been investigated in an adequate number of populations for drawing geographic maps. An exception is given by two restriction polymorphisms of the Y chromosome, for which a reasonable map of Europe can be made[50]. They confirm the genetic pattern that is thought to have been due to the spread of Neolithic farmers from the Middle East towards Europe. It has been stated that the mtDNA D loop does not confirm this migration[51] but, in truth, these data simply do not show any statistically significant variation in Europe. Greater numbers of individuals would be necessary for the reasons discussed earlier[28,44].

Future hopes

A combination of methods of historical and geographical analysis can greatly improve our understanding of human history and evolution. A consensus is slowly developing, but it is essential to reach agreement among results from many different markers and analytical



approaches. Insufficient BAs are currently available, but their number will rapidly increase, given the keen interest in obtaining them, as well as the existence of several new and efficient detection techniques, such as denaturing high-performance liquid chromatography[26].

In-depth analysis of normal human genetic variation is also necessary for understanding and, eventually, treating successfully much disease. So far, medical genetics has dealt most effectively with the simplest hereditary diseases, which are caused by single clear-cut mendelian factors. But many important, often common, diseases, such as diabetes and heart disease, probably depend on the interaction of several genes and external factors. When several different loci contribute to a phenotype (a polygenic system) it is likely that the alleles at loci responsible for such interactions have high frequencies in populations. As an example, if six genes contribute equally to a disease with an incidence of 1.5%, each susceptibility allele must have a population frequency of around 50%. Thus, factors of polygenic disease are often likely to be 'normal' alleles from unsuspected SNPs that have relatively high frequencies. The identification of 'normal' polymorphisms is of greater importance for medical genetics than hitherto believed.

The study of human genome variation is still made in the fragmentary and haphazard manner in which it has been going on for almost 80 years. This is an inefficient and uneconomic way of doing science, and there is little control on the ethical aspects of DNA-sample collection. It could hardly be more urgent to organize ethically and efficiently the collection of samples from populations, and the distribution of samples for testing by interested laboratories and the analysis of results. These are the aims of the Human Genome Diversity Project (HGDP), the usefulness of which has been recently supported by a study by the US National Research Council. The project has already begun in the first four regions in which it has been funded: Europe; China; India; south-west Asia; and only very recently at a pilot level in the USA. It is hoped that it will soon extend to other regions. A small collection of cell lines has been supported by the NIH for some time, and it is planned to form at least

FIGURE 2. The first five principal components of European classical polymorphisms[35,48–50]. The first accounts for about 28% of the total genetic variation; it parallels almost exactly the expansion of Neolithic farmers, which began in the Middle East about 9500 years ago and is accurately dated archeologically. The second accounts for 22% and might represent the spread of Uralic language speakers from the region around the Northern Urals. The third accounts for 11% and can be explained by the spread of pastoral nomads who originally inhabited the Don-Volga region (in the Bronze Age, about 5000 years ago). The fourth accounts for 7% and is in agreement with the colonial expansion of Greeks in the first millennium BC. The fifth accounts for 5%; the darkest area corresponds approximately to the area where Basque language is spoken today. This language is likely to have been spoken in this area for a very long time, as confirmed by the study of place names. (Adapted, with permission, from Ref. 35.). Principal components are expressed in units of standard deviation and vary around a mean of zero. They are represented graphically in a scale of shadings varying from lightest to darkest. The most likely area of origin of expansion is the extreme PC value represented by the darkest shading in (a), (b), (c), and by the lightest one in (d).

two general repositories of cell lines in duplicate (one in Europe and one in the USA) for the storage of cell lines in duplicate, the generation and distribution of DNA samples to scientists, and the collection of results (for general circulation). The Center for the Study of Human Polymorphisms (CEPH) founded in Paris by Jean Dausset has played a major role in the establishment of the human genetic map, and a collaborative plan of HGDP and CEPH is under consideration.

Finally, the genetic study of fossil material has lagged behind, because of technical difficulties, but has now made a good start and will undoubtedly continue, even if it is unquestionably very demanding, and problems of DNA conservation will often limit the chance of obtaining results from older samples. Technological breakthroughs in the analysis of genetic variation have truly revolutionized our ability to examine our own past.

References

1 Nei, M. and Roychoudhury, A.K. (1988) *Human Polymorphic Genes: World Distribution*, Oxford University Press
2 Cavalli-Sforza, L.L. (1990) *Am. J. Hum. Genet.* 46, 649–651
3 Chee, M. *et al.* (1996) *Science* 274, 610–614
4 Dib, C. *et al.* (1996) *Nature* 380, 152–154
5 Armour, J.A.L. *et al.* (1996) *Nat. Genet.* 13, 154–160
6 Tishkoff, S.A. *et al.* (1996) *Science* 271, 1380–1387
7 Erlich, H.A., Bergstrom, T.F., Stoneking, M. and Gyllensten, U. (1996) *Science* 274, 1552–1554
8 Cann, R.L., Stoneking, M. and Wilson, A.C. (1987) *Nature* 325, 31–36
9 Templeton, A.R. (1993) *Am. Anthropol.* 95–72
10 Horai, S. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 532–536
11 Klein, R.G. (1989) *The Human Career: Human Biological and Cultural Origins*, University of Chicago Press
12 Nei, M. and Roychoudhury, A.K. (1982) *Evol. Biol.* 14, 1–59
13 Slatkin, M. and Hudson, R.R. (1991) *Genetics* 129, 555–562
14 Rogers, A.R. and Harpending, H. (1992) *Mol. Biol. Evol.* 9, 552–569
15 Harpending, H. *et al. Proc. Natl. Acad. Sci. U. S. A.* (in press)
16 Krings, M. *et al.* (1997) *Cell* 90, 1–20
17 Handt, O. *et al.* (1994) *Science* 264, 1775–1778
18 Whitfield, L.S., Sulston, J.E. and Goodfellow, P.N. (1995) *Nature* 378, 379–380
19 Dorit, R.L., Akashi, H. and Gilbert, W. (1995) *Science* 268, 1183–1185
20 Jobling, M.A. and Tyler Smith, C. (1995) *Trends Genet.* 11, 449–456
21 Hammer, M.F. (1995) *Nature* 378, 376–378
22 Hammer, M.F. and Zegura, S.L. (1996) *Evol. Anthropol.* 5, 116–134
23 Deka, R. *et al.* (1996) *Genome Res.* 6, 1177–1184
24 Seielstad, M.T. *et al.* (1994) *Hum. Mol. Genet.* 3, 2159–2161
25 Underhill, P.A. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 196–200
26 Underhill, P.A. *et al.* (1997) *Genome Res.* 7, 996–1005
27 Ruiz-Linares, A. *et al.* (1996) *Am. J. Hum. Genet.* 60, 401–408
28 Cavalli-Sforza, L.L. and Minch, E. (1997) *Am. J. Hum. Genet.* 61, 247–250
29 Lewontin, R.C. (1972) *Evol. Biol.* 6, 381–398
30 Nei, M. and Roychoudhury, A.K. (1972) *Science* 177, 434–435
31 Barbujani, G. *et al.* (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 4516–4519
32 Ayala, F.J. (1995) *Science* 270, 1930–1936
33 Harding, R.M. *et al.* (1997) *Am. J. Hum. Genet.* 60, 772–789
34 Nei, M. (1987) in *Molecular Evolutionary Genetics*, Columbia University Press
35 Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A., (1994) *The History and Geography of Human Genes*, Princeton University Press
36 Mountain, J.L., Lin, A.A., Bowcock, A.M. and Cavalli-Sforza, L.L. (1992) *Philos. Trans. R. Soc. London Ser. B* 377, 159–165
37 Weber, W. and Wong, C. (1993) *Hum. Mol. Genet.* 2, 1123–1128
38 Goldstein, D.B., Ruiz-Linares, A., Cavalli-Sforza, L.L. and Feldman, M.W. (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 6723–6727
39 Bowcock, A.M. *et al.* (1994) *Nature* 36, 455–457
40 Felsenstein, J. (1985) *Evolution* 9, 783–791
41 Sober, E. (1988) *Reconstructing the Past: Parsimony, Evolution and Inference*, Massachusetts Institute of Technology Press
42 Cummings, M.P., Otto, S.P. and Wakeley, J. (1995) *Mol. Biol. Evol.* 12, 814–822
43 Ruiz-Linares, A., Minch, E., Meyer, D. and Cavalli-Sforza, L.L. (1995) in *The Origin and Past of Modern Humans* (Brenner, S. and Hanihara, K., eds) pp. 123–148, World Scientific
44 Mountain, J.L. *et al.* (1995) *Am. J. Hum. Genet.* 56, 979–992
45 Jorde, L.B. *et al.* (1995) *Am. J. Hum. Genet.* 57, 523–538
46 Bendall, K.E., Macaulay, V.A., Baker, J.R. and Sykes, B.C. (1996) *Am. J. Hum. Genet.* 59, 1276–1287
47 Menozzi, P., Piazza, A. and Cavalli-Sforza, L.L. (1978) *Science* 201, 786–792
48 Piazza, A. *et al.* (1995) *Proc. Natl. Acad. Sci. U. S. A.* 92, 5836–5840
49 Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1993) *Science* 259, 639–646
50 Semino, O. *et al.* (1996) *Am. J. Hum. Genet.* 59, 964–968
51 Richards, M. *et al.* (1996) *Am. J. Hum. Genet.* 59, 185–203

L.L. Cavalli-Sforza *is in the Department of Genetics, Stanford School of Medicine, Stanford, CA 94305-94293, USA.*