Geographic patterns of nuclear and mitochondrial DNA variation in Europe

Barbujani, Guido
Universitá di Ferrara
Dipartimento di Biologia
Via L. Borsari 46
44100 Ferrara. Italia

Gizabanako modernoen arteko bariazio genetikoa aztertzeak informazioa ematen digu historiaurreko gertakariez. Dibertsitate genetikoaren gradu-arauak eta proba arkeologikoetan oinarrituriko ereduen aurreikuspenekin alderatzeak aditzera ematen duenez, neolitoan gertatu Sortaldeko lehen nekazarien hedapena izan zen segur aski Europako populazioaren historiaren gertakari nagusia, baina hala Homo sapiens sapiens-ek abiaturiko lehen kolonizazioa nola geroagoko migrazioak dokumentaturik daude egungo datu genetikoen bankuetan.

Giltza-Hitzak: Dibertsitate genetikoa. ADN. Graduak. Genealogia genetikoak. Populazioen barreiadura neolitoan. Paleolitoko kolonizazioa.

El análisis de la variación genética entre individuos modernos proporciona información sobre acontecimientos prehistóricos. La comparación de niveles pautas de diversidad genética con las predicciones de modelos basados en pruebas arqueológicas sugiere que la extensión de los primeros granjeros desde el Levante mediterráneo durante el neolítico fue probablemente el episodio principal de la historia de la población europea, pero tanto la colonización inicial por el Homo sapiens sapiens como migraciones más recientes están documentadas en los actuales bancos de datos genéticos.

Palabras Clave: Diversidad genética. ADN. Grados. Genealogías genéticas. Diseminación de poblaciones en el neolítico. Colonización en el paleolítico.

L'analyse de la variation génétique parmi des individus modernes fournit une information sur des événements préhistoriques. La comparaison de niveaux, modèles de diversité génétique avec les prédictions de modèles basés sur des preuves archéologiques suggère que l'extension des premiers fermiers depuis le Levant méditerranéen durant le néolothique fut probablement l'épisode principal de l'histoire de la population européenne, mais aussi bien la colonisation initiale par l'Homo sapiens sapiens que des migrations plus récentes sont documentées sur les banques de données génétiques actuelles.

Mots Clés: Diversité génétique. ADN. Degrés. Généalogies génétiques. Dissémination de populations à l'époque néolithique. Colonisation à l'époque paléolithique.

XV Congreso de Estudios Vascos: Euskal zientzia eta kultura, eta sare telematikoak = Ciencia y cultura vasca, y redes telemáticas = Science et culture basque, et réseaux télématiques = Basque science and culture, and telematic networks (15. 2001. Donostia). – Donostia : Eusko Ikaskuntza, 2002. - P 161-169. - ISBN: 84-8419-949-5.

1. INTRODUCTION: DIFFERENCES AMONG AND WITHIN POPULATIONS

Natural species, including humans, are genetically polymorphic, in two related, but conceptually distinct ways. Different individuals may carry different alleles at the same locus (polymorphism within populations), and different populations may be associated with different alleles or with different frequencies of the same alleles (polymorphism between populations).

The relative importance of polymorphism within and between populations varies across species. In many species the largest share of genetic diversity occurs among groups of populations, whereas individuals of the same group tend to be rather similar. When that is the case, individuals of unknown origin can be assigned to their group on the basis of their genotype, with little error. Examples include most plants, and animals such as honey bees (Estoup et al. 1995), sea urchins (Palumbi et al. 1997), chimpanzees (Morin et al. 1994; Goldberg and Ruvolo 1997), and lowland gorillas (Ruvolo et al. 1994). When morphologically or genetically distinct groups occupy geographically distinct regions, it is customary to define them as subspecies. This corresponds to a classical definition of races in anthropology: each race occupies a different territory, and differs from other races "in measurable characteristics to a considerable degree" (Coon 1963; see also Cohen 1991). Evolutionary trees summarising variation in those species typically show few, well-separated major clusters of haplotypes or genotypes (fig. 1A).

Other species, on the contrary, show little or no internal structuring (a term we shall use to indicate genetic diversity between populations or groups). Populations of Antarctic crustaceans separated by thousands of kilometers do not differ significantly for allozyme frequencies (Fevolden and Schneppenheim1989), and monarch butterflies living on opposite slopes of the Rocky Mountains have the same mitochondrial haplotypes at the same frequencies (Brower and Boyce 1991). Evolutionary trees summarising variation in those species do not show clear clustering of haplotypes or genotypes of the same geographical origin (fig. 1B).

Such different patterns of genetic diversity reflect different evolutionary histories. Although it is difficult to generalise, the degree of structuring is largely and inversely related to the levels of gene flow. In practice, if population sizes are roughly equivalent, the level of genetic differentiation between groups will reflect their reciprocal isolation. This well-known principle of population genetics has been confirmed in many field studies.

What is the place of humans in this picture? We differ from each other, and this is a common observation. But, to cite just one example among many, *Drosophila pseudoobscura* shows a ten-fold greater nucleotide diversity than our species (Li and Sadler 1991). Thus, we are not among the most variable organisms. Nevertheless, the question remains of how these limited differences are distributed, within human populations and between them. Until recently, this was not an interesting question to most scientists. Even after the theories postulating independent origins for the different human groups had lost their appeal (see Cohen 1991, pp. 23-28), the existence of highly differentiated human races was not a matter of discussion. However, retrospectively, it is clear that some major problems had been overlooked.

The starting point for all systems attempting to classify humans is the identification of a few basic human types. There are two problems, though; any groups defined on the basis of morphological traits harbours some degree of variation, and different criteria, say skin color and skull shape, lead to different classifications. Where should racial boundaries be put, and what if different markers show different patterns of variation? No simple answer has been found, and so it comes as no surprise that, although the notion that human races exist has been challenged only recently (Livingstone 1962), no agreement has ever been found on how many races exist, and which. Even if one considers only the 20th century, the numbers of races proposed in various studies (partly outlined in Table 1) vary from 3 to 200 (Brown and Armelagos 2001).

Understanding whether deeply differentiated groups exist in our species is crucial to understanding human evolutionary history. Indeed, the two questions "How different are we?" and "How come we are the way we are"? are closely related. Models assuming that humans were subdivided in distinct groups over much of their evolution, such as Wolpoff et al.'s (1984) Multiregional Theory, predict a high degree of differentiation among major human groups, whereas models assuming a recent human origin, such as the Out-of-Africa theory (Stringer 1989) predict that genetic differences among human groups did not have sufficient time to accumulate, and thus that most human diversity falls within populations, not between them

The large-scale study of genetic polymorphisms has made it possible to assess in a quantitative, less subjective way, the pattern and the extent of genetic diversity among humans. For that purpose, the overall genetic variation can be broken down into three components, representing respectively: 1. Individual differences within samples; 2. Differences between samples belonging to the same race; and 3. Differences between races. Lewontin (1972) collected data on nine blood groups and eight serum proteins and blood cell enzymes, in almost all populations of the world that had been typed at that time, grouped in seven major racial groups.

The average proportion of the overall variance that fell within populations was 0.854. Additional differences between populations of the same race represented on average, 0.083 of the total, and differences between races represented the remaining

0.063 of the total. At the protein level, therefore, variation among seven races seems to represent a minor component of the overall genetic diversity of our species.

Lewontin's conclusion that the human racial classification has virtually no genetic or taxonomic significance was not unanimously accepted (see e.g. Nei and Roychoudhury 1982). One serious objection was that electrophoretic or serologic polymorphisms may not fully reflect the underlying genetic diversity. A more reliable answer should be sought at the DNA level.

Genetic diversity at 109 DNA loci was apportioned at the same levels of population structuring in 16 populations from five continents (Barbujani et al. 1997), by means of an analysis of molecular variance, or AMOVA (Excoffier et al. 1992). The populations were chosen so as to form well-separated groups over the map of the world, whereas geographically and ethnically intermediate populations were not considered. Therefore, if sampling affected the genetic variances estimated, it did so by increasing the weight of the between-continent components. The loci chosen were 30 short-tandem repeats (STR or microsatellite loci) of chromosomes 13 and 15, and 79 restriction-fragment length polymorphisms (RFLP). For 16 RFLPs, compound haplotypes referring to subsamples of individuals were available. Although such haplotypes are not statistically independent from the RFLPs, as analysed individually, variance components were also independently estimated from them. Overall, 1109 individuals were considered, 321 of them having been characterised for RFLP multilocus haplotypes.

The within-population variance component appeared significantly greater than 0 at almost all loci, representing a fraction of the overall variance comprised between 54.4% and 94.6% of the total. Very similar results were obtained in the analysis of single STR loci, of the multilocus STR haplotypes, and of RFLP polymorphisms (Barbujani et al. 1997). Table 2 summarises the results of the analysis of DNA data, along with comparable studies of the same scope. It is evident that the within-population component of the overall genetic diversity of our species is close to 85 per cent, no matter which kind(s) of genetic markers and which populations are studied. With respect to Lewontin's result, the analysis of DNA diversity suggests that the among-group (races or continents) component is somewhat greater, close to 11 per cent, whereas differences between populations of the same continent or race account for a lesser fraction. This difference is presumably due to the elimination of intermediate populations from the DNA study, and may or may not be evolutionarily significant.

What is clear, however, is that, despite the limited polymorphism of the protein markers employed, and despite the choice of a different set of populations, the results obtained by Lewontin are fully confirmed at the DNA level. STR loci are among the most variable components of the genome, and yet their analysis confirms that the largest fraction of human genetic diversity falls within populations. Belonging to different populations of the same race or continent adds betwen 3 and 8 per cent to the expected random diversity between individuals; and if two individuals come from different races or continents, a further fraction estimated between 6 and 11 per cent must be added. But in all studies carried out so far, the differences between racial or continental groups appear to add less than 20 per cent to the random genetic differences observed between members of the same community.

The studies of protein and DNA variances carried out so far indicate that there are no major genetic differences between broad human groups, as defined on the basis of a racial or geographical classification system. In our species, genetic diversity largely reflects individual differences among members of the same population (Barbujani et al. 1997). In practice, we could expect a random population on earth to contain more than four-fifths of the current overall genetic diversity. Another way to put it is, if we pose equal to 100 the genetic difference between us and the individual of the world who seems to least resemble us, the genetic difference between us and another member of our community (who is not a close relative of ours) is expected to be 85, and not 10 or 20, as many would suppose.

What do these results imply for the controversy about modern human origins? The multiregional model predictes a substantial genetic differentiation of continental groups, whereas little differentiation is expected under the out-of-Africa model. Studies of genetic variances, therefore, support the latter. By itself this fact would not be sufficient to take a strong standing in favour of a recent African origin of humankind. However, almost all recent studies of DNA diversity in humans are easier to reconcile with a recent African origin than with alternative possibilities.

The absence of major genetic differences among continental human groups seems therefore a consequence of the peculiar evolutionary history of our species. Our common origin is too recent, and the genetic exchange between groups has been too extensive, to permit a substantial diversification of a few, broad races. As Ruvolo et al. (1994) put it, on the average even very distant members of our species are less genetically different than random pairs of gorillas inhabiting the same forest.

2. GENETIC VARIATION IN EUROPE

Genetic variation in Europe is a subset of that 5 to 10 per cent that separates populations of different continents. We have seen the largest share of human diversity falls within populations, and hence it is clear that the largest share by far of the

European diversity is accounted for by individual differences among members of the same population. Still, by analysing differences among populations important historical inferences cam be drawn.

Archaeological evidence shows that two main migrational processes, both occurring from the Near East, have been important in the peopling of Europe. Although both probably entailed several, sometimes complicated, subprocesses, including choice of favourable habitats, local extinctions, and spatially random gene flow, for the sake of clarity it is useful to speak of an initial Palaeolithic colonisation, and of a later Neolithic expansion.

Palaeolithic people were hunter-gatherers who dispersed in Europe starting around 45,000 years ago, probably following the first wave of human migrations out of Africa (Mellars 1992). The first archaeological evidence of farming activities is in Anatolia and in the Levant, at can be dated around 10,000 years ago (Ammerman and Cavalli-Sforza 1984). Later specimens demonstrate that food-producing technologies spread north and west. The establishment of farming societies across Europe, from the Levant all the way to Iberia and the British Isles, lasted from 8,000 to 3,000 BC (Ammerman and Cavalli-Sforza 1984), or slightly less (Whittle 1996). However, in each locality the transition from hunting and gathering societies to farming communities based on fully domesticated crops and animals may have taken only a few centuries (Heun et al. 1997; Diamond 1997). This set of modifications has been termed the Neolithic transition (Ammerman and Cavalli-Sforza, 1973 Diamond 1997), and the people who carried into Europe the farming technologies are referred to as Neolithic people.

Many studies of genetic variation in Europe described broad gradients, centred in the Near East, and extending all the way to the Atlantic shores and to Scandinavia (Menozzi et al. 1978;. Sokal and Menozzi 1982; Sokal et al. 1991; Cavalli-Sforza et al. 1993; Barbujani et al. 1994, Chikhi et al. 1998; Rosser et al. 2000). Because of the striking similarity with the known routes of farming diffusion, these gradients are generally regarded as a consequence of the Neolithic transition. Altogether, they imply that that transition entailed a population expansion, and was not simply due to some form of cultural transmission, i.e. imitation or acculturation. In the latter case, we should see no genetic consequences of a process that would have implied limited migration, or none at all. In addition, theory shows that gradients are not easily established and maintained over such large distances. To account for their origin, three assumptions are necessary (Menozzi et al. 1978; Ammerman and Cavalli-Sforza 1984): (1) that the populations of early Near Eastern farmers grew in numbers (presumably because storable food surpluses were available) and gradually expanded wherever they coul find more arable land; (2) that expanding farmers kept growing in numbers as they came to occupy localities further west and north of their origin; (3) that their expansion was accompanied by limited (Ammerman and Cavalli-Sforza 1984; Rendine et al. 1986) or even no (Barbujani et al. 1995) admixture with the hunting-gathering populations encountered in the process.

By and large, therefore, the overall genetic structure of the European population is consistent with the expected consequences of a Neolithic process. If so, most ancestors of current Europeans should have lived out of Europe, in the Near East or beyond, until 10,000 years ago or so. But are there alternative explanations? And can we safely assume that every European population was affected to the same extent by the Neolithic transition? The answer to the second question is certainly: no. The relative role of Palaeolithic and Neolithic people in the formation of the gene pool of current European populations must have varied. Three main regions where Neolithic technologies developed differently have been identified based on archaeological evidence (Whittle 1996), and differences exist within these regions too. Most likely, some Palaeolithic groups were essentially unaffected by the incoming agriculturalists, admixture was the rule in some regions, and Neolithic immigrants totally replaced preexisting groups in some other regions. Although no simple, general answer is thus likely to be satisfactory, the widespread genetic gradients observed suggest that the main single process that led to the current genetic diversity was a directional expansion from the Levant, which many independent pieces of evidence suggest to locate in Neolithic times.

Additional evidence in favour of a major impact of Neolithic expansions in the peopling of Europe comes from simulation studies. Rendine et al. (1986) demonstrated that under the assumptions of dispersal from the Near East, population growth, and little admixture, clines are generated, and they are very similar to those observed in empirical studies. Barbujani et al. (1995) showed that a Neolithic dispersal associated with language diffusion accounts for the observed genetic distances in Europe better than any other competing model, and that clines are generated even if there has been no admixture between expanding farmers and hunter-gatherers, but a total replacement (such an extreme scenario seems unlikely in the light of archaeological evidence). Recent analyses of Y-chromosome diversity in Europe confirm that clinal variation is more the rule than the exception (Semino et al. 1996, 2000; Rosser et al. 2000).

In conclusion, the model of Neolithic demic diffusion has found support from various independent types of evidence, and essentially from:

(i) the presence if Europe of wide allele frequency clines;

(ii) the correlation between archaeological and genetic maps;

(iii) the simulations of a Neolithic expansion in Europe, upon the assumption of limited or no admixture with hunter-gatherers.

Recently, however, the view whereby the European gene pool largely originated in Neolithic populations of the Levant has been challenged by studies of mitochondrial DNA (mtDNA). Patterns of mtDNA variation seem to differ from the patterns described for allozymes and other protein markers. Sequence diversity is high in Europe for mtDNA, but at the global scale little geographical patterning is apparent among the samples studied (Excoffier 1990; Pult et al. 1994; Richards et al. 1996, 2000). Even populations that are known to substantially differ from most other European populations at the allele-frequency level (including the speakers of non Indo-European languages) do not greatly differ at the mitochondrial level (Simoni et al. 2000). Also, factors that have been shown to affect allele frequencies such as geographical and linguistic barriers, appear to be seldom associated with significant mtDNA sequence change, although a few exceptions exist (Simoni et al. 2000).

In a series of studies based on increasing numbers of individuals, Richards et al. (1996, 2000) analysed sequences of the mtDNA control region. They confirmed the absence of clear geographical trends and, using the method of median networks (Bandelt 1994), they were able to define several groups of evolutionarily-related haplotypes, or haplogroups. Diversity within these haplogroups was estimated, and their age (or coalescence time) was inferred, assuming a constant rate of evolution, and under the hypothesis that each haplogroup's diversity developed in Europe from a single founder. Most haplogroups were shown to be derived from ancestral sequences dating back to the Palaeolithic, which led to the conclusion that the people carrying those haplogroups arrived in Europe at that time. Because 85% of mtDNA lineages had a Palaeolithic origin, Richards et al. (1996) estimated the contribution of hunter-gatherers as being approximately 85%.

The relationship between coalescence times and population divergence has been discussed by several authors, who showed that identifying the origin of a population with the common molecular ancestor of a DNA sequence is simply wrong (Pamilo and Nei 1988; Barbujani et al. 1998; Edwards and Beerli 2000; Nichols 2001; Bertranpetit and Calafell 2001), unless it can be proved that that population originated from a very small number of genetically identical individuals (Barbujani and Bertorelle 2001). However, Richards' et al. studies, besides contributing to a deeper understanding of the phylogenetic relationships among mitochondrial haplotypes, triggered a critical review of the existing data. In particular, the relative contributions of the two gene pools, Neolithic and Palaeolithic, had never been made explicit. Ammerman and Cavalli-Sforza's (1984) initial idea seemed to be that Neolithic farmers were outnumbering hunter-gatherers. However, replying to Richards et al. (1996), Cavalli-Sforza and Minch (1997) wrote that Neolithic genes can represent approximately 26% of the current European gene pool, a figure corresponding to the percentage of the total genetic variance explained by the clinal component of genetic variation observed in their analysis (technically speaking, a principal component analysis). Accordingly, the genes contributed by Neolithic farmers should not represent the absolute majority of the current European genome.

It is not clear whether and how a genetic variance could represent an estimate of the relative contribution of two groups to a common gene pool. Be that as it may, given the approximate nature of any such figure, the 26% of Neolithic genes proposed by Cavalli-Sforza and Minch (1997) might not differ significantly from the 15% which Richards et al. (1996, 1997) proposed. However, the only way to evaluate to what extent two or more groups contributed to a hybrid population is to use explicit admixture models. These models consider one or more populations as the result of admixture among groups whose characteristics are approximated by those observed in suitable contemporary populations that are then labelled 'parental' populations. The most likely contribution of each parental group is then estimated. Because admixture is a process affecting all regions of the genome, the estimates will be more accurate when several genes will be jointly considered, a substantial advantage with respect to studies of a single marker, such as mitochondrial DNA.

Figure 2 represents the result of an admixture analysis based on a model in which both the allele-frequency differences between populations, and the DNA sequence differences between alleles, contribute to the estimation of the relevant parameters (Excoffier and Bertorelle 1998; Dupanloup and Bertorelle 2001). For this test, Isabelle Dupanloup considered three datasets, one of mitochondrial sequences (Simoni et al. 2000), and two of Y-chromosome biallelic polymorphisms (Semino et al. 2000; Rosser et al 2000). The populations were pooled according to geographical nearness, so as to define 12 potentially hybrid groups. The current Basque population was considered, according to the opinion of most investigators (see Cavalli-Sforza et al. 1994), as the most direct descendant of the people who inhabited Europe in Palaeolithic times. The best approximation to the genetic features of the Neolithic early farmers was similarly identified in current populations of the Levant (Cavalli-Sforza et al. 1994). It is evident how most of the gene pool in Southern and Eastern Europe resembles more closely the supposed Neolithic features of Near eastern populations. Genetic characteristics compatible with a major contribution of Paleolithic people are found in the Iberian peninsula and in the British isles.

The results shown in Figure 2 are only preliminary, and probably suffer from two kinds of problems. First, it is unlikely that only two groups, the ones we defined as Paleolithic and Neolithic, contri-

buted to the European gene pool. A more realistic model will have to consider the input of genes from, at least, North Africa, and Northen Asia, as has already been observed for the Y chromosome (Rosser et al. 2000). Second, a greater number of independently transmitted loci will make these estimates more accurate. Still, once the appropriate methods are used, a large Neolithic component is evident, even using the loci, especially mitochondrial DNA, whose previous analysis led to opposite conclusions.

Why have mitochondrial allele genealogies been interpreted as supporting a greater age of the European population? One answer can be found in the assumptions of the models estimating the ages of alleles. As we have stressed above, there is no guarantee that the age of a group of alleles will approximate the age of the population where they have been found, unless a group of genetically identical founders has started the entire population. Any immigration will bring new alleles which, if not properly considered, will make the allele genealogy deeper, and the estimated population's age older. If we go back to figure 1, it is evident that only in case A, where genetically similar individuals are also part of the same population, the allele's age is a good estimate of the population's age. However, as we have seen at the beginning of this article, human genealogies are seldom or never of that type. On the contrary, because of the already described distribution of genetic variances, human genealogies tend to be much more like in the B tree, with related or even identical alleles being observed over a great range of localities. The age of those alleles will contain little information on the time at which a population has been established. At present, it seems obvious that Neolithic processes have deeply affected the European's gene pool. It may not make much sense to give a global figure representing their impact, because different populations have different histories. However, if one is tempted by that exercise, there is little doubt that the best estimates of the Neolithic contribution to the global European gene pool are well above 50%. It is also important to stress that no historically or archaeologically documented process after the Neolithic seems to have had the potential for exerting genetic consequences at a continental scale. Successive demographic processes, albeit sometimes dramatic, affected only more limited areas of Europe. Therefore, the clines that were observed cannot have originated after the Neolithic period.

3. A BRIEF NOTE ABOUT THE BASQUES

The Basque population is traditionally regarded as a genetic outlier in Europe. Calafell and Bertranpetit (1994) showed that, for several loci and not only for the well-known Rh-negative allele, the Basques differ substantially from their neighbours. However, using a different collection of markers, Barbujani and Sokal (1990) concluded that it is all the northern part of the Iberian peninsula, and not simply the Basque country, that stands out as a genetic outlier. More recently, mitochondrial studies have not identified substantial differences between the Basques and the bulk of the Europeans. At the mitochondrial level, the main European outliers are other populations, notably the Saami (Sajantila et al. 1995) and the Ladin speakers of the Eastern Alps (Stenico et al. 1996). That is not surprising at all, in fact. Given what we know on human variation, i.e. on the limited differences between populations, and on the limited degree of isolation of European communities, a globally differentiated population is hard to envisage in Europe. The analysis of broader genome regions may somewhat change this view in the future, but for the time being one can safely conclude that the evident cultural and linguistic differences among Europeans, including Basques, are reflected in their genome only to a limited extent.

REFERENCES

AMMERMAN A.J. and CAVALLI-SFORZA L.L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press, Princeton

BANDELT H.J. (1994) Phylogenetic networks. *Verh. Naturwiss. Ver. Hamburg* 34, 51-72.

BARBUJANI G. and BERTORELLE G. (2001) Genetics and the population history of Europe. *Proc. Natl. Acad. Sci. USA* 98, 22-25.

BARBUJANI G., BERTORELLE G., and CHIKHI L. (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *Am. J Hum. Genet.* 62, 488-491.

BARBUJANI G., MAGAGNI A., MINCH E. and CAVALLI-SFORZA L.L. (1997) An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci. USA* 94, 4516-4519.

BARBUJANI G., PILASTRO A., DE DOMENICO S., and RENFREW C. (1994) Genetic variation in North Africa and Eurasia: Neolithic demic diffusion versus Paleolithic colonisation. *Am. J Phys. Anthropol.* 95, 137-154.

BARBUJANI G. and SOKAL R.R. (1990) Zones of sharp genetic change in Europe are also language boundaries. *Proc. Natl. Acad. Sci. USA* 87, 1816-1819

BARBUJANI G., SOKAL R.R., and ODEN N.L. (1995) Indo-European origins: A computer-simulation test of five hypotheses. *Am. J Phys. Anthropol.* 96, 109-132.

BERTORELLE G., and BARBUJANI G. (1995) Analysis of DNA diversity by spatial autocorrelation. *Genetics* 139, 811-819

BERTORELLE G. and EXCOFFIER L. (1998) Inferring admixture proportions from molecular data. *Mol. Biol. Evol.* 15, 1298-1311.

BERTRANPETIT J. and CALAFELL F. (2001) Genome versus population understanding in human geentic studies. In Donnelly P And Foley R.A. (eds.) Genes, Fossils and Behaviour, pp. 49-62. IOS Press, Amsterdam.

BERTRANPETIT J., CALAFELL F., COMAS D., PEREZ-LEZAUN A., and MATEU E. (1996) Mitochondrial DNA

sequences in Europe: An insight into population structure. In Boyce A.J. and Mascie-Taylor C.G.N. (eds) *Molecular Biology and Human Diversity*, pp 112-129. Cambridge University Press, Cambridge

BIASUTTI R. (1959) *Razze e Popoli della Terra*. Turin: UTET

BROWER A.V.Z. and BOYCE T.M. (1991) Mitochondrial DNA variation in monarch butterflies. Evolution 45, 1281-1286.

BROWN R.A., and ARMELAGOS G.J. (2001) Apportionment of racial diversity: A review. *Evol. Anthropol.* 10, 34-40.

CALAFELL F., BERTRANPETIT J. (1994) Principal component analysis of gene frequencies and the origin of Basques. *Am. J. Phys. Anthropol.* 93, 201-215

CAVALLI-SFORZA L.L., MENOZZI P., and PIAZZA A. (1993) Demic expansions and human evolution. *Science* 259,639-646.

CAVALLI-SFORZA L.L., MENOZZI P., PIAZZA A. (1994) *The History and Geography of Human Genes*. Princeton, Princeton University Press.

CAVALLI-SFORZA L.L., MINCH E. (1997) Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. *Am. J. Hum. Genet.* 61, 247-251.

CHIKHI L., DESTRO-BISOL G., BERTORELLE G., PASCALI V. and BARBUJANI G. (1998) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc. Natl. Acad. Sci. USA* 95, 9053-9058.

COHEN C. (1991) Les races humaines en histoire des sciences. In Hublin J.J. and Tillier A.M., (eds.): *Aux Origines d'Homo sapiens*, pp. 9-47. Paris: Presses Universitaries de France.

COMAS D., CALAFELL F., MATEU E., PEREZ-LEZAUN A., BOSCH E., and BERTRANPETIT J. (1997) Mitochondrial DNA variation and the origin of the Europeans. *Hum. Genet.* 99, 443-449

COON C.S. (1963) *The Origin of Races*. New York: Knopf.

DIAMOND J. (1997). Location, location, location: the first farmers. *Science* 278:1243-1244.

DUPANLOUP I. and BERTORELLE G. (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol. Biol. Evol.* 18, 672-675.

EDWARDS S.V. and BEERLI P. (2000) Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54, 1839-1854

Estoup A.L., Garnery L., Solignac M. and Cornuet J.M. (1995) Microsatellite variation in honey bee (*Apis mellifera*) populations: hierarchical genetic structure and test of the infinite alleles and stepwise mutation models. *Genetics* 140, 679-695.

EXCOFFIER L. (1990) Evolution of human mtochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J. Mol. Evol.* 30, 125-139.

EXCOFFIER L., SMOUSE P.E. and QUATTRO J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131, 479-491.

FEVOLDEN S.E. and SCHNEPPENHEIM R. (1989) Genetic homogeneity of krill (*Euphausia superba*) in the Southern Ocean. *Polar Biol.* 9, 533-539.

GOLDBERG T.L. and RUVOLO M. (1997) The geographic apportionment of mitochondrial genetic diversity in east African chimpanzees, *Pan troglodites schweinfurthii*. *Mol. Biol. Evol.* 14, 976-984.

HEUN M., SCHÄFER-PREGL R., KLAWAN D., CASTAGNA R., ACERBI M., BORGHI B., SALAMINI F. (1997). Site of Einkorn wheat domestication identified by DNA fingerprinting. *Science* 278, 1312-1314.

LATTER B.D.H. (1980) Genetic differences within and between populations of the major human subgroups. *Am. Nat.* 116, 220-237.

LEWONTIN R.C. (1972) The apportionment of human diversity. *Evol. Biol.* 6, 381-398.

LI W.H. and SADLER L.A. (1991) Low nucleotide diversity in man. *Genetics* 129, 513-523.

LIVINGSTONE F. (1962) On the non-existence of human races. *Curr. Anthropol.* 3, 297-281.

MELLARS P.A. (1992) Archaeology and the population dispersal hypothesis of modern human origins in Europe. *Phil Trans Royal Soc London* B 337, 225-234.

MENOZZI P., PIAZZA A., CAVALLI-SFORZA L.L. (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201, 786-792

MORIN P.A., MOORE J.J., CHAKRABORTY R., JIN L. GOODALL J. and WOODRUFF D.S. (1994) Kin selection, social structure, gene flow and the evolution of chimpanzee. *Science* 265, 1193-1201.

NEI M. and ROYCHOUDHURY A.K. (1982) Genetic relationships and evolution of human races. *Evol. Biol.* 14, 1-59

NICHOLS, R. (2001) Gene trees and species trees are not the same. Trends Ecol. Evol. 16, 358-364.

PALUMBI S.R., GRABOWSKY G., DUDA T., GEYER L. and TACHINO N. (1997) Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution* 51, 1506-1517.

PAMILO P. and NEI M. (1988) Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5, 568-583.

PULT I., SAJANTILA A., SIMANAINEN J., GEORGIEV O., SCHAFFNER W., and PÄÄBO S. (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. *Biol. Chem. Hoppe-Seyler* 375, 837-840

RENDINE S., PIAZZA A. and CAVALLI-SFORZA L.L. (1986) Simulation and separation by principal components of mutiple demic expansions in Europe. *Am. Nat.* 128, 681-706.

RICHARDS M., CORTE-REAL H., FORSTER P., MACAULAY V., WILKINSON-HERBOTS H., DEMAINE A., PAPIHA S., et al (1996) Palaeolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 58, 185-203.

RICHARDS M., MACAULAY V., SYKES B., PETTIT P., FORSTER P., HEDGES R., and BANDELT H.J. (1997) Palaeolithic and Neolithic lineages in the European mitochondrial gene pool: a response to Cavalli-Sforza and Minch. *Am. J. Hum. Genet.* 61, 251-254.

RICHARDS M., MACAULAY V., HICKEY E., VEGA E., SYKES B., GUIDA V., RENGO C., et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am. J. Hum. Genet.* 67, 1251-1276

ROSSER Z.H., ZERJAL T., HURLES M.E., ADOJAAN M., ALAVANTIC D., AMORIM A., AMOS W. et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* 67, 1526-1543

RUVOLO M., PAN D., ZEHR S., GOLDBERG T., DISOTELL D.R. & VON DORNUM M. (1994) Gene trees and hominoid phylogeny. *Proc. Natl. Acad. Sci. USA* 91, 8900-8904

SAJANTILA A., LAHERMO P., ANTTINEN T., LUKKA M., SISTONEN P., SAVONTAUS M.L., AULA P. et al. (1995) Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* 5, 42-52

SEMINO O., PASSARINO G., BREGA A., FELLOUS M., and SANTACHIARA-BENERECETTI A.S. (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* 59, 964-968

SEMINO O., PASSARINO G., OEFNER P.J., LIN A.A., ARBUZOVA S., BECKMAN L.E., DE BENEDICTIS G., et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* 290, 1155-1159

SIMONI L., CALAFELL F., PETTENER D., BERTRANPETIT J. and BARBUJANI G. (2000a) Geographic patterns of mtDNA diversity in Europe. *Am. J. Hum. Genet.* 66, 262-278

SOKAL R.R., and MENOZZI P. (1982) Spatial autocorrelation of HLA frequencies in Europe support Demic Diffusion of early farmers. *Am. Nat.* 119, 1-17.

SOKAL R.R., ODEN N.L., WILSON C. (1991) New genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351, 143-145

STENICO M., NIGRO L., BERTORELLE G., CALAFELL F., CAPITANO M., CORRAIN C., and BARBUJANI G. (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. *Am. J. Hum. Genet.* 59, 1363-1375.

STRINGER C. (1989) The origin of early modern humans: A comparison of European and non-European evidence. In Mellars P.A. and Stringer C. (eds.): *The Human Revolution*, pp. 232-244. Edinburgh, Edinburgh University Press.

WHITTLE A. (1996) *Europe in the Neolithic: The Creation of New Worlds*. Cambridge University Press, Cambridge.

WOLPOFF M.H., WU X. and THORNE A. (1984) Modern Homo sapiens origins: A general theory of hominid evolution involving the fossil evidence from East Asia. In Smith F. and Spencer F., (eds.): The Origins of Modern Humans: A World Survey of the Fossil Evidence, pp. 411-483. New York: Alan Liss.

TABLE 1: Some systems of classification of human races (compiled from Biasutti 1959, Coon 1963, and Cohen 1991).

Author	Nr of races	Races proposed
Linnaeus (1758)	4	Homo americanus, H. Europaeus, H. asiaticus, H. afer
Blumenbach (1775)	5	Caucasian, Mongolian, Ethiopian, American, Malay
Cuvier (1828)	3	Caucasoid, Negroid, Mongoloid
Huxley (1875)	4	Xanthochroid, Mongoloid, Negroid, Australoid
Deniker (1900)	29	
Weinert (1935)	17	
Von Eickstedt (1937)	38	
Biasutti (1959)	53	
Coon (1963)	5	Caucasoid, Negroid, Australoid, Congoid, Capoid

TABLE 2: Percentages of overall genetic variances at three levels of population subdivision, as estimated in studies based on protein (P), DNA RFLP (R) or DNA microsatellite (M) markers; H are multilocus RFLP haplotypes. Latter (1980) gave three sets of values, based on three different statistical approaches.

Authors	N Loci	Type	Within samples within groups	Between samples	Between groups
Lewontin (1972)	17	P	85.4	8.3	6.3
Latter (1980)	18	P	83.8 – 87.0	5.5 – 6.6	7.5 – 10.4
Ryman et al. (1983)	25	P	86.0	2.8	11.2
Barbujani et al. (1997)	79	R	84.5	3.9	11.7
"	30	M	84.5	5.5	10.0
"	16	H	83.6	8.4	8.0
"	109	M+R	84.4	4.7	10.8

Fig. 1: Examples of evolutionary trees summarising genetic variation in a geographically structured species (A) and in a species where geographical groups do not differ much from each other (B). Haplotypes of the same geographical origin are represented by the same symbols (stars, circles and triangles).

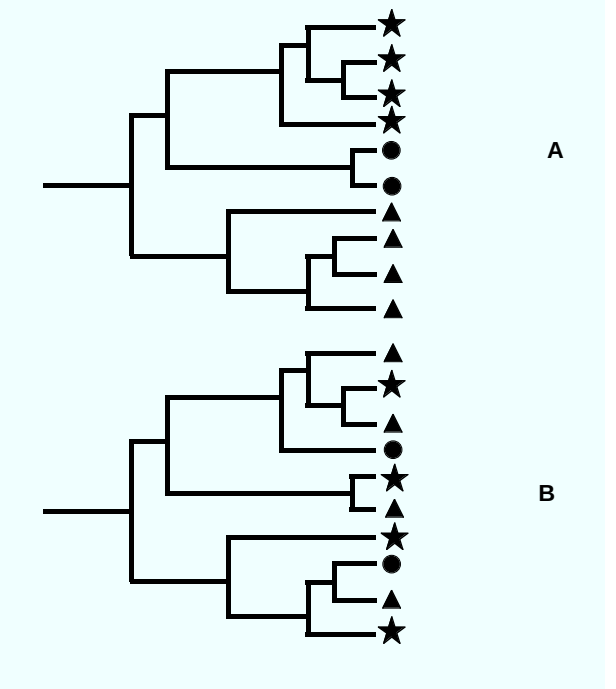

Fig. 2: Estimated contributions of Paleolithic colonizers (white sections of the pies) and Neolithic early farmers (black sections) to the European gene pool. Figure courtesy of Isabelle Dupanloup.