The basque population and ancient migrations in Europe

Luca L. Cavalli-Sforza*

KEYWORDS: Rh genes dynamics, Origin or Basques, Origin of Indoeuropean Languages, Neolithic spread in Europe, Cultural conservation.

SUMMARY

Basques have a unique position in Europe not only because of their language but also because of some genetic differences with their neighbors. One of the most characteristic ones is the relatively high frequency of Rh-negative genes. The simplest interpretation of the origin of Basques is that they were in their present location already in upper paleolithic and mesolithic times, when they probably occupied a wider area than now. They may represent the most direct descendants of Cro-Magnon people. Early neolithic farmers began expanding in all directions from the Middle East or Asia Minor around 9000 years ago or earlier; they mixed with local mesolithic hunter-gatherers. Those who came to Europe may have spoken Indoeuropean languages. If this is correct, the Basque language may be the only one that survived among those spoken earlier. Other non-Indoeuropean languages spoken in Europe came later from the outside. The reasons for conservation of the Basque language and culture and their interaction are discussed.

THE GENETIC BACKGROUND

If one looks at a tree of the genetic origin of European populations, one finds considerable homogeneity; but four people stray in different directions: Lapps, Icelanders, Sardinians and Basques. For the first three, we have historical or linguistic clues that help us understand at least part of the story. No such clues are available for Basques: here, on the contrary, the linguistic evidence seems to obscure the picture. The Basque language is described as an isolate, with no living relatives. While there are some linguists who take exception to this conclusion, as we shall discuss later it remains reasonable to conclude that the Basque language is a relic. It, or languages ancestral to it, may have once been spoken over vast regions. All historical data, however, attest to the ancient presence of the Basque language in the region, so that the Basque people may have been in their present location for a very long time.

Given their linguistic isolation, it is not at all surprising that Basques also show genetic peculiarities that indicate their relative genetic isolation from their neighbors. The existing data have been summarised and discussed by Alberto Piazza in this symposium, and I will refer readers to his excellent paper. One of the outstanding genetic differences from other populations is in the frequency of Rh genes; this is particularly striking as the evolutionary dynamics of this gene is rather peculiar. It is well known that Rh positives will tend to increase if they are more than 50% of the population, or decrease if they are less than 50%. Obviously, the same can be said of the alternative type of genes, the Rh negatives. Basques tend to be slightly above 50% Rh negatives (fig. 1) and, therefore, Rh negatives should increase while Rh positives should decrease and eventually disappear. Other Europeans cluster at 40% negatives or below, and should decrease (fig. 2). How does Europe remain so close to the 50% point, the most unstable of all?

A possible explanation is that once Europe was inhabited mainly or entirely by Rh negative people, and later an invasion of predominantly or entirely Rh positive people took place. The pressure of invaders was less important in the present day Basque area than in the rest of Europe; and therefore, the mixture of earlier inhabitants to invaders was in favor of the former. The existence of an unstable point at 50%-positives: 50%-negatives has helped maintain the gene frequencies near the Western Pyrenees around the present value. It is of interest to note in fig. 2 that the evolutionary process of gene frequencies is fairly slow, so that in the absence of any migration it takes 10- to 25,000 years (400-1,000 generations) to have important changes in gene frequencies if at the start they are near 50%.

In order to discover when this «Invasion» of Europe might have happened, we must look to the archaeological and historical records. It is believed that Neanderthals disappeared from Europe, in particular Western Europe, between 31- to 33,000 years ago. (F.C. Howell, 1984). There is no reason to think

* Departament of Genetics. Stanford University. Stanford, California USA.
that modern people in Western Europe have any special relationship to Neanderthals. Human remains north of the Pyrenees define the Cro-Magnon type, who is clearly modern post-Neanderthal.

Lascaux caves (dated around 15- to 17,000 BP) are believed to have been inhabited by Cro-Magnon-like people. The similarity of art work north and south of the Pyrenees supports the inference that north and south of the West Pyrenees the people were similar. The poverty of the fossil record south of the Pyrenees does not offer much further clarification.

Figure 1. A. E. Mourant was among the first to draw the attention on the high frequency of Rh negatives among Basques and the hypothesis that Basques may be proto-europeans, least admixed with later immigrants. This figure is abstracted from the geographic distribution of the Rh-negative gene published in A.E. Mourant, A.C. Kopec, K. Domaniewska-Sobazak. The Distribution of the Human Blood groups, Oxford University Press, London 1976.

We will assume that the entire Basque region at that time may have been inhabited by a relatively homogeneous population, and will search for evidence of later migrations.

On the basis of our work, the most important migration to Europe in the last 10,000 years was that of Neolithic people originating in the Middle East. It would be more correct to say they originated in Asia Minor, since the first appearance of the Neolithic in Europe, at the beginning of the preceramic type is in Macedonia. But undoubtedly at the time there was communication between the Middle East and Asia Minor. The theory of Neolithic expansion to Europe has been presented in some detail by Ammerman and Cavalli-Sforza (1984). The theory was engendered by the a priori consideration that the development of food production lifted the constraints on population density existing in foraging (hunting-gathering) populations. Demographic increase was generated and was responsible for the geographic expansion of an agricultural population, which proceeded slowly to occupy available niches in vast regions, from Europe to Africa and Asia.

Obviously, there were other independent agricultural developments in many parts of the world; but the Middle Eastern Neolithic was particularly early, rich and complex, including both plant and animal breeding. The crops domesticated in the Near East (essentially wheat and barley) proved capable of growth in a very wide area.

The evidence for this theory comes from two sources. First, the archeological data (Ammerman and Cavalli-Sforza, 1971, 1973). As is well known, archeology cannot prove or disprove migrations; and in fact, after the excessive «migrationist» enthui-
siasm of archeologists of the thirties, there has been a widespread tendency to completely avoid migrationist explanations. The evidence, however, shows quite clearly that following the introduction of agriculture, food products typical of the Neolithic economy were found at greater and greater distances from the place of origin, the Middle East, at later and later dates. It has been possible to map geographically the advance of Neolithic agriculture in Europe by plotting the times of its first arrival in well known archeological regions (fig. 3). Also, by plotting the first time of arrival to a given spot in Europe versus its geographic distance from the origin, one observes a substantially constant but slow rate of radiation of agriculture from the the Middle East (on average one Km per year as the crow flies). It was also shown (see AMMERMAN and CAVALLI-SFORZA, 1984) that this radial rate of advance is compatible with the demographic increase of Neolithic populations. Therefore, the hypothesis that farmers themselves spread throughout the area, and not simply the idea of farming is not in contradiction with the archeological data.

In the absence of direct evidence on migration which archeology cannot provide, our attention was turned to the genetic markers of human populations. Unfortunately, no fossil data are satisfactory for our purpose. Skeletal measurements are unreliable, being highly sensitive to dietetic and other environmental conditions, and are, in any case, unavailable in sufficient numbers. Common genetic markers, whose inheritance is clearly proved, cannot be detected on fossil material except in very rare and insufficient cases. The only remaining possibility is that of examining living populations, on which there is a wealth of genetic data. The hypothesis will have to be made, and tested, that there were no important changes in genetic markers in the last millennia before the present.

No single gene is sufficiently informative for this analysis, and a multivariate approach was therefore necessary. A new method introduced by Menozzi, Piazza and Cavalli-Sforza (1978) allowed researchers to create geographic maps of genetic patterns shown by European populations. The most important pattern (the geographic map of the first principal component of gene frequencies) shows a major gradient with a center in the Middle East, in very close agreement with our hypothesis of a «demic» diffusion (i.e.: geographic expansion of the population)

Figure 3. Computer map of the dates of first arrival of neolithic formers from their area of origin. Recalculated, from Ammermann and Cavalli-Sforza, 1984.
of early Neolithic farmers from the Middle East. The next two lower principal components (p.c.) generated different patterns. The geographic map of the second p.c. demonstrates a gradient East-to-West. It probably describes the effect of migration from Central Asia, which were numerous and almost continuous from 5,000 years ago to recent times. The third p.c. displays an almost circular gradient with a center in Poland-Ukraine. It may describe a center of radiation of Indo-European speakers, postulated by Gimbutas (1973) to have started in early Bronze Age (5,000 years ago approximately). It is difficult to exclude that it can also represent the effects of barbarian invasions at the end of the Roman Empire, which may have had a geographically similar origin.

The analysis of the genetic patterns of modern European populations thus seems to support entirely the hypothesis that there was a slow migration of Neolithic farmers. It should be stressed that there certainly was not a complete replacement of old occupants by the new ones; but rather, the migrating farmers must have intermingled with the local inhabitants, the Mesolithic occupants of Europe, an admixture which probably varied in importance from place to place. Only in this way could the observed genetic gradients arise, as described by a diffusion model with interbreeding of the faster-growing farmer population with the lower-density Mesolithic local populations (see Ammerman and Cavalli-Sforza, 1984; Rendine, Piazza and Cavalli-Sforza, 1986). To strengthen these conclusions, it was necessary to test the validity of the Menozzi et al method of detecting old migration patterns with simulation experiments (Rendine et al, 1986). This helped to prove on one hand that the method of geographic maps of principal components can separate old migrations, and on the other that the later events do not usually blur the picture. It is, therefore, reasonably safe to use genetic data from modern populations for an analysis of major ancient migrations. It should be remembered that in the last two or three millennia Europe already had a fairly high population density, and that therefore latecomers were usually responsible for minor contributions in terms of relative population numbers. Although their genetic impact might be felt with increased genetic knowledge allowing a higher level of resolution, they could hardly alter the general patterns established by the major migrations into Europe, from 9,000 years ago (the beginning of the Neolithic infiltration from West Asia) to the latest Asian invasions of the first millennium A.D.

Therefore, the first migration into Europe by Neolithic farmers may have been, in part, replacing the older European substratum; and there still remains a genetic residual of these «protoeuropeans», especially in the regions with higher Rh- areas. Of these, the Basque region is the most important, followed by the north of Europe. Naturally, these preneolithic Europeans from these two distant areas could not be expected to show much similarity. The density of most upper Paleolithic and Mesolithic populations was low enough that they must have had major genetic differences due to random genetic drift. Piazza points to an overall similarity between Basques and Sardinians. For Rh genes, however, Sardinians and Basques are at opposite extremes. Genetic differences of this magnitude are not unexpected between populations separated for many millennia. They may also have been amplified by the later Phoenician colonization of Sardinia. When isolated populations of Central and Southern Europe will be better known, a clearer picture of the genetics of preneolithic Europe may emerge.

In conclusion, the genetic evidence leads to the inference that Basques may have descended from a Mesolithic population which has survived the impact of genetic admixture with later comers better than other old European populations. Among these arrivals, the first and most important were probably the Neolithic people which entered Europe from Asia Minor in the ninth millennium before the present. The Neolithics flourished in Europe, thanks to their economy of food production, and arrived slowly perhaps by way of the Mediterranean to the Basque region. In part, they may have come also from Africa.

The Mesolithic populations of the Basque region had presumably been in situ for many millenia prior to the arrival of Neolithic farmers and culture in Spain. By the time of their arrival in Spain, three thousand years after their first passage to Europe, the Neolithic genotype had probably been seriously diluted through mixing with local Paleolithics or Mesolithics in the approximately one hundred and fifty generations during which the slow Neolithic diffusion toward the West took place.

THE LINGUISTIC INFORMATION

In our search for the ancient origins of the Basque people, we have inferred that the Basques may have been in place since Mesolithic times and mixed somewhat less with latecomers than other European populations. We carefully note that this makes no assumption as to the language they spoke prior to the arrival of Neolithics. We know, however, that names of places and of populations show Basque-speaking people once covered a much broader area (see J. Allègres, this symposium).

Let us return to linguistics; does it truly offer no help? There are two major hypotheses: that there is a connection of Basque 1) with South Caucasian
languages, of which Georgian is a preeminent one, and 2) with Iberians. Iberian origin offers very little help, the origin of Iberian itself being obscure. While Caucasian languages are many and show considerable variation, there are, however, some clear lexical resemblances and important similarities in verb conjugation with Basque (see Allières, 1986).

The considerable geographic distance between the Caucasus and the Basque country may seem puzzling. In this respect, it is useful to draw a parallel with the residual geographic distribution of languages of one linguistic family once spread over a wide area. Celtic languages, for example, are spoken today only in the most peripheral parts of the British Isles (essentially Scotland, Wales and Ireland) having been replaced by Anglo-Saxon languages spoken by invaders from Central Europe. Celtic languages were spoken in France and in the whole of Central Europe, including Northern Italy for more than five centuries. (The Celtic language spoken in Brittany, Northwestern France, may have been reimported there later from Cornwall, where a Celtic dialect was spoken and is now extinct). There are other instances in which a group of languages, when supplanted by another, tends to persist in the peripheral areas. A most dramatic example is that of Dravidian languages, spoken today almost entirely in Southeastern India. Earlier, it was spoken throughout the north of India, prior to the arrival of the Indo-European languages with the Aryans. The invaders probably came from the Asian steppes; and before entering India, they most likely caused the fall of the Harappan civilization of the Indus Valley in Pakistan around 1750-1500 B.C. There are indications that the Harappans spoke a Dravidian language; in any case, there is still one Dravidian language spoken in Pakistan today (Brahui). Possibly Dravidian languages were also spoken throughout most of Iran. It has recently been shown that Elamite, spoken in Southwestern Iran around 3000-2500 B.C. and now extinct, belonged to the Dravidian family which now, because of this reason, is called Elamo-Dravidian (see a summary and references in Ruhlen). Elamite is known to us because it was written in cuneiform script. It is likely that Dravidian languages were spoken over a vast expanse, from Western Asia to almost all of India. As in the case of the Celtic branch of the Indo-European group, the Dravidian family survives today only at the periphery of the original expansion, in addition to a few pockets in Pakistan and Northern India.

The spread of farmers from the Middle East that started in Neolithic times around 9,000 years ago probably took place in all directions: we have been able to validate on archeological and genetic bases the spread to Europe (i.e.: in the Northwestern direction). Data for studying the possible expansions in other directions are much less satisfactory, both at the archeological and at the genetic level. But the linguistic evidence should be kept in mind. In Europe, with the exception of Basque and the languages of Uralic origin in Scandinavia and Hungary (Lapp, Fin, Hungarian and others), all languages belong to the Indo-European family (see fig. 4).


The spread of Indo-European languages was certainly due to a series of different radiations that may have taken place at widely different times. The branch of Indo-European languages spoken today in Iran and India (Indo-Iranian) arrived in place fairly late (the second millennium B.C.) with the Aryan invasion. This migration most probably originated in the Asian steppes and must have been responsible for suppressing or confining to peripheral areas Dravidian languages previously spoken much more widely. From the evidence given before, it is possible that Dravidian languages were originally spoken in the Near East and Asia Minor (along with other languages) and were spread eastwards by Neolithic farmers. The languages spoken in the Middle East, Ara-
bria, North Africa and Ethiopia, belonging to the Afro-Asiatic family, may well have spread South and West from Mesopotamia with Neolithic people, more or less at the same time as Neolithic people spread to Northwestern Europe, and toward the Southeast, to Iran and India.

Recently, Russian linguists (in particular, ILLYCH-SVITICH and DOLGOPOLSKY; see RUHLEN, 1987) have become very interested in Nostratic languages, a group of families including Altaic, Uralic, South Caucasian, Indo-European, Afro-Asiatic and Dravidian in the Near East around 14,000 year ago. Independently, J. GREENBERG (see RUHLEN, 1987) suggested an Eurasian superfamily overlapping with Nostratic, but excluding the South Caucasian and Dravidian families and including Chuckchi-Eskimo.

The inclusion of the South Caucasian family in Nostratic is especially interesting because of the postulated relationship with Basque (see ALLIERES, 1986). The Caucasus is a mountainous area in which populations were less exposed to invasions that would have affected their culture and language, while all areas surrounding them had a heavy turnover of people or languages.

The Nostratic hypothesis may need further refinement before it is widely accepted. It is, however, refreshing that the creation of linguistic superfamilies seems to find more supporters than it ever did. When trying to connect the linguistic and genetic facts, the simplest hypothesis that comes to mind is that languages of Basque type, with area of origin unknown, were spread over a vast area in Europe and West Asia prior to the Neolithic radiation originating from West Asia, and that this radiation spread to Europe old Indo-European languages originally spoken in Asia Minor. One can make this hypothesis even more suggestive by adding to it that the Neolithic radiation starting in the tenth millennium before the present may have spread at least three major linguistic families from a putative common center of origin in West Asia (fig. 5): Indo-European initially at least towards the Northwest and therefore Europe, Afro-Asiatic towards the South and Southwest and therefore Arabia and North Africa, and Dravidian towards the Southeast (Iran and India). This differs from the Nostratic hypothesis mostly in a somewhat later timing of the expansion, and in not taking a definite position on the other families connected with the Nostratic hypothesis. The advantage of linking some linguistic expansions from Western Asia with the spread of the Neolithic is that there is archaeological and genetic evidence for this expansion. Other languages of unknown origin were

Figure 5. A hypothesis on the spread of tree linguistic families with neolithic farmers. The areas shaded are those of origin of neolithic plant and animal domestication: the Fertile Crescent (shaded horizontally) and its Anatolian appendix (oblique shading).
also spoken in late Neolithic and early bronze age in Mesopotamia and neighboring regions, like Sumerian, Hurrian, and Urartian. Greater understanding of their relationships could help draw a more complete picture than is possible today. Also of interest is that the Afro-Asiatic families may well have penetrated into Spain from Northwestern Africa, probably with early Neolithics. If so, then the presumed connections of the Iberian language with Berber (Afro-Asiatic) languages from Northern Africa, as suggested by the common «ber» (ALLIÈRES, 1986), may find a simple explanation. The Iberian language may also have contributed to the origin of Basque.

The Nostratic hypothesis that Indo-Europeans spread from Western Asia is based on the recognition of early borrowings between Indo-European and Semitic (Shevoroshkin, personal communication). If Indo-European actually spread from Western Asia at the time of the Neolithic radiation, then the view that Basque was a mesolithic language of older origin seems very reasonable. However, it is difficult at this stage to completely reject an alternative hypothesis: that a protobasque language was spoken by early Neolithics and spread by them to Europe, where it is still spoken only in the most peripheral areas. In this case, Indo-European languages entered later from other sources, as postulated for instance by Gimbutas. Nevertheless, even accepting Gimbutas center of origin (the Kurgan culture in Southern European Russia), it is not necessary that this is the primary center of expansion of Indo-European languages; it could be secondary. The coincidence of the time of the postulated spread of Indo-European languages, about 5000 years ago, from glottochronological sources, and that of the Kurgan culture to which she has linked her postulated spread of people is unconvincing. The calibration curve of glottochronology is far from linear for older times and may cause serious underestimation (Kruskal et al).

Investigations to distinguish the two hypotheses, namely if Neolitics spoke 1) an early Indo-European or 2) Basque language, and therefore Mesolitics from the Basque region spoke 1) an early Basque or 2) some other unknown language, might profit from a study of the origins of Basque words indicating cereals and early agriculture and animal breeding practices and tools. In addition to linguistics, further ethnographic study of legends, etc. could be targeted specifically to distinguish if Basque culture conserves elements clearly related to pre-neolithic farming. My own preference goes to the Mesolithic hypothesis of Basque language origin, which seems in greater agreement with the general picture, even if I cannot find hard evidence to exclude the second or other hypotheses.

**WHY DID BASQUES MAINTAIN THEIR LANGUAGE?**

It seems likely that the Basque language has local roots which may be 10- to 15,000 years old if the first hypothesis just mentioned above is correct, and in any case not younger than 6000 or 7000 years ago. A very interesting question is why it was maintained. This also raises a problem of interaction between sociological and linguistic factors.

This example of extreme conservation of language and probably of other cultural aspects as well indicates that very powerful mechanisms of «cultural transmission» have been at work. In a mathematical study of these mechanisms, in collaboration with M. FELDMANN (summarised in our joint 1981 book), we find that two mechanisms in particular favor cultural conservation: vertical transmission (from parent-to-child) and social pressure. We have called this type of transmission «many-to-one», because many individuals influence every individual of the social group in a coherent fashion.

We have shown that vertical transmission tends to maintain the cultural state of a population unaltered, without affecting existing individual differences. The reasons are very similar to those that make biological characters extremely stable over generations. One should emphasize, however, that the traits in question are not biological or biologically determined, but determined by direct or indirect teaching of values, customs, beliefs, attitudes and behaviors.

The second mechanism of transmission: social pressure by other members of the social group, which is also powerful and conservative, leads not only to conservation but also to uniformity by discouraging innovation in cultural traits of which the group promotes the conservation. In a society in which family ties are especially powerful and which is closely knit (i.e.: in which there is great respect for the opinion of members of the group even outside that of closer relatives), social pressures are especially powerful.

One would expect that these two mechanisms have been very active in maintaining both language and culture in the Basque country. It is unlikely that there exist quantitative data on these transmission mechanisms, which have been studied only to a very limited extent in any population (see e.g. HEWLETT and CAVALLI-SFORZA, 1986). This type of analysis should, however, be encouraged; and it would be very interesting if it could be carried out comparatively also on other neighboring populations that have or have not shown equal conservation of certain social values and norms. It is worth noting that the description of the Basque family (ALLIÈRES, 1986) accords at least qualitatively with some of the expectations set forth above. The Basque family is described as patriarchal, with strong attachment to
the house and the land, and an overall democratic structure. This description also accords with that of the family structure in the whole of the southwest of France, as demonstrated by studies of demographic parameters which can be expected to be influenced by family patterns (Le Bras and Todd, 1981). These authors note that family structure in France follows a pattern which is anterior to the barbarian invasions at the fall of the Roman empire. The family structure pattern might, in fact, be much older. For the reasons I briefly summarised in speaking about cultural transmission, one can expect that family structure itself is highly conserved for two reasons: 1) it must be transmitted vertically given that things about the family are learned from parents and close relatives, and 2) the family is itself a social group, although small, but certainly very powerful; it, therefore, acts in cultural transmission by the «many-to-one» mechanism which is like the vertical one very effective in cultural conservation. One demonstration of this fact is to be found in an analysis of ethnographic data in Subsaharan Africa (Málatti, Guglielmino, Viganotti and Cavalli-Sforza, 1983).

In the paper given by Allières at this symposium, we notice that toponomastic studies indicate that the Basque speaking area in France was much wider than its present extension would indicate; a similar situation is likely to be true in Spain. The area that may have been speaking Basque in France, according to the map given by Allières, is fairly similar to that of the patriarchal democratic family structure in the southwest of France.

In conclusion, it seems reasonable to infer that a common culture has helped maintain Basque language, and the common language has probably helped preserve important aspects of the original culture. It is likely that the strength of family ties in Basque society are responsible. Probably national pride is also a strongly transmitted, highly conserved, value; and substantial satisfaction with Basque social life must have helped maintain it, and with it the language. Also, common culture and language have certainly affected the genetic population structure of the Basques, limiting exogamy, and thus helping to maintain at least to some extent Basque genetic identity.

NOTE ADDED IN PROOFS

A recent book by C. Renfrew (Archeology and Language, Cambridge Univ. Press 1987) proposes the idea that Indo-European languages spread with early farmers from the Middle East and Anatolia towards Europe, Iran and India. This differs from the hypothesis here made, that it was Dravidian languages which spread in the eastern direction.

BIBLIOGRAFIA

ALLIÈRES, J.

AMMERMAN, A.J. and CAVALI-SFORZA L.L.

CAVALI-SFORZA, L.L. and FELDMAN, M.

DOLGOPOL’SKIJ, A.B.

GIMBUTAS, M.

GREENBERG, J.
1987. Citation in Ruhlen.

HEWLETT, B.S. and CAVALI-SFORZA, L.L.

HOWELL, F.C.

ILLYICH-SVITICH, V.M.

KRUSKAL, J.B., DYEN, I. and BLACK, P.

LE BRAS, H. and TODD E.
MATESSI, G., GUGLIELMINO, C., VIGANOTTI, C. and CAVALI-SFORZA, L.L.


MENOZZI, P., PIAZZA, A. and CAVALI-SFORZA, L.L.


RENDINE S., PIAZZA, A., and CAVALI-SFORZA, L.L.


RUHLEN, M.