DIFFUSION OF GENES AND LANGUAGES
IN HUMAN EVOLUTION

ALBERTO PIAZZA

Dipartimento di Genetica, Biologia e Biochimica,
Università di Torino,
via Santena 19, 10126 Torino, Italy
alberto.piazza@unito.it

LUIGI CAVALLI SFORZA

Department of Genetics,
Stanford University,
Stanford, CA 94305,USA
cavalli@stanford.edu

In a study by Cavalli-Sforza et al. (1988), the spread of anatomically modern man was reconstructed on the basis of genetic and linguistic pieces of evidence: the main conclusion was that these two approaches reflect a common underlying history, the history of our past still frozen in the genes of modern populations. The expression `genetic history' was introduced (Piazza et al. 1988) to point out that if today we find many genes showing the same geographical patterns in terms of their frequencies, this may be due to the common history of our species. A deeper exploration of the whole problem can be found in Cavalli-Sforza et al. (1994). In the following, some specific cases of structural analogies between linguistic and genetic geographical patterns will be explored that supply further and more updated information. It is important to emphasize at the outset that evidence for coevolution of genes and languages in human populations does not suggest by itself that some genes of our species determine the way we speak; this coevolution may simply be due to a common mode of transmission and mutation of genetic and linguistic units of information and common constraints of demographic factors.

1.   The Genetic Analysis of a Linguistic Isolate: The Basques

The case of the Basques, a European population living in the area of the Pyrenees on the border of Spain and France who still speak a non-Indo-European language, is paradigmatic. What are the genetic relations between the

Basques and their surrounding modern populations, all of whom are Indo-European speakers?

Almost half a century ago it was suggested (Bosch-Gimpera 1943) that the Basques are the descendants of the populations who lived in Western Europe during the late Paleolithic period. Their withdrawal to the area of the Pyrenees, probably caused by different waves of invasion, left the Basques untouched by the Eastern European invasions of the Iron Age. In their study of the geographic distribution of Rh blood groups, Chalmers et al. (1948) pointed out that the Rh negative allele, which is found almost exclusively in Europe, has its highest frequency among the Basques.

Chalmers et al. hypothesized that modern Basques may consist of a Palaeolithic population with an extremely high Rh negative frequency, who later mixed with people from the Mediterranean area. In more recent times genetic analyses have produced the following conclusions:

(a) Mitochondrial and Y-chromosome DNA polymorphisms support the idea that the Basques are genetically different from the other modern European populations (Richards et al., 2000, 2002; Semino et al. 2000).

(b) Mitochondrial and Y-chromosome DNA polymorphisms support the idea that the Basques are the descendants of a Palaeolithic population (Richards et al., 2000, 2002; Semino et al. 2000). The main haplogroups contributing to the European mitochondrial geography are H, pre-V, and U5. Haplogroup H is the most frequent haplogroup in both Europe and the Near East but occurs at frequencies of only 25% 30% in the Near East and the Caucasus, whereas the frequency is generally 50% in European populations and reaches a maximum of 60% in the Basque country. The age ranges of the mitochondrial founders of these lines are mostly palaeolithict: specifically the age ranges of the mitochondrial haplogroup V which is found at the highest frequency among the Basques and the Saami are pre-neolithic. In agreement with the suggestion proposed to explain the distribution of mtDNA haplogroup V (Torroni et al. 1998), the distributions of Y chromosome groups R* and R1a have been interpreted by Semino et al. (2000) to be the result of postglacial expansions from refugia within Europe. European mtDNA estimates the Neolithic component in the Basques to be the lowest for any region in Europe. Although the criteria used to identify Near Eastern founder types are somewhat heuristic and involve many assumptions, the relative number of types in different European populations should still be informative, and the Basque component, estimated at 7%, clearly lies outside the distribution for

the rest of Europe, estimated to range between 9% and 21% (Richards et al. , 2000).

(c) The linguistic hypothesis originally put forward by Trombetti (1926) that Basques share a common ancestry with the modern Caucasian speaking people living in the northern Caucasus (see Ruhlen 1991) to form the Dene-Caucasian linguistic macrofamily according to Greenberg is in agreement with some genetic evidence: Wilson et al. (2001) report that the paternal ancestors of modern Basques could have shared a common genetic origin with Celtic speaking populations. In fact, the Y chromosome complements of Basque- and Celtic-speaking populations are strikingly similar. The similarity and homogeneity of the Basque, Welsh and Irish samples suggest one of two explanations: (i) pre-agricultural European Y chromosomes were homogeneous or (ii) there was a specific connection between the Basques, the pre-Anglo-Saxon British, and the Irish. With regard to the latter hypothesis, it is interesting that a northward expansion from a glacial refugium in Iberia has been postulated from the diffusion of Magdalenian industries (Otte et al., 1990) and patterns of Y-chromosome (Semino et al., 2000) and mtDNA variation (Torroni et al., 1998). More detailed investigation of the genetic diversity present in and around Europe may allow these hypotheses to be distinguished.

2. Coevolution of Genes and Languages: The Origin of Indo-European

Barbujani and Sokal (1990) found a correlation between linguistic and genetic boundaries in Europe. In the majority of cases (22 out of 33) there were also physical barriers that may have caused both genetic and linguistic boundaries. In nine cases there were only linguistic and genetic boundaries but not physical ones: three of them (northern Finland vs. Sweden, Finland vs. the Kola peninsula, Hungary vs. Austria) separate Uralic from Indo-European languages. It remains to be determined whether in these cases linguistic boundaries have generated or enhanced genetic boundaries, or if both are the consequence of political, cultural, and social boundaries that have played a role similar to that of physical barriers.

The problem of the origin of the Indo-European linguistic family and of the people speaking its languages has roused much more interest over the last years than in earlier times partly owing to the book by Renfrew (1987), who suggested that farmers, beginning to spread from Anatolia around 9,000 years ago, spoke Indo-European languages. His hypothesis was based on the suggestion originally

put forward by Ammerman and Cavalli-Sforza (1984) that the spread of Neolithic farming from the Fertile Crescent was due to the spread of the farmers themselves and not only of the farming technology, and on the consideration that migrating people retain their language, if at all possible. Renfrew's hypothesis was criticized by most Indo-European linguists (for a review, see Mallory 1989, Lehmann 1993: 283-8) and did not fare well when contrasted with earlier hypotheses, now identified with the name of another archaeologist, Marjia Gimbutas (1985), that Indo-Europeans migrated to Europe from the Pontic steppe area of south Russia from Dniepr to the Volga (which she called `Kurgan' from the Russian name of mounds covering the graves), beginning with the early Bronze Age, that is, around 5,500 years ago.

Genetic data cannot give strong evidence on dates of migration, especially since the `Kurgan' area, one of the largest pre-historic complexes in Europe, probably remained very active in generating population expansions for a long time after the Bronze Age. In that area we find at c. 6,000 years ago the Sredni-Stog culture, later (5,500-4,500 years ago) the Yamnaya cultures (formerly called pit-grave cultures) which stretched from the Southern Bug River over the Ural River and which dates from 5,600 to 4,200 years ago. From about 5,000 years ago we begin to find evidence for the presence in this culture of two and four-wheeled wagons (Anthony 1995).

Genetic data on European populations using blood typing (Piazza et al. 1995) and Y-chromosome DNA markers (Semino et al. 2000) have strongly supported a centre of radiation in the Ukraine. It has been suggested (Cavalli-Sforza et al. 1994, Piazza et al. 1995) that the hypotheses of Renfrew and Gimbutas should not be treated as mutually exclusive; they may be compatible, as Schrader anticipated as long ago as 1890: `the Indo-Europeans practiced agriculture at a site between the Dniepr and the Danube where the agricultural language of the European branch was developed' (quoted from Lehmann 1993, p. 279). The settling of the steppe by Neolithic farmers must have occurred after the beginning of their migration from Anatolia, and if the expansions began at 9,500 years ago from Anatolia and at 6,000 years ago from the Yamnaya culture region, then a 3,500-year period elapsed during their migration to the Volga-Don region from Anatolia, probably through the Balkans. There a completely new, mostly pastoral culture developed under the stimulus of an environment unfavourable to standard agriculture, but offering new attractive possibilities. Our hypothesis is, therefore, that Indo-European languages derived from a secondary expansion from the Yamnaya culture region after the Neolithic

farmers, possibly coming from Anatolia and settled there, developing pastoral nomadism.

A new treatment of the problem has been given in a still unpublished analysis (Piazza et al., but see Cavalli-Sforza, 2000 where main results are anticipated) of a set of lexical data (200 words) in 63 Indo-European languages published by Dyen et al. (1992). From a linguistic distance matrix whose elements are the fraction of words with the same lexical root for any pair of languages and its transformation to make the matrix elements proportional to time of differentiation, we were able to reconstruct a linguistic tree. The root of the tree separates Albanians from the others, with a reproducibility rate (the error in reconstructing the tree) of 71 percent. The next oldest branch is Armenian. The simplest interpretation is that the language of the first migrant Anatolian farmers survives today in two direct descendants, Albanian and Armenian, which diverged from the oldest pre-Indo-European languages in different directions but remained relatively close to the point of origin.

If we give to the first split the time depth of the beginning of the expansion of the pre-Indo-European Anatolian farmers, about 9,000 years ago, we can then calculate that the origin of the European branch dates to about 6,000 years ago. The four major branches (pre-Celtic, pre-Balto-Slavic, pre-Italic, pre- Germanic) may correspond to some extent to different migratory waves, but archaeological dating is too scanty to provide unambiguous associations. It is reasonable to suggest that a first migration corresponds to the first branch, the pre-Celts (6,000 years ago, according to the tree), who settled first and went further west. Their only linguistic remnants are still alive today at the extreme of their original range. They profited from being among the first to develop an Iron Age culture, and were able to develop a wide community that spoke their language. Before Roman rule they spread to half of Europe, extending from Spain to France, most of the British Isles, northern Italy, and central Europe.

Very recently Gray and Atkinson (2003) have analyzed the same data set. They generated a tree of 87 languages which can be compared with our tree. We eliminated a small number of modern languages of the Dyen et al. set (1992), and Gray and Atkinson added interesting information on three extinct languages, Hittite and Tocharian A and Tocharian B, which we did not include. Their inclusion may have the advantage of providing some support for the root, but the noticeable shortening of the Hittite branch in their tree introduces some doubt on its usefulness. We believe it is worth discussing the differences between the two trees in some detail, because they are relevant to the problem of Indo-European origins and also to the general problem of evolutionary tree analysis.

Both approaches used for inferring the tree use information on the variation of evolutionary rates of different words. This is essential but very rarely done, because it affects strongly the shape of the curve formed by the rate of cognate retention rates (C) versus separation times t, causing a serious underestimation of longer times when compared with the standard glottochronological approach (Swadesh, 1952), that assumes a proportionality of log C and t. They use a method that assumes a normal distribution of log C of the retention rate of individual words and estimate it directly from the data, while in our analysis we estimate it using another source of information: the number of the different roots used to express the same meaning of each word.There remain a few differences between the trees, and it is worth considering them in detail. Gray and Atkinson seem to agree with us that there could be two origins of Indo-European languages, the first in coincidence with the origin of agriculture as suggested by Renfrew (1987), to be located in the Middle East or Anatolia, and a later one in the Ukraine, as suggested by Gimbutas (1985). The oldest languages, Armenian, Albanian and Greek, are among the oldest in both trees, but there is some disagreement in the relevant dichotomies. These are, however, those that have the highest errors in both trees, as shown by the percentage of agreement among repetitions of the analysis.

The other discrepancy is the dichotomy of Celtic, which in our tree is the oldest of the European subfamilies, while in theirs the oldest is Balto-Slavic. Our bootstrap value is higher than in their tree, indicating our method has smaller error in this part of the tree. There is information from other disciplines that supports our tree for both discrepancies.

If history can support some separation dates, though very weakly, geography may again be of help. Albanian is weakly related to Indic Iranian, while in our tree it is nearest to the root, closest to Armenian and Greek, in agreement with geography. Given the long distance between Albania and south Asia, and the local tree uncertainty it may be better to make the first dichotomy of the tree as a branch leading to a trichotomy of Albania, Greece and Armenia, corresponding with what remains of the first spread of farmers from Anatolia, and another branch leading to all the rest, reflecting later farmers expansions starting from the Ukraine, that gave rise to an early split into the Indic-Iranian branch going east and south, and the European branch, with the splitting sequence in time Celtic/Italic-Germanic/ Balto-Slavic. Making the Celtic branch the eldest is in agreement with other information : 1) Celtic languages are believed to have been spoken in Austria, Switzerland and northern Italy by the La Tene culture at least in the early part of the third millennium BC ; 2) in

Julius Caesar's time Celtic languages were spoken in France and Great Britain, while Germanic languages were spoken east of the Rhine; the later spread northwards and westwards of Germanic languages and southwards and westwards of Italic languages confined Celtic languages to the most peripheral parts of the British Isles, with Brittany speaking Celtic because of a secondary migration from the British Isles at the time of the Anglo-Saxon invasion, in the V-VI century AD. A remarkable help from weavings: the La Tène culture used Scottish style tartans , which were found over 3000 years ago also in the clothes of the mommies of west China. It is not entirely clear but these people may have spoken Tocharian in later times.

From a methodological points of view it is clear the retention rates of the Indo-European core vocabulary of 200 meanings considered in the analysis not only are heterogeneous but also fit to a bimodal gamma distribution and this adds further uncertainty to the dates associated to the major branchings in the tree.

From a general point of view it is of some interest to explore how the linguistic classification correlates with genetic data. Poloni et al. (1997) showed, for the Y chromosome, an important level of population genetics structure among human populations, mainly due to genetic differences among distinct linguistic groups of populations. A multivariate analysis based on genetic distances between populations shows that human population structure inferred from the Y chromosome corresponds broadly to language families (r = .567, P < .001), in agreement with autosomal and mitochondrial data. Times of divergence of linguistic families, estimated from their internal level of genetic differentiation, are fairly concordant with current archaeological and linguistic hypotheses. Variability of the p49a,f/TaqI Y polymorphic marker is also significantly correlated with the geographic location of the populations (r = .613, P < .001), reflecting the fact that distinct linguistic groups generally also occupy distinct geographic areas. Comparison of Y-chromosome and mtDNA polymorphisms in a restricted set of populations shows a globally high level of congruence, but it also allows identification of unequal maternal and paternal contributions to the gene pool of several populations.

3. Towards a Global Perspective

More than 5,000 languages are spoken today in the world, and it does not take a linguist to recognize that some languages are more closely related than others due to history. The official origin of historical linguistics can be dated to 1786, when the English judge Sir William Jones advanced the idea that Sanskrit, a

classical language in India, Greek, Latin, and possibly Celtic and Gothic (the ancestor of Germanic languages) shared a common origin. These old languages were the first members of a family of languages that would become known as the `Indo-European' family (or `phylum'). As Indo-European is the earliest and best studied linguistic family, coevolution of genes and languages has been documented. Since the eighteenth century, however, many other linguistic families or superfamilies have been recognized. The most complete classification on a world basis was proposed by Ruhlen (1994) on the basis of Greenberg's published and unpublished writings: he lists 12 linguistic families (Khoisan, Niger-Kordofanian, Nilo-Saharian, Afro-Asiatic, Dravidian, Kartvelian, Euroasiatic, Dene-Caucasian, Austric, Indo-Pacific, Australian, Amerind).

The reconstruction of the relationships above the family level is hotly debated among historical linguists who have yet to agree on the existence of a single tree linking all the existing language families, that is on the possible differentiation of modern languages from a single ancestor language. Even unification at a lower level such as that of the (pre-Columbian) American languages proposed by Greenberg (1987), who grouped them into just three macro-families (Eskimo-Aleut, Na-Dene, and Amerindian), has been strongly opposed by the majority of American linguists. Interestingly, Greenberg's proposal seems to agree with the analysis of genetic markers in extant Native Americans (Cavalli-Sforza et al. 1994) and these three families seem to identify three major migrations suggested by archaeological data. Amerindian speakers appear to have come first (between 30,000 and 15,000 years ago according to genetic data), followed by Na-Dene speakers and finally Eskimo-Aleut (both in a period between 15,000 and 10,000 years ago). It must be said, however, that at a finer level of classification contemporary Amerindian speakers show high genetic variability, and this is not easy to reconcile with linguistic taxonomy.

Even without an agreed genealogy of the linguistic families covering all tongues spoken today, it is relevant to note the impressive one-to-one correspondence of the genetic phylogeny of the world populations with the classification into the 12 large linguistic families listed above (Cavalli-Sforza et al. 1988). This correspondence is expected because there are important similarities between the evolution of genes and languages. In either case: (a) a change which first appears in a single individual can subsequently spread throughout the entire population (for genes they are called mutations; they are rare, are passed from one generation to the next and can, over many generations, eventually replace the ancestral type; linguistic innovations are much more

frequent and can also pass between unrelated individuals); and (b) the dynamics of change is affected by the same demographic pressures, isolation, and migration. Two isolated populations differentiate both genetically and linguistically because isolation, which could result from geographic, ecological, or social barriers, reduces the likelihood both of marriages and cultural exchanges and, as a common result, reciprocally isolated populations will evolve independently and gradually become different. Both genes and languages will drift apart regularly over time, the former slowly, the latter much more quickly.

In principle, therefore, the linguistic tree and the genetic tree of human populations should agree since they reflect the same history of population splitting and subsequent independent evolution. The different rate of change, however, is a major source of divergence: one language can be replaced by another in a relatively short time. In Europe, for example, Hungarian is spoken in a land surrounded by Indo- European speakers but it belongs to the Finno-Ugric subdivision of Uralic. At the end of the ninth century AD, the nomadic Magyars left their land in Russia and invaded Hungary. The number of conquerors was probably less than 30 percent of the conquered population so that their genetic contribution was limited, but they imposed their language on the local Romance-speaking population. Today all Hungarians speak a Uralic language, but barely 10 percent of their genes can be attributed to the Uralic conquerors.

Generally it is intuitive that the total substitution of one language for another occurs more easily under the pressure of a strong political power of the newcomers, as witnessed in the Americas. The case of Basques, on the other hand, shows that separate languages spoken in nearby countries can remain relatively unaffected for thousands of years, even when their genes experience a partial substitution. It is remarkable that, despite the above sources of confusion, the correlation between genes and languages has been maintained through the centuries until today and is still statistically significant. The ties between biology and linguistics were already evident since the times of Darwin, who in chapter XIV of his The Origin of Species wrote:

> "If we possessed a perfect pedigree of the mankind, a genealogical arrangement of the races of man would afford the best classification of the various languages now spoken throughout the world; and if all extinct languages, and all intermediate and slowly changing dialects, were to be included, such an arrangement would be the only possible one. Yet it might be that some ancient language had altered very little and had given rise to few new languages, whilst others had altered much owing to the spreading, isolation, and state of civilization of the several co-descended races, and had

thus given rise to many new dialects and languages. The various degrees of difference between the languages of the same stock, would have to be expressed by groups subordinate to groups; but the proper or even the only possible arrangement would still be genealogical; and this would be strictly natural, as it would connect together all languages, extinct and recent, by the closest affinities, and would give the filiation and origin of each tongue."

The increasing resolving power of modern genetic data makes it possible to follow Darwin and to use the genetic phylogeny of our species to infer the earliest branches of a hypothetical linguistic tree. The most comprehensive genetic phylogeny reconstructed in Cavalli-Sforza et al. (1988) was used by Ruhlen (1994) to draw the tree of origin of human languages (some reference dates from genetic and archaeological evidence have been added). The oldest linguistic families must be African: Khoisan is probably the oldest and Afro-Asiatic the most recent, while Niger-Kordofanian and Nilo-Saharian believed by some linguists to descend from an ancestor tongue, and the Congo- Saharan, were probably spoken at an intermediate time. A more exhaustive discussion of this hypothetical tree can be found in Cavalli-Sforza (2000). As the genetic data improves with the inclusion of more representatives from those geographical areas of the world where the sampling is still scanty, the tree will be more complex but it is likely that its main features will remain unchanged.

In conclusion, our present genome keeps the record of its past evolution with an impressive richness of detail that is also reflected by our languages. Genes and languages contribute to the understanding of human history by highlighting human diversity; both are instrumental in giving some of the silent voices of our past a chance to be heard.

References

Ammerman, A.J., Cavalli-Sforza, L.L (1984). *Neolithic Transition and the Genetics of Populations in Europe*. Princeton University press, Princeton, NJ

Anthony, D.W. (1995). Horse, wagon & chariot: Indo-European languages and archaeology. Antiquity 69: 554-65

Barbujani, G., Sokal, R.R. (1990). Zones of sharp genetic change in Europe are also linguistic boundaries. *Proceedings of the National Academy of Sciences* 87: 1816-9

Bosch-Gimpera, A. (1943). El problema de los origines vascos. *Eusko-Jakintza* 3: 39

Capelli, C., Redhead, N., Abernethy, J.K., Gratrix, F., Wilson, J.F., Moen, T., Hervig, T., Richards, M., Stumpf, M.P., Underhill, P.A., Bradshaw, P., Shaha, A., Thomas, M.G., Bradman, N., Goldstein, D.B. (2003). A Y chromosome census of the British Isles. *Current Biology*, **13,** 979 - 984.

Cavalli-Sforza, L.L. (2000). *Genes, Peoples, and Languages*. North Point Press, New York

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A. (1994). *The History and Geography of Human Genes*. Princeton University Press, Princeton, NJ

Cavalli-Sforza, L.L., Piazza, A., Menozzi, P., Mountain, J. (1988). Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proceedings of the National Academy of Sciences* 85: 6002-6

Chalmers, J.N.M., Ikin, E.W., Mourant, A.E. (1948). Basque blood groups. *Nature* 162: 27

Dyen, I., J.B Kruskal & P.Black. (1992). An Indoeuropean classification: a lexicostatistical experiment, *Transactions of the American Philosophical Society* 82: Part 5. Philadelphia, American Philosophical Society.

Gimbutas, M. (1985) Primary and secondary homeland of the Indo-Europeans. *Journal of Indo-European Studies* 13: 185-202

Gray, R.D., Atkinson, Q.D. (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435-439.

Greenberg, J.H. (1987). *Language in the Americas*. Stanford University Press, Stanford, CA

Lehmann, W.P. (1993). *Theoretical Bases of Indo-European Linguistics*. Routledge, London

Mallory, J.P. (1989). *In Search of the Indo-Europeans: Language, archaeology and myth*. Thames and Hudson, London

Otte, M. Soffer, O., Gamble, C., (eds.) (1990). In *The World at 18,000 BP*, Unwin Hyman, London.

Piazza, A., Cappello, N., Olivetti, E., Rendine, S. (1988). A genetic history of Italy. *Annals of Human Genetics 5*2: 203-13

Piazza, A., Rendine, S., Minch, E., Menozzi, P., Mountain, J., Cavalli-Sforza, L.L. (1995). Genetics and the origin of the European languages. *Proceedings of the National Academy of Sciences* 92: 5836-40

Piazza, A., Minch, E., Cavalli-Sforza, L.L. (in preparation). The Indo-Europeans: Linguistic tree and genetic relationships. Manuscript

Poloni, E.S., Semino, O., Passarino, G., Santachiara-Benerecetti, A.S., Dupanloup, I., Langaney, A., Excoffier, L. (1997). Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. *Am J Hum Genet.* 61, 1015-35.

Renfrew, C. (1987). *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge University Press, New York

Richards, M., Macaulay, V., Hickey, E., Vega, E., Syke,s B., Guida, V., Rengo, C., Sellitto, D., Cruciani,. F, Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Calò, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., Hatina, J., Belledi, M., Di Rienzo, A., Oppenheim, A., Novelletto, A., Nùrby, S., Al-Zaheri, N., Santachiara-Benerecetti, S., Scozzari, R., Torroni, A., Bandelt, H.-J. (2000). Tracing European founder lineages in the near eastern mtDNA pool. *The American Journal of Human Genetics* 67, 1251-76.

Richards, M., Macaulay, V., Torroni, A., Bandelt, H.-G. (2002). In Search of Geographical Patterns in European Mitochondrial DNA. *Am. J. Hum. Genet.* 71,1168-1174.

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. *The American Journal of Human Genetics* 67, 1526-1543.

Ruhlen, M. (1994). *On the Origin of Languages: Studies in Linguistic Taxonomy.* Stanford University Press, Stanford, CA

Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhill, P.A. (2000). The genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* 290: 1155-1159

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society* 96:452-463.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistical dating. *International Journal of American Linguistics* 21:121-137.

Torroni, A., Bandelt, H.-J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savantaus, M.-L., Bonné-Tamir, B., Scozzari, R. (1998). mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152.

Trombetti, A. (1926). *Le origini della lingua Basca.* Bologna, Italy

Wilson, J.F., Weiss, D.A., Richards, M., Thomas, M.G., Bradman, N., Goldstein, D.B. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci USA* 98:5078‒5083.