

Geographic Patterns of mtDNA Diversity in Europe

Lucia Simoni,^{1,2} Francesc Calafell,² Davide Pettener,¹ Jaume Bertranpetit,² and Guido Barbujani³

¹Department of Evolutionary and Experimental Biology, University of Bologna, Bologna; ²Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona; and ³Department of Biology, University of Ferrara, Ferrara, Italy

Summary

Genetic diversity in Europe has been interpreted as a reflection of phenomena occurring during the Paleolithic (~45,000 years before the present [BP]), Mesolithic (~18,000 years BP), and Neolithic (~10,000 years BP) periods. A crucial role of the Neolithic demographic transition is supported by the analysis of most nuclear loci, but the interpretation of mtDNA evidence is controversial. More than 2,600 sequences of the first hypervariable mitochondrial control region were analyzed for geographic patterns in samples from Europe, the Near East, and the Caucasus. Two autocorrelation statistics were used, one based on allele-frequency differences between samples and the other based on both sequence and frequency differences between alleles. In the global analysis, limited geographic patterning was observed, which could largely be attributed to a marked difference between the Saami and all other populations. The distribution of the zones of highest mitochondrial variation (genetic boundaries) confirmed that the Saami are sharply differentiated from an otherwise rather homogeneous set of European samples. However, an area of significant clinal variation was identified around the Mediterranean Sea (and not in the north), even though the differences between northern and southern populations were insignificant. Both a Paleolithic expansion and the Neolithic demic diffusion of farmers could have determined a longitudinal cline of mtDNA diversity. However, additional phenomena must be considered in both models, to account both for the north-south differences and for the greater geographic scope of clinal patterns at nuclear loci. Conversely, two predicted consequences of models of Mesolithic reexpansion from glacial refugia were not observed in the present study.

Introduction

Broad clines encompassing much of Europe have been observed for many classes of genetic markers, including blood groups and allozymes (Menozzi et al. 1978; Sokal et al. 1989), histocompatibility alleles (Menozzi et al. 1978; Sokal and Menozzi 1982), and Y-chromosome and autosomic DNA variants (Semino et al. 1996; Chikhi et al. 1998a, 1998b; Seielstad et al. 1998; Casallotti et al. 1999). Menozzi et al. (1978) identified, in a multilocus gradient extending from the Levant into northern and western Europe, the main component of European genetic diversity, accounting for 26% of the overall allele-frequency variance (Cavalli-Sforza et al. 1994). That figure may have been somewhat inflated by the fact that principal component (PC) analysis requires interpolation of data, possibly resulting in artificial clinal patterns (Sokal et al. 1999). However, using spatial autocorrelation analysis, a different approach that does not require previous manipulation of data, Sokal et al. (1989) concluded that one-third of the European alleles are clinally distributed over the entire continent and that most loci are affected by clinal variation.

If the existence of clines spanning much of Europe is undisputed, little consensus has been reached on their origin. The archaeological record shows traces of two demographic transitions that could have affected genetic variation on a continental scale—namely, the first *Homo sapiens sapiens* colonization (starting ~45,000 years before the present [BP], in the early Upper Paleolithic) and the Neolithic spread of farming (starting ~10,000 years BP) (Cavalli-Sforza et al. 1994). In both cases, Near Eastern populations expanded into much of western and northern Europe (fig. 1). Between those expansions, in the last glacial period, peaking 18,000–20,000 years ago (Mellars 1994), populations may have withdrawn into a few (perhaps three) warmer areas, or *glacial refugia*, from which they reexpanded as the climate improved, during what is called the “late Upper Paleolithic,” or simply “Mesolithic,” period. Thus, Mesolithic gene flow may also have affected the pattern of genetic affinities among European populations (Richards et al. 1998; Torroni et al. 1998), but, because this gene flow was caused by dispersal from several centers, continentwide clines are not among its expected consequences. Conversely,

Received July 15, 1999; accepted for publication September 9, 1999; electronically published December 15, 1999.

Address for correspondence and reprints: Dr. Guido Barbujani, Dipartimento di Biologia, Università di Ferrara, via L. Borsari 46, I-44100 Ferrara, Italia. E-mail: bjg@dns.unife.it

© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6601-0028\$02.00

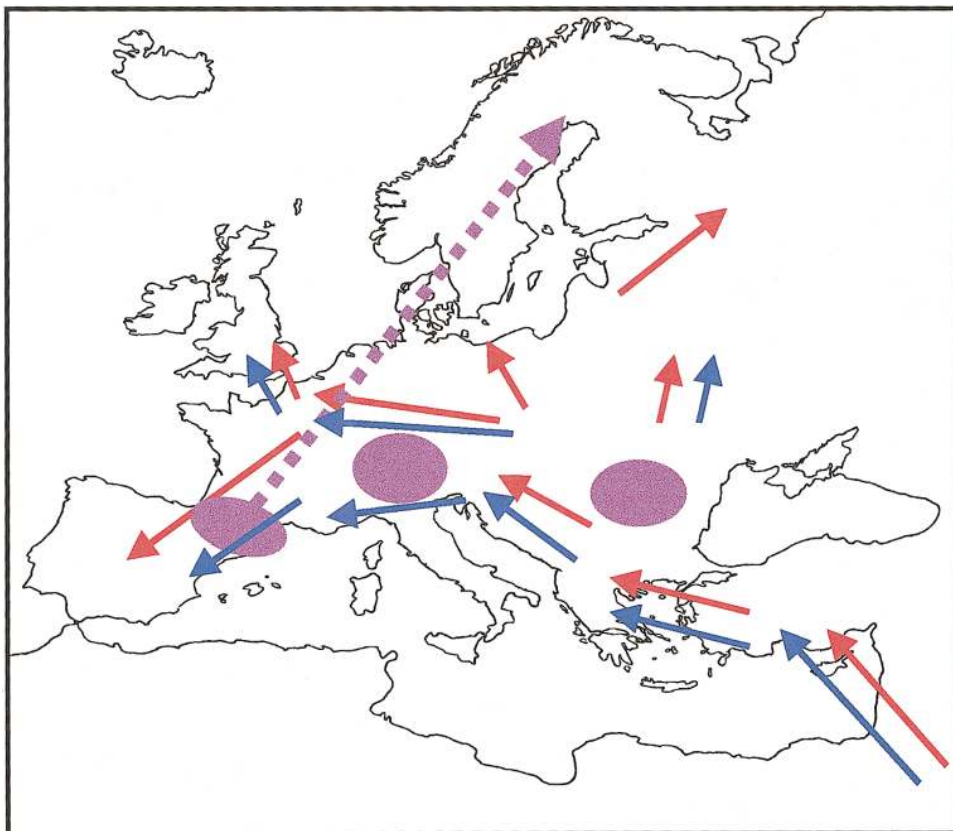


Figure 1 Scheme of the main dispersal processes supposed to have occurred during the Paleolithic first colonization of Europe (red arrows [from Richards et al. 1997, p. 253]) and during the Neolithic demic diffusion (blue arrows [from Renfrew 1987, p. 160]). The approximate location of glacial refugia is represented (violet ovals), as is a major Mesolithic expansion from Iberia (dashed violet arrow [proposed by Torroni et al. 1998]).

Paleolithic and Neolithic dispersals were directional processes, the potential effects of which include the establishment of southeast-northwest patterns at many loci.

Until recently, most agreed on the notion that the European clines were a trace of the demic diffusion of the first Neolithic farmers (Ammerman and Cavalli-Sforza 1984; Renfrew 1987). Communities that could produce (as opposed to collect) food tended to grow in size and to disperse wherever suitable land was available (Pennington 1996). If Near Eastern Neolithic farmers did not mix much with the hunters and gatherers whom they met during their dispersal, broad gradients are expected at many loci. Comparisons of patterns of linguistic and genetic diversity (Sokal 1988; Barbujani and Pilastro 1993), as well as simulation studies (Rendine et al. 1986; Barbujani et al. 1995), agreed with this view.

However, as data on mtDNA variation accumulated, a different picture took shape. Not only were European levels of nucleotide and gene diversity much lower than those in Asia and Africa (Jorde et al. 1995), but European populations appeared extremely similar to each

other (Pult et al. 1994; Torroni et al. 1994; Sajantila et al. 1996). At the mtDNA level, even well-known isolated groups, such as the Basques, did not differ much from their neighbors (Bertranpetit et al. 1996), and only a subset of the major haplogroups (we will use this term to define a group of sequences sharing deep nucleotide substitutions in the allele genealogy) appeared to show any degree of geographic clustering in mitochondrial networks (Richards et al. 1996). Although a few differentiated communities were detected, notably the Saami (Sajantila et al. 1995) and the Ladins (Stenico et al. 1996), the study of mtDNA variation led to the proposal that the Neolithic contribution to the European gene pool had been overestimated in studies of protein markers (Richards et al. 1996; Comas et al. 1997). Instead, the European population structure would reflect the first Paleolithic colonization of the continent and the repeated founder effects that accompanied it. A process of this kind has only been modeled in simulations of a Neolithic expansion with no population admixture (Barbujani et al. 1995), but it also seems a plausible model

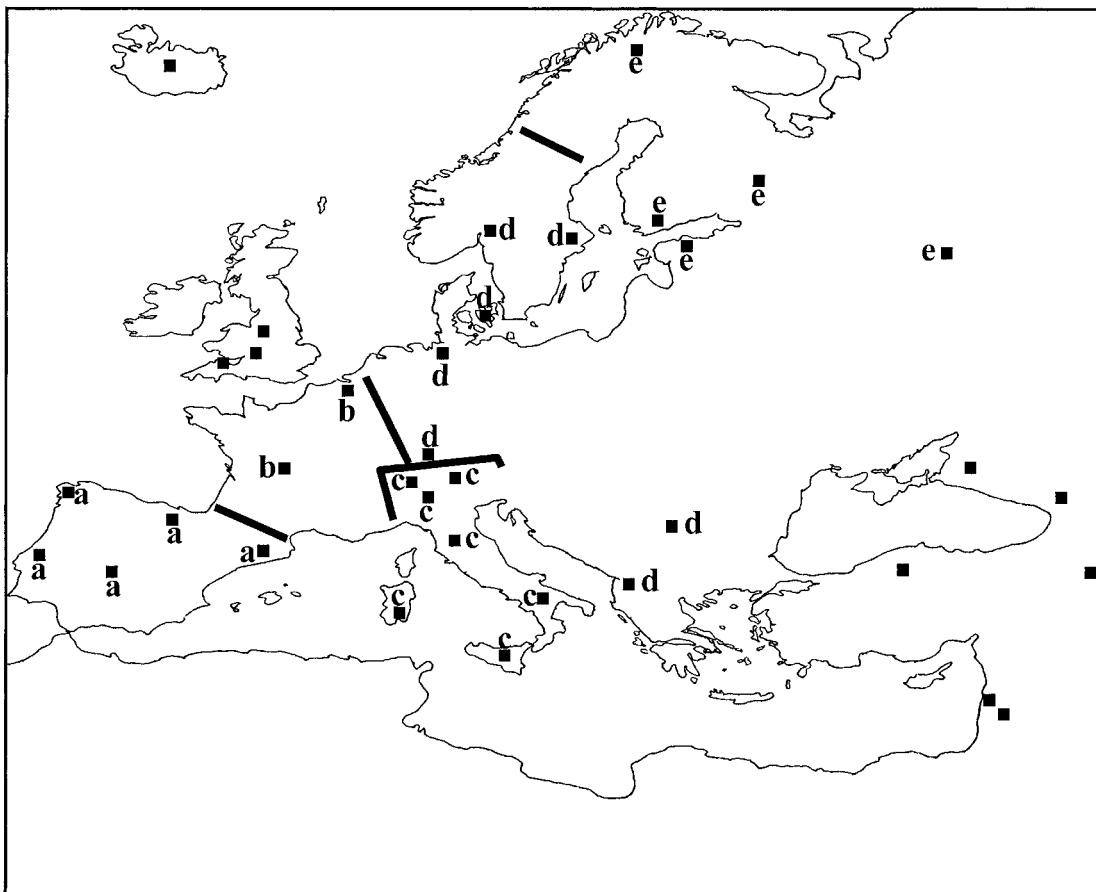


Figure 2 Geographic distribution of sampled populations in Europe. Unbroken lines represent the consensus suture zones recognized, in a study of 20 nonhuman species, by Taberlet et al. (1998, fig. 6). The groups of samples used in AMOVA are identified by different letters, *a–e*; samples not associated with a letter were not considered in that analysis.

of Paleolithic colonization (Richards et al. 1996). The controversy that followed (Cavalli-Sforza and Minch 1997; Richards et al. 1997; Barbujani et al. 1998; Richards and Sykes 1998) has not been settled yet.

One crucial question at this stage is to objectively define how mtDNA diversity is distributed in Europe. Indeed, quantitative methods for identification of geographic patterns have not been applied yet to the available European mitochondrial data. Once a pattern of population diversity has been described, a second question is whether mitochondrial and nuclear patterns are the same. Finally, a third question is whether these patterns are easiest to reconcile with the effects of demographic phenomena dating back to the Paleolithic, the Mesolithic, or the Neolithic period. We tried to address these questions and to test two expected consequences of Mesolithic dispersal. One is the suggestion, made by Torroni et al. (1998), of a major demographic expansion from southwestern Europe, a suggestion based on the sharing of some mitochondrial alleles (haplogroup V)

between Iberians and Saami. The second is the existence of sharp genetic change at the zones (suture zones) where populations expanding from different refugia should have met, which have been identified on the basis of a study by Taberlet et al. (1998). Other implications of the model of postglacial expansions will be tested in a study currently in progress. Finally, the existence of a Mediterranean cline proposed by Comas et al. (1997) was checked by use of the extended database and proper quantitative methods.

Material and Methods

Database

We collected 2,619 mtDNA sequences of the hypervariable region I (HVR-I), typed for positions 16024–16383 of the Cambridge reference sequence (Anderson et al. 1981). Samples are from all over Europe, the Near East, and the Caucasus (fig. 2). Europe was divided into

36 regions, corresponding generally to entire countries or, when more detailed information was available, either to more-restricted areas (such as Cornwall, Sardinia, and the eastern Alps) or to well-defined population groups (Adygh, Druzes, Saami, Basques, and Catalans). Geographic coordinates were inferred from the original papers and were associated with each population (table 1). Sample sizes vary from 15 (Catalans) to 249 (individuals from southern Germany) individual sequences, with an average of 72.8. Data on high-resolution RFLP typing of the entire mitochondrial chromosome (as by Macaulay et al. 1999) are available only for a small subset of the samples and therefore could not be considered for this analysis.

The number of different sequences observed is 852, and the number of polymorphic sites is 241. This implies high statistical noise caused by rare substitutions, which contain little phylogenetic information. Richards et al. (1998) identified phylogenetically informative polymorphic sites in HVR-I, which allow one to divide the European sequences into various clusters, or haplogroups. Therefore, we focused on 22 such variable positions, which are listed in table 2. Note that some of these sites have been defined as fast evolving, in a worldwide analysis of mtDNA variation (Meyer et al. 1999). However, on a European scale, that is not always the case. An example is site 16223, which is very stable in Europe. In this way, 203 distinct 22-nucleotide haplotypes were obtained from the initial 2,619 sequences, each haplotype being present in one or more individuals. Variation at other sites was disregarded for spatial autocorrelation analysis but not for the study of suture zones and genetic boundaries (see below).

Spatial Autocorrelation Analysis

Patterns of mitochondrial variation were summarized by two spatial autocorrelation methods. Spatial autocorrelation compares data (here, DNA sequences and haplogroup frequencies) within arbitrary space lags. Measures of overall genetic similarity are evaluated in each distance class, and inferences are based on the degree of genetic similarity at various geographic distances. A variable is autocorrelated, positively or negatively, if its value at a given point in space is associated with its values at other locations.

Sequence and frequency data were analyzed, respectively, by use of AIDA—an approach specifically designed for DNA analysis (Bertorelle and Barbujani 1995)—and SAAP—a classical spatial autocorrelation approach (Sokal and Oden 1978). SAAP was also used to describe the pattern of variation of Nei's (1987) gene diversity, D , which is equivalent to the expected heterozygosity for diploid data. The autocorrelation statistics calculated by AIDA and SAAP are called " II " and " I ,"

respectively. Both are defined between +1 and -1, and, for large sample sizes, their expected value is close to 0. I is independently calculated for each allele (or, in this case, for each haplogroup), and its values reflect the degree of frequency similarity between samples at a given distance. II is calculated by comparison of sequences, and so its values reflect both frequency and sequence similarity. Since neither I nor II is normally distributed, their significance was assessed by permutation tests. II can also be estimated at distance 0, yielding a measure of sequence similarity within populations; it can be regarded as the increase in the probability to sample the same nucleotide twice in the same population, with respect to random sampling over the entire continent. Geographic distances were calculated as air distances, which are known to correlate well with road distances in Europe (Crumpacker et al. 1976).

The set of spatial autocorrelation coefficients (II or Moran's I) evaluated at various distance classes, or the *correlogram*, can be associated with one or more likely generating processes (Sokal and Oden 1978). A spatially random distribution results in a series of insignificant autocorrelation coefficients at all distances. A decreasing set of coefficients, from positive significant values to negative significant values, describes a cline, whereas a correlogram decreasing from positive significant values through insignificant values at large distances is expected under isolation by distance—that is, when genetic diversity is the product of the interaction between genetic drift and short-range gene flow (Barbujani 1987). Finally, negative coefficients in the last distance classes reflect some kind of long-range differentiation (i.e., the pattern typical of a subdivided population in which geographically extreme samples are also the most differentiated but in which there is no overall gradient).

Haplogroup Definitions in SAAP

As shown in table 2, each haplogroup defined by Richards et al. (1998) is associated with a specific set of substitutions involving some of the 22 nucleotide sites considered in our analysis. On the basis of polymorphism at those sites, ~78% of the sequences of our database could be unambiguously assigned to one haplogroup. In the 29 cases in which a sequence contained mutations characteristic of two different haplogroups, a parsimony criterion was followed, to assign each case to the most probable haplogroup. For instance, a sequence bearing substitutions at sites 16069, 16126, and 16343 was allocated to haplogroup J, because the 16069 and 16126 substitutions are distinctive for the J motif, whereas 16343 is part of the U3 motif. An additional 17% of the sequences presented one to four substitutions, which, however, did not allow one to associate them with certainty to any haplogroup; therefore, they

Table 1**Populations Considered in Present Study**

Population (Size)	Latitude	Longitude	Reference(s)
Albania (42)	41°20' N	19°50' E	Belledi et al. (in press)
Austria (117)	47°16' N	11°24' E	Handt et al. (1994), Parson et al. (1998)
Belgium (33)	50°50' N	4°20' E	De Corte et al. (1996)
Great Britain:			
Cornwall (69)	50°18' N	5°03' W	Richards et al. (1996)
Mainland (100)	52°30' N	1°48' W	Richards et al. (1996)
Wales (92)	51°30' N	3°12' W	Piercy et al. (1993)
Bulgaria (30)	42°48' N	23°18' E	Calafell et al. (1996)
Caucasus-Adygei (50)	44°38' N	40°04' E	Macaulay et al. (1999)
Denmark (32)	55°45' N	12°30' E	Richards et al. (1996)
Estonia (28)	59°24' N	24°45' E	Sajantila et al. (1995)
Finland (79)	60°09' N	24°57' E	Sajantila et al. (1995), Richards et al. (1996)
France (111)	47°05' N	2°24' E	M. Le Roux (personal communication)
Georgia (45)	41°43' N	44°49' E	D. Comas (personal communication)
Germany:			
Northern (108)	53°36' N	10°00' E	Richards et al. (1996)
Southern (249)	48°03' N	9°47' E	Richards et al. (1996), Lutz et al. (1998)
Iceland (53)	64°06' N	22°00' W	Sajantila et al. (1995), Richards et al. (1996)
Israel-Druze (45)	38°04' N	35°37' E	Macaulay et al. (1999)
Italy:			
Alps (115)	46°03' N	11°03' E	Stenico et al. (1996)
Sardinia (73)	39°12' N	9°06' E	Di Rienzo and Wilson, (1991), O. Rickards (personal communication)
Sicily (63)	37°09' N	14°27' E	O. Rickards (personal communication), L. Nigro (personal communication)
Southern (37)	40°30' N	15°50' E	O. Rickards (personal communication)
Tuscany (49)	43°18' N	11°15' E	Francalacci et al. (1996)
Karelia (83)	61°54' N	34°06' E	Sajantila et al. (1995)
Kurds (29)	37°00' N	43°00' E	D. Comas (personal communication)
Near East (42)	32°00' N	36°00' E	Di Rienzo and Wilson (1991)
Norway (30)	59°55' N	10°45' E	Dupuy and Olaisen (1996)
Portugal (54)	38°39' N	9°09' W	Corte-Real et al. (1996)
Saami (240)	68°54' N	27°00' E	Sajantila et al. (1995), Dupuy and Olaisen (1996)
Spain:			
Basques (106)	43°24' N	2°00' W	Bertranpetit et al. (1996), Corte-Real et al. (1996)
Catalunya (15)	41°18' N	2°12' E	Corte-Real et al. (1996)
Central (74)	40°24' N	3°42' W	Corte-Real et al. (1996), Pinto et al. (1996)
Galicia (92)	42°53' N	8°33' W	Salas et al. (1998)
Sweden (32)	59°20' N	18°03' E	Sajantila et al. (1995)
Switzerland (72)	46°00' N	8°57' E	Pult et al. (1994)
Turkey (96)	40°00' N	32°48' E	Calafell et al. (1996), Comas et al. (1996), Richards et al. (1996)
Volga-Finnic (34)	56°38' N	47°52' E	Sajantila et al. (1995)

were included in a group called “Other.” Finally, 49 sequences could belong to haplogroups I, W, or X, and 70 sequences could belong to either J or T. These 119 sequences were considered only in the analysis of superhaplogroup frequencies (see below). In summary, for each population, we had frequencies of 11 European haplogroups—namely, H, I, J, K, T, U3, U4, U5, V, W, X, and Other. Haplogroup H contains all sequences, including the Cambridge reference sequence (Anderson et al. 1981), that show none of the 22 substitutions considered in this study. Since haplogroups H and Other together account for ~48% of all European sequences, the remaining haplogroups tend to be either rare or absent altogether in some populations. To reduce the statistical effect of random fluctuations around such very low values, we also analyzed by SAAP the frequencies

of combinations of these haplogroups, or superhaplogroups—namely, IWX, HV, KU, and JT (table 3)—which are considered to be monophyletic in Europe (Richards et al. 1998).

Testing for the Effects of Suture Zones

Taberlet et al. (1998) looked for concordant geographic patterns of DNA variation in 20 animal and plant species whose distribution has probably been affected by postglacial (i.e., Mesolithic) expansions. On the basis of paleontological and genetic evidence, four suture zones (fig. 2) were identified as the places where different waves of expanding animal and plant populations presumably came into contact. Those suture zones subdivide Europe into five regions. If, after the last

Table 2
Haplogroup and Superhaplogroup Definitions for 2,619 Individuals

Haplogroup (No. of Individuals)	Variable Position(s) in HVR-I (16024–16383) ^a
H (783)	...
I (72)	16223, 16129
J (180)	16126, 16069
K (145)	16224, 16311
T (197)	16126, 16294
U3 (38)	16343
U4 (80)	16356
U5 (298)	16270
V (176)	16298
W (43)	16223, 16292
X (39)	16223, 16278
Other (449)	Several
IWX (49)	16223
JT (70)	16126

^a In addition to those listed, positions 16145, 16163, 16172, 16186, 16189, 16193, 16222, 16231, and 16261 also were considered in spatial autocorrelation analysis, for a total of 22 segregating sites.

glacial maximum, humans expanded along the same routes, each of these five regions should show some degree of genetic homogeneity and should be rather differentiated from the other regions. This prediction was tested by a form of analysis of variance, AMOVA (Excoffier et al. 1992), that estimates the fraction of the total genetic variance that can be attributed to three sources: (1) individual differences among members of the same sample, (2) differences among samples of the same region, and (3) differences among the five regions defined above. By means of a randomization procedure, AMOVA also evaluated the statistical significance of the variance components thus estimated.

Genetic Boundaries

The zones of sharpest mtDNA change in Europe, or genetic boundaries, were identified by use of a method based on genetic distances (see Bosch et al. 1997; Stenico et al. 1998). Localities were connected according to adjacency criteria, thus defining a so-called Delaunay triangulation (fig. 3). Genetic distances between populations connected by single edges of the network were calculated. Two genetic-distance measures were used, one (d_{AB}) giving equal weight to any substitutions (Nei 1987) and another weighting transversions 15 times as much as transitions (Tamura and Nei 1993). From the edges of the network associated with the highest genetic distances, an arbitrary number (six, in this study) of lines of maximum genetic differentiation, or genetic boundaries, was traced. The significance of each boundary thus identified was eventually tested by AMOVA, comparing the samples on either side of that boundary.

Results

Spatial Autocorrelation Analysis: AIDA

AIDA was independently run six times, initially considering all samples available and then removing some of them or separately analyzing specific regions. In the global analysis, *II* is positive and highly significant at distance 0 (table 4, group A); that is, genetic similarity is higher within than between samples. At increasing distances, the level of autocorrelation tends to decrease, in a rather irregular way. Although *II* coefficients are positive and significant for distances <1,500 km and are negative and significant for distances >2,000 km, the overall trend is not one of monotonic decrease. The greatest divergence is observed at 2,500–3,800 km, whereas sequences sampled at greater distances show lower numbers of substitutions. Therefore, this pattern is not strictly clinal, although it shows significant long-distance differentiation. Many comparisons in the extreme-distance classes involve either Georgians or Kurds, populations that, in the neighbor-joining tree (Saitou and Nei 1987) estimated from Nei’s (1987) d_{AB} distances (not shown), cluster with northern Europeans. As a consequence, the highest level of sequence differentiation is found in comparisons between southeastern Europeans and Saami, mostly falling in the distance interval of 2,500–3,000 km.

Well-known mitochondrial outliers in Europe are Saami and Ladins; here the latter were pooled with other groups from the Alps. Removal of the Alps sample from the database did not change the AIDA profile in essence (data not shown). Conversely, a new pattern emerged when Saami were excluded (table 4, group B). *II* at distance 0 dropped drastically ($II_0 = .0104$); the autocorrelation at short distances became insignificant; differences >3,000 km, although still significant, were reduced to one-fifth or less; and sequences in the extreme-distance class no longer differed significantly. This suggests that the previously observed pattern in Europe is largely the result of two facts: (1) that Saami differ from all other Europeans (resulting in long-distance negative autocorrelation, because of their extreme geographic position, and in short-distance positive autocorrelation, because all other samples are comparatively similar to each other) and (2) that Saami are very homogeneous internally (resulting in increased autocorrelation at distance 0).

Because correlation statistics are heavily affected by extreme values, another possible distortion of the global autocorrelation pattern in Europe may come from the geographic position of Icelanders. The Icelandic population did not evolve *in loco* for long but was established comparatively recently, through immigration from Scandinavia and, possibly, admixture with groups of Irish

Table 3

Haplogroup Frequencies, as Inferred from Sequence Data in HVR-I, and Genetic Diversity

SOURCE	FREQUENCY OF HAPLOTYPE														GENETIC DIVERSITY (D)		
	H	I	J	K	T	U3	U4	U5	V	W	X	Other	IWX	HV		KU	JT
Albania	.524	.071	.009	.000	.000	.000	.000	.143	.000	.000	.000	.119	.071	.524	.143	.143	.906
Austria	.325	.034	.000	.103	.085	.009	.043	.068	.009	.009	.009	.188	.051	.333	.222	.205	.959
Belgium	.406	.000	.030	.125	.031	.000	.063	.031	.094	.000	.000	.156	.000	.500	.219	.125	.991
Great Britain:																	
Cornwall	.348	.058	.000	.029	.087	.000	.058	.043	.014	.000	.000	.145	.058	.362	.130	.304	.965
Mainland	.350	.030	.011	.100	.070	.000	.040	.080	.030	.000	.030	.140	.070	.380	.220	.190	.976
Wales	.478	.033	.000	.076	.043	.000	.000	.043	.033	.000	.011	.130	.043	.511	.120	.196	.931
Bulgaria	.233	.000	.067	.133	.100	.100	.067	.033	.000	.000	.067	.167	.067	.233	.333	.200	.977
Caucasus-Adygey	.220	.060	.000	.020	.140	.140	.040	.080	.000	.000	.000	.260	.060	.220	.280	.180	.951
Denmark	.344	.000	.000	.031	.063	.000	.000	.063	.031	.031	.000	.250	.031	.375	.094	.250	.934
Estonia	.214	.000	.000	.000	.107	.000	.071	.179	.000	.071	.000	.250	.107	.214	.250	.179	.989
Finland	.278	.063	.044	.051	.051	.000	.025	.139	.089	.076	.000	.127	.152	.367	.215	.139	.970
France	.450	.018	.009	.045	.099	.000	.000	.054	.027	.036	.000	.180	.072	.477	.099	.171	.964
Georgia	.178	.022	.000	.111	.244	.044	.111	.067	.000	.022	.044	.089	.111	.178	.333	.289	.964
Germany:																	
Northern	.250	.009	.178	.093	.083	.019	.009	.074	.065	.009	.009	.269	.046	.315	.194	.176	.973
Southern	.309	.028	.000	.052	.088	.024	.052	.092	.048	.016	.004	.177	.056	.357	.221	.189	.977
Iceland	.226	.019	.014	.038	.075	.057	.019	.113	.019	.000	.000	.245	.038	.245	.226	.245	.979
Israel-Druze	.244	.044	.000	.156	.044	.000	.000	.000	.000	.000	.178	.178	.267	.244	.156	.156	.952
Italy:																	
Alps	.184	.026	.032	.088	.219	.018	.035	.061	.061	.000	.000	.193	.026	.246	.202	.333	.993
Sardinia	.384	.027	.067	.055	.110	.000	.000	.082	.027	.014	.014	.219	.068	.411	.137	.164	.956
Sicily	.397	.016	.027	.048	.016	.032	.016	.000	.079	.048	.032	.190	.143	.476	.095	.095	.962
Southern	.378	.027	.041	.027	.108	.000	.000	.081	.027	.054	.027	.108	.189	.405	.108	.189	.969
Tuscany	.224	.041	.000	.061	.061	.000	.041	.061	.000	.020	.041	.224	.143	.224	.163	.245	.969
Karelia	.313	.024	.000	.024	.072	.000	.084	.181	.060	.036	.000	.120	.096	.373	.289	.120	.964
Kurdistan	.276	.069	.095	.172	.034	.069	.000	.000	.000	.034	.000	.276	.172	.276	.241	.034	.958
Near East	.095	.071	.000	.000	.119	.024	.000	.000	.000	.000	.095	.190	.238	.095	.024	.452	.995
Norway	.467	.067	.019	.133	.000	.000	.000	.100	.033	.033	.000	.133	.133	.500	.233	.000	.954
Portugal	.444	.000	.000	.074	.111	.019	.056	.000	.037	.000	.019	.130	.056	.481	.148	.185	.934
Saami	.033	.025	.019	.000	.004	.000	.000	.529	.346	.000	.000	.058	.029	.379	.529	.004	.799
Spain:																	
Basques	.509	.000	.000	.047	.047	.000	.000	.104	.066	.000	.019	.179	.019	.575	.151	.075	.936
Catalunya	.267	.000	.000	.000	.000	.000	.133	.000	.267	.133	.067	.067	.200	.533	.133	.067	.952
Central	.522	.011	.004	.033	.022	.000	.011	.043	.033	.022	.011	.185	.135	.297	.176	.189	.987
Galicia	.243	.041	.000	.054	.095	.014	.027	.081	.054	.041	.027	.203	.054	.554	.087	.120	.930
Sweden	.250	.000	.027	.031	.125	.000	.094	.063	.063	.000	.000	.281	.000	.313	.188	.219	.988
Switzerland	.278	.014	.000	.069	.028	.000	.056	.083	.056	.014	.000	.292	.042	.333	.208	.125	.965
Turkey	.221	.032	.000	.032	.074	.053	.032	.021	.032	.042	.032	.168	.200	.253	.137	.242	.988
Volga-Finnic	.176	.029	.032	.029	.118	.000	.147	.118	.029	.000	.000	.176	.029	.206	.294	.294	.982

origin. Data for group C in table 4 show that, when both Saami and Icelanders were excluded, there was virtually no positive short-distance autocorrelation left. However, *I* regained significance in the extreme-distance class. This is the pattern that we previously had defined as long-distance differentiation. Its simplest interpretation is that the extreme geographic position of Icelanders does not correspond to their mtDNA characteristics, which seem comparable to those of most other Europeans. When they are removed from the analysis, that distortion is removed, and a weak—but not insignificant—differentiation between groups separated by >3,000 km becomes evident.

Cavalli-Sforza and Minch (1997), using the data set of Richards et al. (1996), proposed that European mi-

tochondrial variability is basically a result of the differences between Near Easterners and all other populations. When we excluded from the analysis the two Near Eastern samples, we obtained the same autocorrelation profile observed for the whole data set (table 4, group D). Thus, the pattern of European mtDNA diversity does not seem to depend crucially on differences between Europeans proper and people from the Levant. On the contrary, some degree of genetic continuity between these populations seems to be apparent.

In synthesis, the overall patterns of mtDNA diversity appeared to be poorly significant in Europe. It made sense, however, to test whether they could reach significance within certain regions. In a study of eight European samples, six of them from the Mediterranean area,

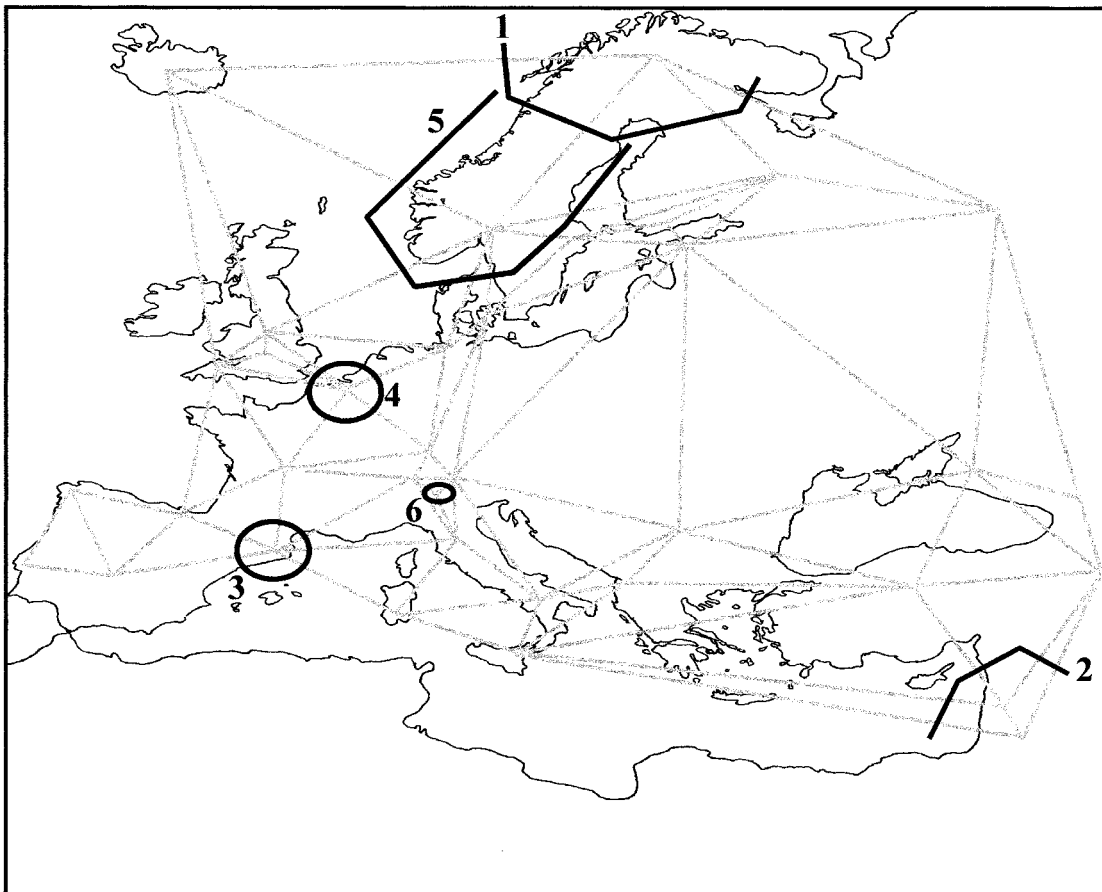


Figure 3 Zones of maximum genetic change in Europe, inferred from mtDNA variation (*thick lines*); numbers refer to their ranking. Localities are connected by a Delaunay network (*thin lines*).

Comas et al. (1997) described an east-west gradient of mean pairwise sequence differences and suggested that immigration from the Near East could account for it. Also, Malyarchuk (1998) described a gradient of mitochondrial RFLPs in southern but not in northern Europe. To quantitatively test these observations, we independently ran AIDA on groups from the Mediterranean region and on groups from central northern Europe; 14 and 19 samples were taken into account, respectively, whereas the Kurds and two samples from the Caucasus were excluded from both analyses.

In the region around the Mediterranean Sea, AIDA identified a quasiclinal pattern (table 4, group E). Not only was autocorrelation at distance 0 rather high ($II_0 = .0222$), but genetic resemblance declined smoothly, from positive significant at <1,000 km to negative significant at >2,500 km. A different pattern was observed for northern Europe. When the Saami were included, the correlogram was significant; it decreased at <3,000 km, and it increased at greater distances (data not given). But when Saami and Icelanders were excluded, autocorrelation within samples became very high

($II_0 = .0767$), short-range similarity lost significance, and negative autocorrelation was observed at large distances, even though its significance was low (table 4, group F).

Therefore, separate analyses of groups from northern and southern Europe point to two distinct patterns. Northern Europe shows only moderate differences among the most distant populations, whereas a cline is apparent around the Mediterranean Sea. We tested whether these two regions are genetically differentiated, but AMOVA showed that they are not ($F_{st} = .0027$; $P > .05$). Therefore, what we observed is a result not of the presence of different alleles or of different frequencies of these alleles in northern and southern Europe but of the different ways in which these alleles are distributed in space.

Spatial Autocorrelation Analysis: SAAP

Autocorrelation analysis of haplogroup frequencies shows a remarkable lack of geographic structure. Five *I* correlograms—corresponding to haplogroups H, U3, U4, U5, and W—are significant at the $P = .05$ level. But,

Table 4**Spatial Autocorrelation: AIDA**

Population Group and Upper Limit	<i>I</i>
A. 36 populations:	
0	.0696***
200	.0052***
500	.0043***
1,000	.0081***
1,500	.0042***
2,000	.0005
2,500	-.0048***
3,000	-.0177***
3,800	-.0148***
5,300	-.0095***
B. 35 populations, without Saami:	
0	.0104***
200	.0006
500	.0006
1,000	.0023***
1,500	.0005**
2,000	-.0008
2,500	-.0017
3,000	-.0035***
3,800	-.0064***
5,300	-.0035
C. 34 populations, without Saami and Icelanders:	
0	.0107***
200	.0005
500	.0006
1,000	.0024***
1,500	.0004
2,000	-.0008
2,500	-.0014
3,000	-.0034***
3,800	-.0080***
4,550	-.0054***

(continued)

after Bonferroni correction for multiple tests (Sokal and Rohlf 1995, pp. 702–703), only H and U3 appear to depart significantly from randomness, in a fashion that cannot be defined as clinal, although there is significant differentiation at >3,000 km (see Sokal et al. 1989) (table 5). Note, however, that haplogroup U3 is absent in 22 populations and has an average frequency of 1.7%; therefore, aside from the effect of sampling errors, the significant spatial structure observed is essentially a result of the presence (mostly in southern and eastern Europe, in agreement with the AIDA results for group E in table 4) or absence of that haplogroup in the various localities. Exclusion of Saami and Icelanders did not change the results of SAAP analysis (data not given). Autocorrelation was also insignificant at all distances in the analysis of Nei's *D*.

A somewhat clearer, nonclinal structure was revealed by analysis of superhaplogroup frequencies. Significant patterns, characterized by decreasing negative values at large distances, were observed for superhaplogroups HV,

IWX, and KU; these are the patterns described, by Sokal et al. (1989), as long-distance differentiation. Both for the single haplogroups and for the superhaplogroups, autocorrelation coefficients were positive and significant at ~1,000 km but were insignificant at shorter distances. Under the joint effects of drift and short-range gene flow (i.e., under isolation by distance), populations geographically close to one another should be genetically closer than distant populations (Barbujani 1987), within a geographic range that depends on the mean and SD of migrational distances—in Europe, several hundred kilometers (Wijsman and Cavalli-Sforza 1984). The insignificant autocorrelation consistently observed at short distances for all superhaplogroups suggests the rather puzzling conclusion that short-range gene flow had no role in establishing the observed levels of diversity between populations. In this study, all comparisons between populations separated by <500 km had to be lumped into the first distance class. It may be that the distribution of samples was not sufficient to precisely

Table 4 (Continued)

Population Group and Upper Limit	<i>II</i>
D. 34 populations, without Near Easterners and Druze:	
0	.0693 ^{***}
200	.0053 ^{***}
500	.0030 ^{***}
1,000	.0070 ^{***}
1,500	.0038 ^{***}
2,000	.0004
2,500	-.0051 ^{***}
3,000	-.0195 ^{***}
3,800	-.0154 ^{***}
5,165	-.0057 ^{***}
E. 14 populations from the Mediterranean region ^a :	
0	.0222 ^{***}
500	.0078 ^{***}
1,000	.0044 ^{***}
1,500	-.0001
2,000	.0017 ^{***}
2,500	-.0040
3,000	-.0279 ^{***}
3,857	-.0201 ^{***}
F. 17 populations from central northern Europe, without Saami and Icelanders ^b :	
0	.0767 ^{***}
500	-.0012
1,000	-.0004
1,500	.0006 [*]
2,000	-.0021
2,500	-.0046 ^{**}
3,000	-.0041
3,480	-.0061

^a Samples considered were Albanians, Basques, Bulgarians, Catalans, Galicians, Italians (Sardinia, Sicily, southern Italy, and Tuscany), Near Easterners, Portuguese, Spaniards (mainland), Turks, and Druze.

^b Samples considered were Austrians, Germans (southern and northern), Belgians, Britons (Cornwall), Danes, Estonians, Finns, French, Italians (Alps), Karelians, Norwegians, Swedes, Swiss, Volga-Finnics, and Welsh.

^{*} $P < .05$.

^{**} $P < .01$.

^{***} $P < .005$.

detect short-distance patterns. Once again, however, most other DNA markers studied so far have shown significant positive autocorrelation at distances well beyond 500 km (Chikhi et al. 1998a). Therefore, patterns of mitochondrial allele frequencies in Europe really appear to differ from those observed at nuclear loci. The limited number of samples available made it impossible to analyze northern and southern Europe separately by SAAP.

Spatial Autocorrelation Analysis: Testing for the Effects of a Mesolithic Expansion

In order to test whether there is evidence of a post-glacial expansion from Iberia into northeastern Europe, the sequences of haplogroup V were independently analyzed along a transect corresponding to the dashed arrow in figure 1. The AIDA correlogram calculated from

the HVR-I sequences of the 115 northwestern Europeans listed in table 4 of the report by Torroni et al. (1998) showed a random pattern, with the usual positive correlation of sequences within samples (fig. 4a). No coefficient was significant at large distances, in contrast with the expected consequences of any directional population expansion (Sokal 1979; Bertorelle and Barbujani 1995). For the sake of completeness, we repeated the analysis on *all* sequences (haplogroup V as well as all other haplogroups) from that area of Europe. In the SAAP analysis, the correlogram did not significantly depart from its random expectations ($P = .092$; data not given). AIDA identified a clinal pattern (fig. 4b), which, however, disappeared when Saami were removed from analysis (fig. 4c). Therefore, (i) haplogroup V is not clinally distributed across western and northern Europe, and (ii) an apparent geographic structure in a northwestern

Table 5

Spatial Autocorrelation: SAAP

	AUTOCORRELATION RESULTS WHEN UPPER LIMIT FOR DISTANCE CLASS (NO. OF PAIRS OF POPULATIONS COMPARED) IS								CORRELOGRAM OVERALL PROBABILITY ^a
	500 (31)	1,000 (92)	1,500 (98)	2,000 (96)	2,500 (106)	3,000 (97)	3,800 (79)	5,310 (31)	
	<i>I</i>								
H	.09	.17*	.09	.07	.12*	-.10	-.51**	-.47**	<.005
W	.01	-.03	-.11	-.19*	-.02	.22**	-.11	.07	NS
X	.11	.14*	.05	.10	-.09	-.16*	-.19*	-.25	NS
I	-.16	.13*	.02	.03	-.02	.00	-.23*	-.28	NS
V	.05	.03	-.02	-.01	.04	-.19*	-.08	.00	NS
J	-.06	.01	-.07	-.02	-.17*	.04	.16*	-.18	NS
T	-.21	-.07	.09	-.02	.00	.02	-.16	-.07	NS
U3	.08	.30**	.18**	-.05	-.22**	-.19*	-.25**	.06	<.005
U4	-.07	-.03	-.08	.21**	-.09	-.12	-.06	.08	NS
U5	.04	.19**	.11*	.02	-.04	-.16*	-.26**	-.24*	NS
K	-.04	-.03	.00	-.06	-.13	-.04	.12	.01	NS
Other	.10	-.11	-.02	-.20*	.10	-.11	.11	.00	NS
HV	.17	.21**	.23**	.02	.13*	-.17	-.59**	-.55**	<.005
JT	-.03	-.11	.08	-.06	-.07	.05	-.07	-.04	NS
IWX	.15	.24**	.01	.03	-.05	-.02	-.30**	-.56**	.048
KU	.08	.10	.09	.11*	.09	-.15	-.40**	-.42**	<.005
	<i>D^b</i>								
	.02	-.03	-.08	.00	-.06	-.07	.10	-.10	NS

^a After Bonferroni correction. NS = not significant.

^b $(1 - \sum p^2) / [n(n-1)]$, where p is the frequency of each HVR-I allele.

* $P < .05$.

** $P < .01$.

transect is, once again, only a result of the difference between Saami and all other samples; once the former have been removed, what is left is an insignificant correlogram that does not point to any directional gene-flow process along the transect (as discussed by Sokal 1979, pp. 167-196).

Genetic Variances

We tested by AMOVA whether the mitochondrial structure of European populations resembles the structure determined by postglacial expansions in other species. Significant sequence heterogeneity is apparent among populations within regions (2.66% of the overall genetic variance; $P < .005$) but not among the five regions separated by the suture zones identified in 20 animal and plant species (2.71% of the genetic variance; not significant). In addition, that significance disappears when the Saami samples are excluded from the analysis, confirming that the differences between the Saami and all other Europeans is the main feature of European mitochondrial diversity. These results are based on the entirety of HVR-I haplotypes; they did not change when we took into account only the 22-nucleotide haplotypes used for AIDA analysis (data not shown).

Genetic Boundaries

The genetic boundaries inferred from mtDNA-based distances between populations are shown in figure 3. Regardless of whether the same weight or different weights are given to transitions and transversions, no large-scale subdivision of the European mitochondrial gene pool is apparent, and the same six groups are recognized as genetically differentiated. Saami show the sharpest difference with respect to their neighbors, followed by Near Easterners (including Druze), Catalans, Belgians, Norwegians, and the populations of the eastern Italian Alps (Ladin, German, and Italian speakers). No boundaries were found separating wide regions; the seventh boundary (not shown) separated Turkey from the rest of Europe. Small sample sizes may have determined, at least in part, this result for Catalans ($n = 15$), Norwegians ($n = 30$), and Belgians ($n = 33$). Some sections of boundaries 1 and 3 overlap with some sections of the suture zones identified by Taberlet et al. (1998), but others (especially genetic boundaries 1, 3, 4, and 6) subdivide areas that, under a model of Mesolithic expansions, should have been occupied by populations coming from the same glacial refugia and that are expected to be genetically homogeneous.

The significance of the subdivision thus inferred was tested by estimating the F_{st} between all pairs of popu-

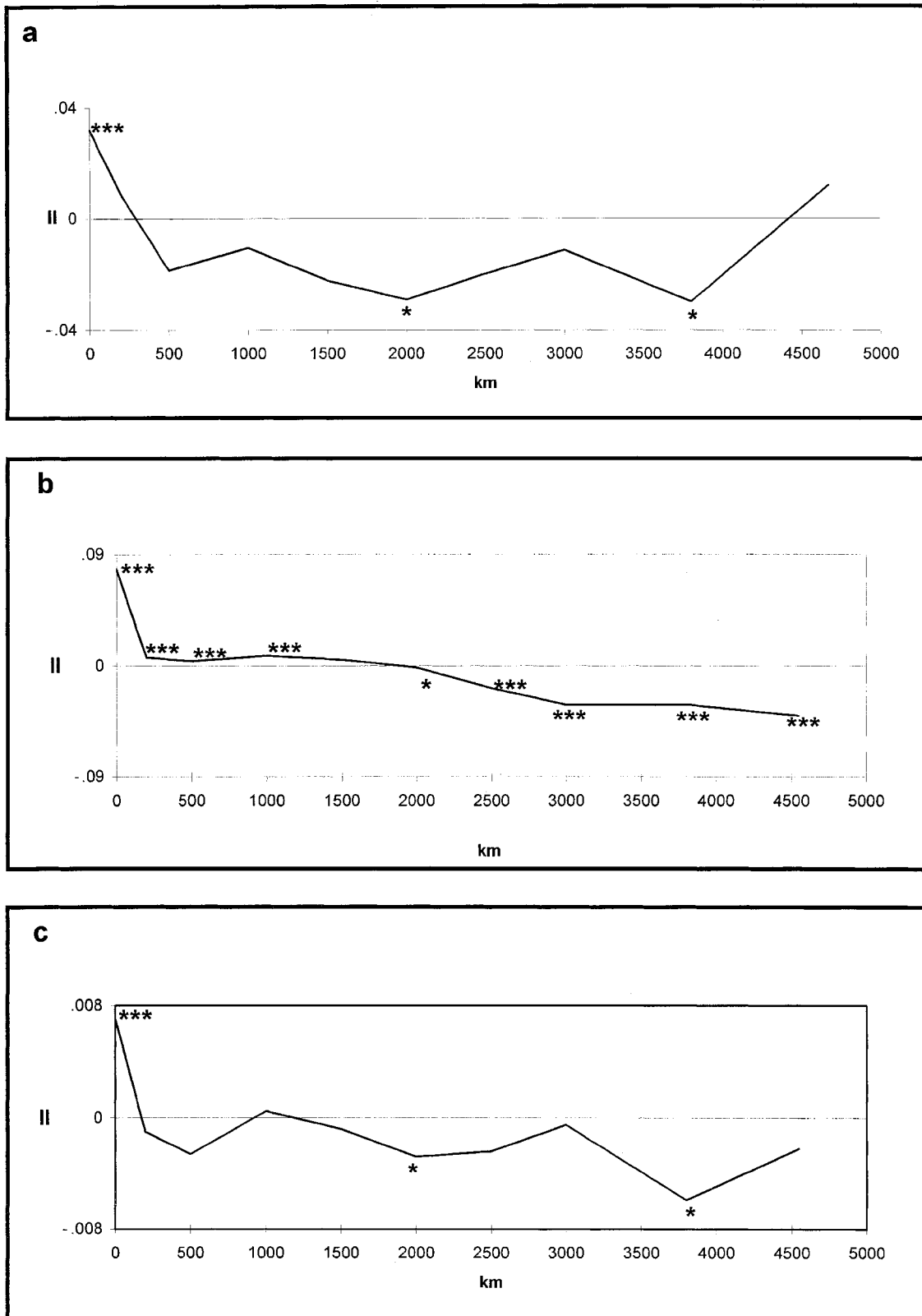


Figure 4 Spatial correlograms calculated along a northwestern-European transect. *a*, Haplogroup V sequences based on table 4 of Torroni et al. (1998) (individuals from North Africa, Sardinia, and Turkey were excluded). *b*, All haplogroups. *c*, All haplogroups with Saami excluded. The X-axis represents geographic distance between samples; the Y-axis represents II ; a single asterisk (*) denotes $P < .05$; triple asterisks (***) denote $P < .005$.

lations separated by a boundary. In all six cases, the F_{st} estimates were significantly >0 for at least two-thirds of the comparisons, with maximums of five out of five and six out of six for boundaries 2 and 4, respectively. When, by Fisher's method, the probabilities of the F_{st} values (Sokal and Rohlf 1995, pp. 794–797) were combined, all six identified boundaries showed highly significant results (although that significance is only nominal, because the comparisons are not independent).

Discussion

Geographic Patterns

Mitochondrial sequences and the frequencies of mitochondrial haplogroups are not clearly patterned over Europe. On a global scale, some degree of short-distance similarity is apparent, as is negative autocorrelation for samples separated by $>2,000$ km. However, if Saami are excluded from the analysis, the overall pattern does not depart significantly from random expectations, and even the genetic resemblance between samples from populations geographically close to one another (which could be regarded as an effect of isolation by distance) disappears. This is confirmed by the geographic position of the most significant genetic boundary, which separates the Saami from all other European groups.

At the regional scale, mitochondrial variation is poorly structured north of an imaginary line corresponding to the latitude of the Pyrenees. Only some degree of east-west divergence is evident, for samples separated by $>2,500$ km. This finding is compatible with an input of Asian genes in northeastern Europe, as has already been proposed (Sajantila et al. 1995). But when the analysis is restricted to southern Europe, a gradient becomes apparent. Note that the analysis of molecular variance failed to identify any significant differences between northern and southern Europe; allele frequencies are roughly the same in the two regions. What is different is their pattern, which is clinal only around the Mediterranean Sea.

Many mtDNA haplogroups are rare, and therefore their geographic pattern depends essentially on their presence or absence in the samples, which is affected by large stochastic variation. But rare alleles also occur at the microsatellite and HLA loci, and yet continentwide clines at those loci were identified by the same statistical methods used in this study (Sokal and Menozzi 1982; Chikhi et al. 1998a, 1998b; Casalotti et al. 1999). In synthesis, (1) many nuclear loci show gradients encompassing much of Europe; (2) the main mitochondrial characteristic of the European population seems to be a marked difference between the Saami and all other groups; and (3) a significant geographic structure is evident for mtDNA around the Mediterranean sea. Any

model trying to explain the origin of the European gene pool must account for all these aspects of genetic variation.

Effects of the Early Upper-Paleolithic Colonization (45,000–30,000 BP)

That the initial colonization of Europe may have generated the longitudinal clines observed for most nuclear markers has been suggested by Richards et al. (1997). One problem with this interpretation is that, according to the map of radiocarbon dates published by Richards et al. (1997), the southern and northern parts of Europe were colonized simultaneously during the Upper Paleolithic; it is unclear why the same process should have left such different genetic traces in the two regions. Another problem is what extent of genealogical continuity can be assumed to exist between the first Paleolithic settlers and the communities currently living in the same regions. Unlikely though it may seem, however, the view that the current genetic structure of the European population was established through repeated founder effects in the first colonization of Europe—and that essentially nothing important happened afterwards—is compatible with the patterns identified in this and previous studies.

Effects of a Late Upper-Paleolithic (Mesolithic) Recolonization (~20,000–15,000 BP)

The extinction of most populations living north of the ice line at the last glacial maximum could reasonably explain the existence of two different geographic patterns of mtDNA diversity in two climatic zones of Europe. The cline around the Mediterranean sea would be the only remaining trace of the initial, early Upper-Paleolithic colonization, whereas similar marks would have been erased in the areas that populations abandoned at the last glacial peak. The lack of continentwide structure for mtDNA could be attributed, perhaps, to the long times that are required for gene flow to determine an isolation-by-distance pattern (see Slatkin 1993). But why, instead, do nuclear markers show significant structuring over the entire continent? To account for that, a higher male than female mobility in the phase of recolonization should be assumed, contrary to what has been suggested by recent studies (Seielstad et al. 1998). In addition, in this study, we found no evidence of northward gradients radiating from Iberia (discussed in the section on Saami), nor did we observe any correspondence between the genetic boundaries and their expected location under the hypothesis of Mesolithic expansions (Taberlet et al. 1998). This implies either that human populations reexpanded after the Ice Age in a different manner than did populations of most other species studied or that the human population structure was not

deeply affected by processes occurring during Mesolithic times.

Effects of the Neolithic Demic Diffusion (10,000–5,000 BP)

A simple demographic expansion from the Levant is easy to reconcile with the gradients observed at many nuclear loci but is not easy to link with the fact that mitochondrial variation is clinal only in southern Europe. One possibility, supported by archaeological evidence, is that farming spread faster along the Mediterranean coastline (Renfrew 1987; Whittle 1994). So far, this point has not been incorporated into models of Neolithic demic diffusion, although a greater gene flow along the coasts resulted in better agreement between simulated and observed allele frequencies (Barbujani et al. 1995). If that were so, then the gradients predicted by Ammerman and Cavalli-Sforza's (1984) model may have been established initially or only in southern Europe. With the shift from a hunting-gathering to a food-producing economy, (1) populations increased in size, and (2) individual mobility was reduced (Renfrew 1987; Pennington 1996). As a consequence, at the end of a Neolithic expansion in the Mediterranean area, southern Europe may have been inhabited by large, stable, and geographically structured populations. On the contrary, in the north, where survival still depended on food collection, populations may have been smaller, mobile, and geographically heterogeneous, because of the strong impact that genetic drift had on them. Under those conditions, a high female mobility (Seielstad et al. 1998) would have had minor consequences in the south, but even limited exchange of women could have radically affected the maternally transmitted fraction of the genome of the small groups dwelling in the north.

More Questions

Discrepancies between the patterns identified at nuclear and mitochondrial loci may be due, in principle, to a number of factors, affecting both the mutation process (Aris-Brosou and Excoffier 1996; Parsons et al. 1997; Jazin et al. 1998) and the selection process. Because we did not estimate mutation rates on the basis of the data of the present study, we prefer not to speculate much on that issue. Rather, selection is a traditional explanation for differences among allelic distributions across loci (Cavalli-Sforza 1966; Lewontin and Krakauer 1973). Wise et al. (1998) suggested that past selective pressures have made human mtDNA unsuitable for evolutionary inferences. That is a rather drastic conclusion, but, indeed, the excess frequency of one mitochondrial type (haplogroup H), with respect to random expectations, may reflect either purifying selection

against some other types or a recent selective sweep, as originally proposed by Excoffier (1990).

Saami and Their Possible Mesolithic Origins

The role of Saami as genetic outliers in Europe is not a new finding (Cavalli-Sforza et al. 1994; Sajantila et al. 1995), and it is here confirmed by the distribution of the genetic boundaries and by the analysis of molecular variance. But this study does not confirm that the mitochondrial characteristics of Saami may reflect expansions from southwestern Europe. The hypothesis of long-range mitochondrial gene flow from Iberia was based on the presence of related haplogroup-V sequences, among Saami, Catalans, and Basques (Torrioni et al. 1998). The authors of that report then interpreted as consistent with that hypothesis one synthetic map obtained from PC analysis of European allele frequencies (Cavalli-Sforza et al. 1994). The latitudinal cline of the second PC across Europe, with the highest values in the Saami and the lowest in northern Iberia, was regarded as parallel to the pattern of haplogroup V, both patterns reflecting what was termed a "major late Paleolithic expansion" (Torrioni et al. 1998).

Two problems seem to persist in this interpretation. First, the second-PC scores have extreme *opposite* values in the Saami and in Iberia, whereas the frequencies of haplogroup V are highest in *both* Saami and Iberians; we do not see how two patterns could be more different. Second, the second PC is clinally distributed, with intermediate localities showing intermediate values, whereas this study shows that haplogroup V (whether investigated over all of Europe or only in the transect between Iberia and northeastern Europe) is not (fig. 4). New mtDNA data may modify this picture; the analysis by Torrioni et al. (1998) was based on a Catalan sample of size 15, in which the frequency of haplogroup V might be within the range .04–.49 (0.27 ± 1.96 standard errors). However, at present, the only possible conclusion is that haplogroup-V mtDNA data and nuclear data disagree, in that only the latter show a gradient—but that they also agree, in not suggesting expansions from Iberia into northern Europe. Data on ancient mtDNA in Basques also challenge that expansion hypothesis (Izaguirre and de la Rua 1999).

Final Remarks

In synthesis, it appears that mitochondrial- and nuclear-DNA variation in Europe have both a point of difference and something in common (question 2 of the Introduction). The main common feature is a significant (if limited, at the mitochondrial level) geographic structuring. However, contrary to the other markers studied so far by spatial autocorrelation, both at the protein and at the DNA level, mitochondrial alleles are distributed

in clines only in the southern part of the continent (question 1 of the Introduction). Globally considered, the single most significant feature of genetic variation in Europe is a continentwide cline, affecting many loci along a southeast-northwest axis. Clines of that breadth generally reflect directional demographic expansions, which archaeological evidence places either during the early upper Paleolithic or during the Neolithic, but not in the Mesolithic (question 3 of the Introduction). On the contrary, a north-south difference seems better explained by a model of postglacial reexpansions, during the Mesolithic, but when we tested for the expected consequences of that model we found no support for it. Therefore, the effects of Mesolithic processes, which, on the basis of mtDNA studies, have been proposed as major determinants of the European population structure, are not apparent in this analysis of mitochondrial diversity, nor were they identified in previous analyses of nuclear diversity (Chikhi et al. 1998a, 1998b; Casalotti et al. 1999).

A major role of Mesolithic demographic phenomena also seems to be rather difficult to reconcile with other aspects of human European diversity. Approximate though they must be, estimates of population divergence that are inferred from genetic distances at microsatellite loci suggest that local European gene pools separated <10,000 years ago (Chikhi et al. 1998a; Casalotti et al. 1999). Migration between regions that are geographically close to one another has probably reduced these genetic distances; it may be that estimated separation times within the Neolithic may reflect Mesolithic population splits, followed by local gene flow. Accordingly, the traces of the initial Paleolithic colonization of Europe would have been erased during the last glaciation, when Europeans may have concentrated within three refugia, south of the Pyrenees, the Alps, and the Balkans. This hypothesis leaves the Neolithic demic diffusion as the only possible cause of the many continentwide genetic gradients. But, be that as it may, clines radiating from a putative center of Mesolithic reexpansion, Iberia, have not been observed in the present study.

More modeling, larger sets of data for different genome regions, and computer simulations seem to be necessary if we are to solve, once and for all, the controversy about the origins of the European gene pool. One additional problem may lie in the fact that European history is comparatively well known. In many cases, it is possible to associate at least one historical phenomenon with any kind of observed genetic pattern. But whether that association is more than a coincidence is generally difficult to say. In the immediate future, significant steps forward in our understanding of human diversity will be possible only if we shall be able to define, a priori, alternative demographic models and to precisely predict their expected consequences, looking for consensus

among the interpretations suggested by the analysis of different loci. After all, the history of human populations has been a single one, and our best reconstruction is the one that jointly accounts for the patterns shown not by single genes but by most of the genome.

Acknowledgments

We are grateful to Giorgio Bertorelle, Eduardo Tarazona Santos, and Lounès Chikhi for many suggestions and for critical reading of the manuscript; the former also developed for this study an upgraded version of the AIDA software. We also thank Michele Belledi, David Comas, Ronny Decorte, Marie Le Roux, Olga Rickards, Michele Stenico, and Bryan Sykes for giving us access to unpublished data. This study was supported by grants from the Italian Ministry of Universities (COFIN 97 and Concerted Actions Italy-Spain), from the Italian National Research Council (contracts 96.01182.PF36 and 97.00683.PF36), and from the Dirección General de Investigación Científico Técnica (DGICT, Spain; grant PB95-0267-C02-01). F.C. is the recipient of a postdoctoral return contract from DGICT.

References

- Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, NJ
- Anderson, S, Bankier T, Barrell BG, De Bruijn MHL, Coulson AR, Drouin J, Eperon IC, et al (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465
- Aris-Brosou S, Excoffier L (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol Biol Evol* 13:494-504
- Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. *Genetics* 117:777-782
- Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* 62:488-492
- Barbujani G, Pilastro A (1993) Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. *Proc Natl Acad Sci USA* 90:4670-4673
- Barbujani G, Sokal RR, Oden NL (1995) Indo-European origins: a computer simulation test of five hypotheses. *Am J Phys Anthropol* 96:109-132
- Belledi M, Poloni ES, Casalotti R, Conterio F, Mikerezi I, Tagliavini J, Excoffier L. Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. *Eur J Hum Genet* (in press)
- Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. *Genetics* 140:811-819
- Bertranpetit J, Sala J, Calafell F, Underhill PA, Moral P, Comas D (1995) Human mitochondrial DNA variation and the origin of Basques. *Ann Hum Genet* 59:63-81
- Bosch E, Calafell F, Perez-Lezaun A, Comas D, Mateu E, Bertranpetit J (1997) Population history of North Africa: evidence from classical genetic markers. *Hum Biol* 69:295-311

- Calafell F, Underhill P, Tolun A, Angelicheva D, Kalaydjieva L (1996) From Asia to Europe: mitochondrial DNA sequence variability in Bulgarians and Turks. *Ann Hum Genet* 60:35–49
- Casalotti R, Simoni L, Belledi M, Barbujani G (1999) Y-chromosome polymorphism and the origins of the European gene pool. *Proc R Soc Lond B Biol Sci* 266:1959–1965
- Cavalli-Sforza LL (1966) Population structure and human evolution. *Proc R Soc Lond B Biol Sci* 164:362–379
- Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
- Cavalli-Sforza LL, Minch E (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61:247–254
- Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998a) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. *Proc Natl Acad Sci USA* 95:9053–9058
- Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, Barbujani G (1998b) Clinal variation in the nuclear DNA of Europeans. *Hum Biol* 70:643–657
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bertranpetit J (1996) Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationship to the European populations. *Mol Biol Evol* 13:1067–1077
- Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. *Hum Genet* 99:443–449
- Corte-Real HB, Macaulay VA, Richards MB, Hariti G, Issad MS, Cambon-Thomsen A, Papiha S, et al (1996) Genetic diversity in the Iberian Peninsula determined from mitochondrial sequence analysis. *Ann Hum Genet* 60:331–350
- Crumpacker DW, Zei G, Moroni A, Cavalli-Sforza LL (1976) Air distance versus road distance as a geographical measure for studies on human population structure. *Geogr Anal* 8: 215–223
- Decorte R, Jehaes E, Xiao FX, Cassiman JJ (1996) Genetic analysis of single hair shafts by automated sequence analysis of the mitochondrial d-loop region. *Adv Forensics Haemogenet* 6:17–19
- Di Rienzo A, Wilson AC (1991) Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* 88:597–601
- Dupuy BM, Olaisen B (1996) mtDNA sequences in the Norwegian Saami and main populations. In: Carracedo A, Brinkmann B, Bär W (eds) *Advances in forensic haemogenetics* 6. Springer-Verlag, Berlin, pp 23–25
- Excoffier L (1990) Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J Mol Evol* 30:125–139
- Excoffier L, Smouse PE, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491
- Francalacci P, Bertranpetit J, Calafell F, Underhill PA (1996) Sequence diversity of the control region of mitochondrial DNA in Tuscany and its implications for the peopling of Europe. *Am J Phys Anthropol* 100:443–460
- Handt O, Richards M, Trommsdorff M, Kilger C, Simanainen J, Georgiev O, Bauer K, et al (1994) Molecular genetic analyses of the Tyrolean Ice Man. *Science* 264:1775–1778
- Izagirre N, de la Rúa C (1999) A mtDNA analysis in ancient Basque population: implications for haplogroup V as a marker for a major Paleolithic expansion from southwestern Europe. *Am J Hum Genet* 65:199–207
- Jazin E, Soodyall H, Jalonen P, Lindholm E, Stoneking M, Gyllenstein U (1998) Mitochondrial mutation rate revisited: hot spots and polymorphism. *Nat Genet* 18:109–110
- Jorde LB, Bamshad MJ, Watkins WS, Zenger R, Fraley AE, Krakowiak PA, Carpenter KD, et al (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *Am J Hum Genet* 57:523–538
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* 74:175–195
- Lutz S, Weisser HJ, Heizmann JP (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111: 67–77
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, et al (1999) The emerging tree of west Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 64:232–249
- Malyarchuk BA (1998) Mitochondrial DNA markers and genetic demographic processes in Neolithic Europe. *Russian J Genet* 7:842–845
- Mellars P (1994) The upper Paleolithic revolution. In: Cunliffe B (ed) *The Oxford illustrated prehistory of Europe*. Oxford University Press, Oxford, pp 42–78
- Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201: 786–792
- Meyer S, Weiss G, von Haeseler A (1999) Pattern of nucleotide substitution and rate heterogeneity in the hypervariable regions I and II of human mtDNA. *Genetics* 152:1103–1110
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York
- Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 111:124–132
- Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, et al (1997) A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15:363–368
- Pennington RL (1996) Causes of early population growth. *Am J Phys Anthropol* 99:259–274
- Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106:85–90
- Pinto F, Gonzalez AM, Hernandez M, Larruga JM, Cabrera VM (1996) Genetic relationship between the Canary Islanders and their African and Spanish ancestors inferred from mitochondrial DNA sequences. *Ann Hum Genet* 60: 321–330
- Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Swit-

- zeland reveal striking homogeneity of European populations. *Biol Chem Hoppe Seyler* 375:837-840
- Rendine S, Piazza A, Cavalli-Sforza LL (1986) Simulation and separation by principal components of multiple demic expansions in Europe. *Am Nat* 128:681-706
- Renfrew C (1987) *Archaeology and language: the puzzle of Indo-European origins*. Jonathan Cape, London
- Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, et al (1996) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185-203
- Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. *Ann Hum Genet* 62:241-260
- Richards MB, Macaulay VA, Sykes BC, Pettitt P, Hedges R, Forster P, Bandelt HJ (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61:247-251
- Richards MB, Sykes BC (1998) Reply to Barbujani et al. *Am J Hum Genet* 62:491-492
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425
- Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus ML, Aula P, et al (1995) Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 5:42-52
- Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci USA* 93:12035-12039
- Salas A, Comas D, Lareu MV, Bertranpetit J, Carracedo A (1998) MtDNA analysis of the Galician population: a genetic edge of European variation. *Eur J Hum Genet* 6:365-375
- Seielstad M, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. *Nat Genet* 20:278-280
- Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am J Hum Genet* 59:964-968
- Slatkin M (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264-279
- Sokal RR (1979) Ecological parameters inferred from spatial correlograms. In: Patil GN, Rosenzweig ML (eds) *Contemporary quantitative ecology and related ecometrics*. International Co-operative, Fairland, MD
- (1988) Genetic, geographic and linguistic distances in Europe. *Proc Natl Acad Sci USA* 85:1722-1726
- Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. *Am J Phys Anthropol* 80:267-294
- Sokal RR, Menozzi P (1982) Spatial autocorrelation of HLA frequencies in Europe supports demic diffusion of early farmers. *Am Nat* 119:1-17
- Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. *Biol J Linn Soc* 10:199-249
- Sokal RR, Oden NL, Thomson BA (1999) The trouble with synthetic maps. *Hum Biol* 71:1-13
- Sokal RR, Rohlf FJ (1995) *Biometry*. WH Freeman, San Francisco
- Stenico M, Nigro L, Barbujani G (1998) Mitochondrial lineages in Ladin-speaking communities of the eastern Alps. *Proc R Soc Lond B Biol Sci* 265:555-561
- Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C, Barbujani G (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. *Am J Hum Genet* 59:1363-1375
- Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Mol Ecol* 7:453-464
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. *Mol Biol Evol* 9:678-387
- Torroni A, Bandelt HJ, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, et al (1998) MtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137-1152
- Torroni A, Lott MT, Cabell MF, Chen YS, Lavergne L, Wallace DC (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. *Am J Hum Genet* 55:760-776
- Whittle A (1994) The first farmers. In: Cunliffe B (ed) *The Oxford illustrated prehistory of Europe*. Oxford University Press, Oxford, pp 136-166
- Wijsman EM, Cavalli-Sforza LL (1984) Migration and genetic population structure with special reference to humans. *Annu Rev Ecol Syst* 15:279-301
- Wise CA, Sraml M, Eastaer S (1998) Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. *Genetics* 148:409-421