Am. J. Hum. Genet. 66:262–278, 2000

Geographic Patterns of mtDNA Diversity in Europe

Lucia Simoni,[1,2] Francesc Calafell,[2] Davide Pettener,[1] Jaume Bertranpetit,[2] and Guido Barbujani[3]

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

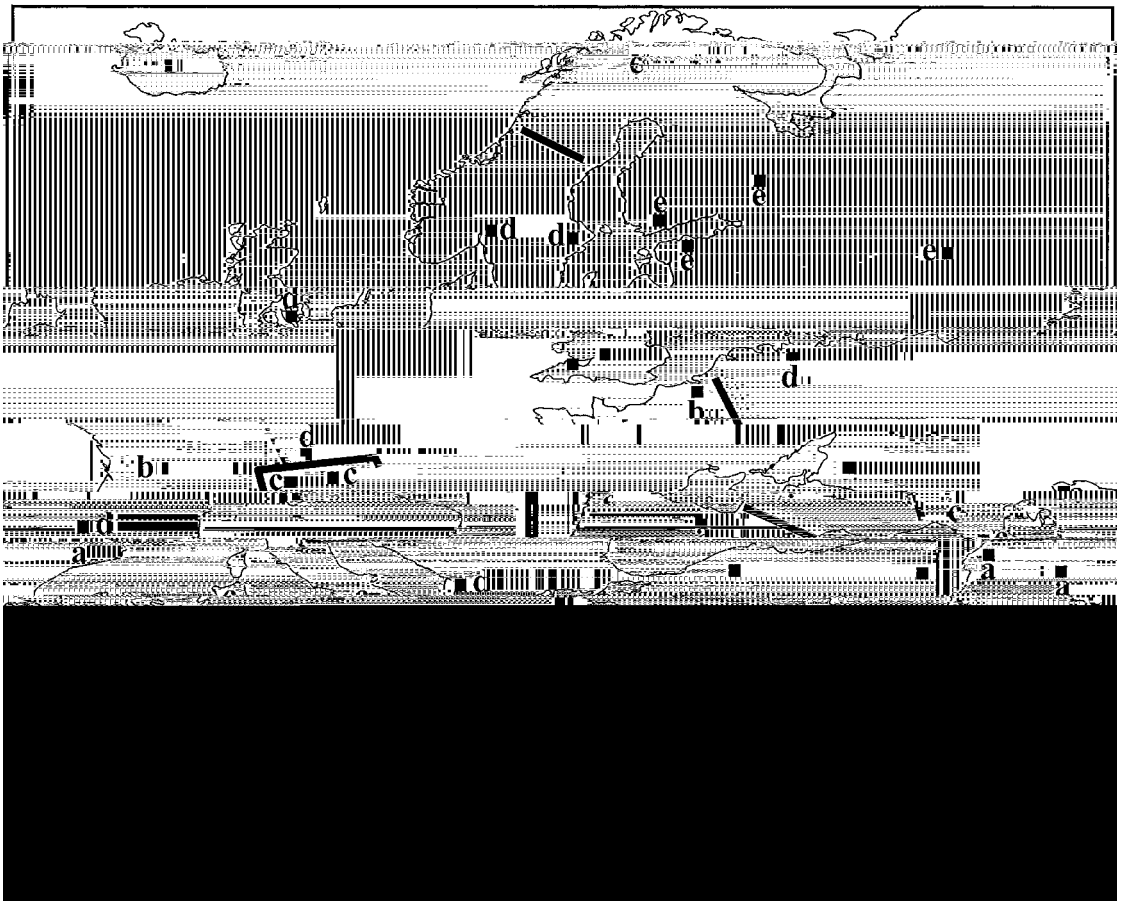

Figure 2 Geographic distribution of sampled populations in Europe. Unbroken lines represent the consensus suture zones recognized, in a study of 20 nonhuman species, by Taberlet et al. (1998, fig. 6). The groups of samples used in AMOVA are identified by different letters, *a–e;* samples not associated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

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, ~

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 rae7,mom(div9uld)-400(bea7 7.9etiated)-345(from)-345(the)-340(other)-340(regions.) sources: (1) individual 9ul 7.9eces among members of bea7the same sample, (2) 9ul 7.9eces among samples of the bea7same reg

Genetic Boundaries

The zon9s of sharpest mtDNA change in Europe, or genetic boundaries, were ideetified by use of a method bea7based on genetic distaeces (see Bosch et al. 1997; Stenico bea7et al. 199 jac9ecy criteria, thus 9efining a so-called Delaunay triangulation (fig. 3). Genetic distaeces between populationsby singleedges the network were calculated. Two genetic-distaece measureswere used, bea7on9 (D_{AB}) giving equal weight to any substitutions (Nei 1987) and another weighting transversions 15 times as bea7much as transitions (Tamura and Nei 1993). From the bea7edges of t aries, was traced. The significance of each boundary thus ideetified was eventually tested by AMOVA, comparing bea7the samples on either side of that boundary.

Results

Spatial Autocorrelation Analysis: AIDA

AIDA was independeetly run six times, initially considering all samples available and then removing some bea7of them or distaece 0 (table 4, group A); that is, genetic similarity is higher withinthanbetween samples. At increasing bea7distaeces, the positive and significant for distaeces <1,500 km and are negative and significant for distaeces >2,000 km, the overall trend is not strictly clinal, although it treme-distaece classes involve eit

Well-known mitochondrial Saami and Ladins; here the latt taece 0 dropped drastically (

relation at short 9ustaeces became insignificant; 9ul 7.-eeces >3,000 km, although still significant, were reduced to on9-fifth or taece class no longer 9ul 7.9d significantly. This suggests bea7that tocorrelation, because of their extreme geographic position, and in short-distaece positive autocorrelation, because all samples are comparatively similar to each other

Table 3

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

Source	Frequency of Haplotype																Genetic Diversity
	H	I	J	K	T	U3	U4	U5	V	W	X	Other	IWX	HV	KU	JT	(*D*)
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	.06tBulgar7Dm80			i8.A4548(.130)-10927 311.574 67w4(.1054(.100)000)-5360(.260)-4(.4(.060)-1Esto.498 r4340)									
4170.103 .067	.103	.067															

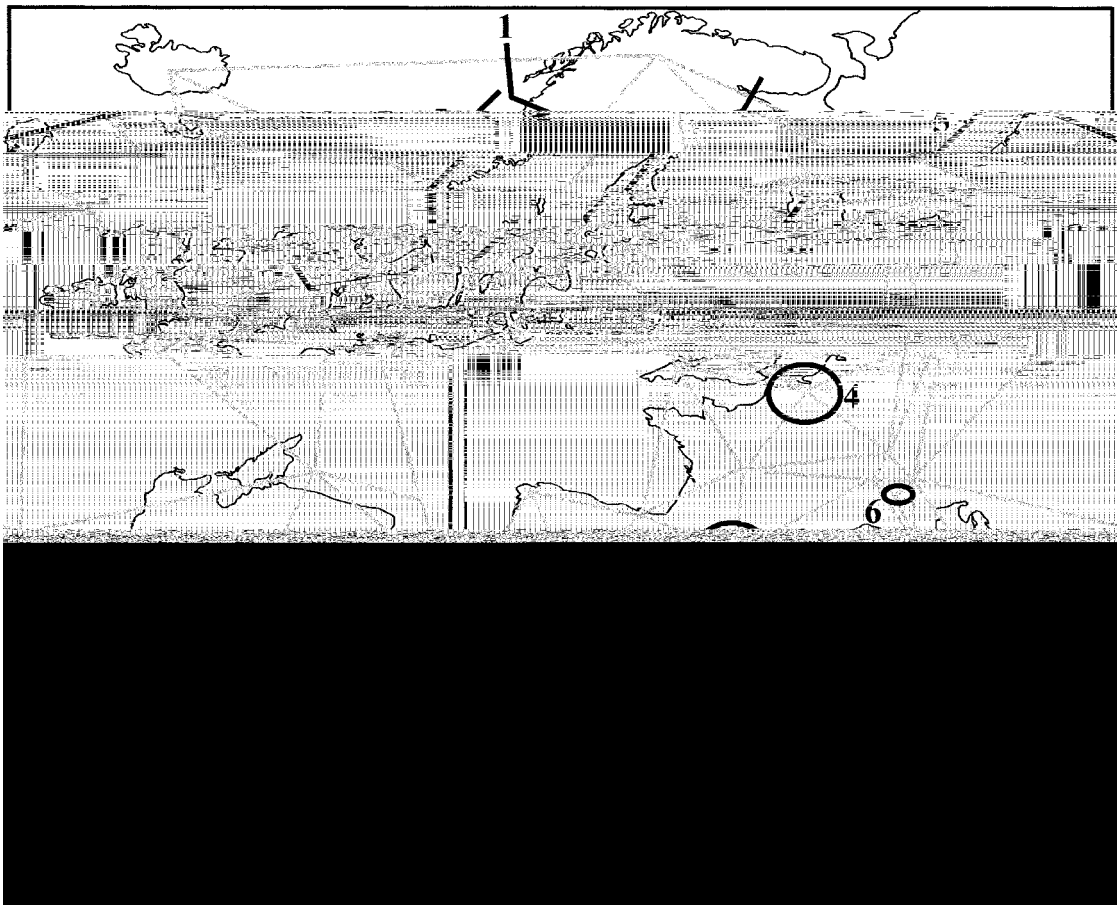

Figure 3 Zones of maximum genetic change in Europe, inferred from mtDNA variation (*thick lines*); numbers refer to their ranking.

Table 4

Spatial Autocorrelation: AIDA

Population Group and Upper Limit	*II*
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***

Table 4 (Continued)

Population Group and Upper Limit	*II*
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. 1998*a*). 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 postglacial 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. 4*a*). 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 285(e11 1 Tfaences)-30-380(a-ucli70(pop9ern,)-360figur.)-

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				
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
					D[b]				
	.02	−.03	−.08	.00	−.06	−.07	.10	−.10	NS

[a] After Bonferroni correction. NS = not significant.

[b] $(1-\Sigma p_i^2)[n/(n-1)]$, 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-

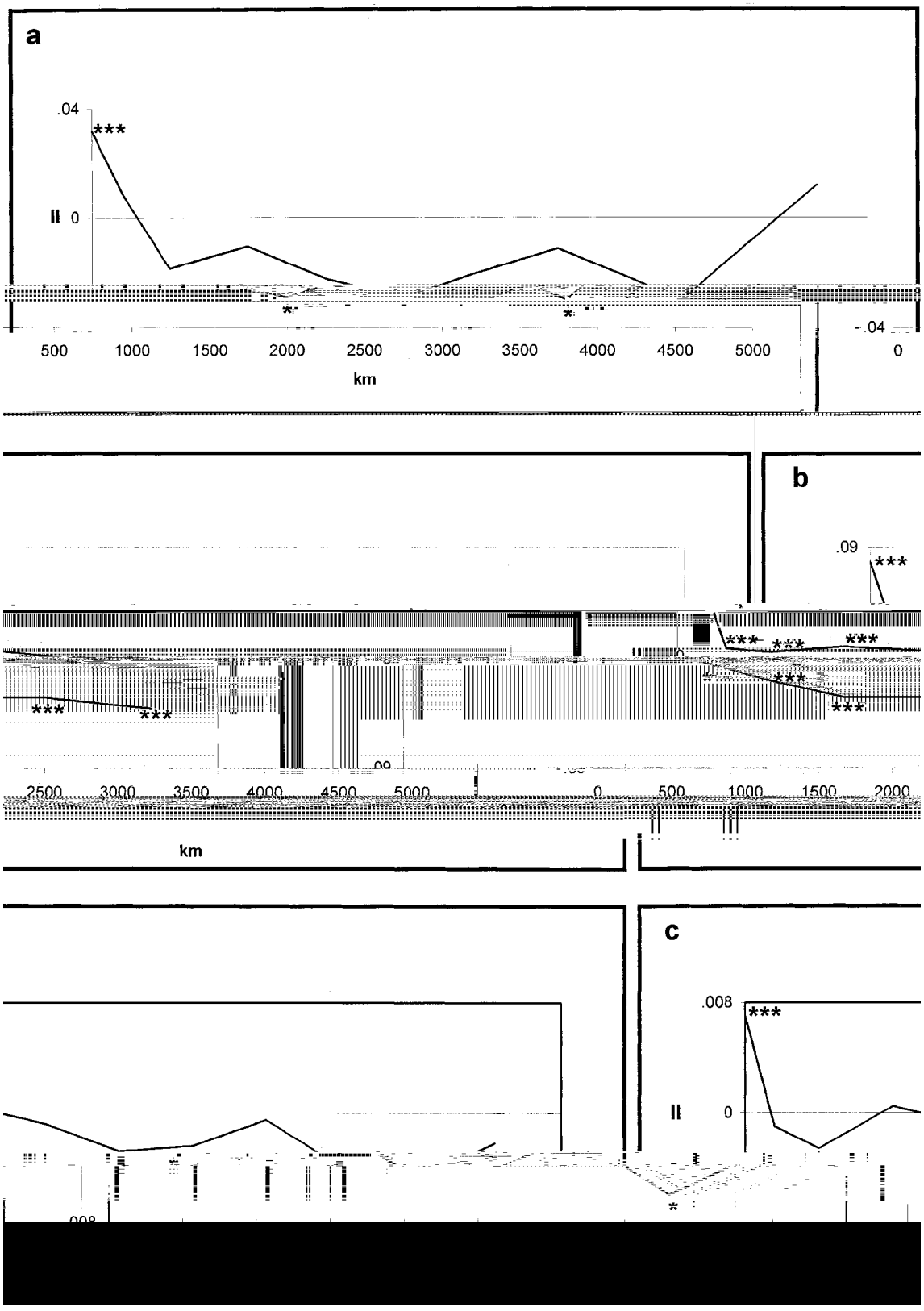

Figure 4 Spatial correlograms calculated along a northwestern-European transect. *a,* Haplogroup V sequences based ontable 4 of Torroni et al. (1998) (individuals from North Africa, Sardinia, and Turkey were excluded). *b,* All haplogroups. *c,* All haplogroups with Saami excluded. The *X*-axis represents 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. 1998*a,* 1998*b;* 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 les bysn po3bi0 U-25ȷT[g to explain the orig280r0-64g280r0*

5(bthele)A -47uld-280(ohav0s345(also)-eftrope;)-[(ch(2)d0(only)-3diter(
–30[(3015)i(zmd300ofol)66or
modl5(b)-2ma-2909e5(bzo50(sa34acteri34ace)-455(are)-4-360.0(e9(bT

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 Eu-

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 (1998*a*) 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,

zerland 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-tranversion 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, Easteal S (1998) Departure from neutrality at the mitochondrial NADH dehydrogenase subunit 2 gene in humans, but not in chimpanzees. Genetics 148:409–421