Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 9171–9175, September 1995
Evolution

Geographical structuring in the mtDNA of Italians

GUIDO BARBUJANI*†‡, GIORGIO BERTORELLE†, GIULIA CAPITANI*, AND ROSARIA SCOZZARI§

*Dipartimento di Scienze Statistiche, Università di Bologna, via Belle Arti 41, 40126 Bologna, Italy; †Dipartimento di Biologia, Università di Padova, via Trieste 75, 35121 Padua, Italy; and §Dipartimento di Genetica e Biologia Molecolare, Università di Roma "La Sapienza," Piazzale Aldo Moro 7, 00100 Rome, Italy

Communicated by Robert R. Sokal, State University of New York, Stony Brook, NY, June 30, 1995

ABSTRACT **Geographical patterns of mtDNA variation were studied in 12 Italian samples (1072 individuals) by two different spatial autocorrelation methods. Separate analyses of the frequencies of 12 restriction morphs show North–South clines, differences between Sardinia and the mainland populations, and the effects of isolation by distance. A recently developed autocorrelation statistic summarizing molecular similarity at all sites (AIDA; autocorrelation index for DNA analysis) confirms the presence of a clinal pattern; differences between random pairs of haplotypes tend to increase with their geographical distance. The partition of gene diversity, however, reveals that most variability occurs within populations, whereas differences between populations are minor (G_{ST} = 0.057). When the data from the 12 samples are pooled, two descriptors of genetic variability (number of polymorphic sites and average sequence difference between pairs of individuals) do not behave as expected under neutrality. The presence of clinal patterns, Tajima's tests, and a simulation experiment agree in suggesting that population sizes increased rapidly in Italy and Sicily but not necessarily so in Sardinia. The distribution of pairwise sequence differences in the Italian peninsula (excluding Sardinia) permits a tentative location of the demographic increase between 8000 and 20,500 years ago. These dates are consistent with archaeological estimates of two distinct expansion processes, occurring, respectively, in the Neolithic and after the last glacial maximum in the Paleolithic. Conversely, there is no genetic evidence that such processes have had a major impact on the Sardinian population.**

Most studies of mtDNA variation in humans have inferred evolutionary processes by reconstructing history—i.e., genealogies of haplotypes (1–4). With one exception (5), geographical information has been disregarded, or it has been used simply to classify populations, as an alternative to the subjective criteria of racial classification (6, 7). There is no doubt, however, that spatial patterns of genetic diversity also contain useful information for evolutionary inferences. One reason for this omission may lie in the relative paucity of statistical tools suited for spatial analysis of molecular data.

Genetic variation in space can be summarized by spatial autocorrelation measures (see, e.g., ref. 8). These statistics have also been used to test hypotheses on past demographic processes (9, 10). Here we apply them to a data base of mtDNA restriction fragment length polymorphism data. We reconstruct patterns of genetic variation from molecular information, and we try to draw inferences about the underlying microevolutionary phenomena.

In the first part of this paper, we analyze the frequencies of 12 common restriction morphs, treating them as if they were allele frequencies. In the second part, we use a recently developed autocorrelation statistic, AIDA (autocorrelation index for DNA analysis) (11), which compares DNA sequences (and not only their frequencies) at several spatial lags. The unit of analysis, then, is the individual haplotype and not the population, which has two main consequences: (*i*) a measure of intrapopulation genetic relatedness is estimated by pairwise comparing all individuals of the same sample; (*ii*) the sample size increases, giving the test higher statistical power. For instance, in a traditional study of 20 samples, autocorrelation statistics are evaluated on the basis of $20 \times 19/2 = 190$ comparisons: but, if the average sample size is 20 individuals, the AIDAs are based on $400 \times 399/2 = 79{,}200$ comparisons.

The results obtained using the two approaches on the same data set are not identical. The frequencies of some mtDNA morphs probably reflect recent processes of drift and gene flow, whereas geographical structuring at the sequence level, strictly depending on the appearance of new mutations, seems to be much more related to differentiation events occurring in a remote past. To interpret these findings, in the final part of this study we calculate some descriptors of genetic heterogeneity. Their distributions suggest that different demographic phenomena affected the peninsular Italian versus the Sardinian populations, with the former, but not the latter, showing evidence of a demographic expansion. Based on mtDNA diversity, such an expansion seems to have occurred either in the early Neolithic or in the late Paleolithic.

SPATIAL AUTOCORRELATION STATISTICS

Spatial autocorrelation is defined as the dependence of one variable upon its values at other localities (12). Patterns of allele frequencies may be summarized by autocorrelation statistics, generally Moran's *I*, calculated in discrete distance classes between all possible pairs of populations. For large samples, Moran's *I* values range from −1 (negative autocorrelation, indicating genetic dissimilarity in a distance class) to +1 (positive autocorrelation, or genetic similarity). The expected value is very close to 0 under a randomization hypothesis (12).

AIDAs are two autocorrelation statistics developed for the study of molecular data (11). They measure whether, and to what extent, pairs of haplotypes (rather than pairs of haplotype frequencies) resemble each other at various distances in space. The AIDA used in this study, called *II* by analogy with Moran's *I*, varies between −1 and +1, with its expected value being close to 0; it can be interpreted in the same way as Moran's *I*. If haplotypes are coded as strings of binary digits (as described in ref. 5), *II* can be calculated in arbitrary distance classes as

$$II = \frac{n\sum_{i=1}^{n-1}\sum_{j>i}^{n} w_{ij}\sum_{k=1}^{S}(p_{ik}-\bar{p}_k)(p_{jk}-\bar{p}_k)}{W\sum_{i=1}^{n}\sum_{k=1}^{S}(p_{ik}-\bar{p}_k)^2},$$

where n is the sample size; W is the number of pairwise comparisons in the distance class of interest; p_{ik} and p_{jk} represent the bases observed at the kth site in the haplotypes

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: AIDA, autocorrelation index for DNA analysis.

‡To whom reprint requests should be sent at the * address.

of the ith and the jth individuals, respectively; $\bar{p}_k$ is the average of the p values at the kth site across all individuals; and the weights w_{ij} are 1 if individuals i and j fall in the distance class of interest, otherwise they are 0. Summation is over the S polymorphic sites for all individuals in the sample. The error of *II* is calculated by repeatedly randomizing haplotypes with respect to their spatial location, each time calculating a pseudovalue of *II*; an empirical null distribution of pseudovalues is thus constructed, and the significance of the observed value is assessed by comparison with it.

The overall significance of the entire set of autocorrelation coefficients at different distances (correlogram) was evaluated by the Bonferroni criterion (13). Under the null hypothesis of isolation by distance, a correlogram is expected to show a decrease of genetic relatedness, from positive significance at a short distance to insignificant (14).

THE DATA

We collected mtDNA restriction fragment length polymorphism data from six studies in the literature (15–20) (Fig. 1). All these samples had been typed using *Bam*HI, *Hae* II, *Msp* I, *Ava* II, and, with one exception, *Hinc*II restriction endonucleases. Because for *Hinc*II there is no polymorphism in Italy (17), we focused on the other four enzymes. The frequencies of the 12 most common restriction morphs were analyzed by traditional spatial autocorrelation.

Haplotype sequences were inferred from the restriction patterns following the scheme proposed by Excoffier (2, 5). Each individual haplotype (or mtDNA type) was represented by an array of 0s or 1s indicating, respectively, that each DNA site was identical to, or different from, the corresponding site in the sequence published by Anderson *et al.* (21). The choice of that particular sequence does not affect any of the statistics we evaluated. The data thus transformed were then used to calculate AIDAs. Geographical distances between localities were great-circle distances.

RESULTS

Twenty-six sites appeared polymorphic among the 1072 individuals studied, defining 42 different mtDNA types (Table 1). Twelve morphs showed substantial variation in frequency among populations.

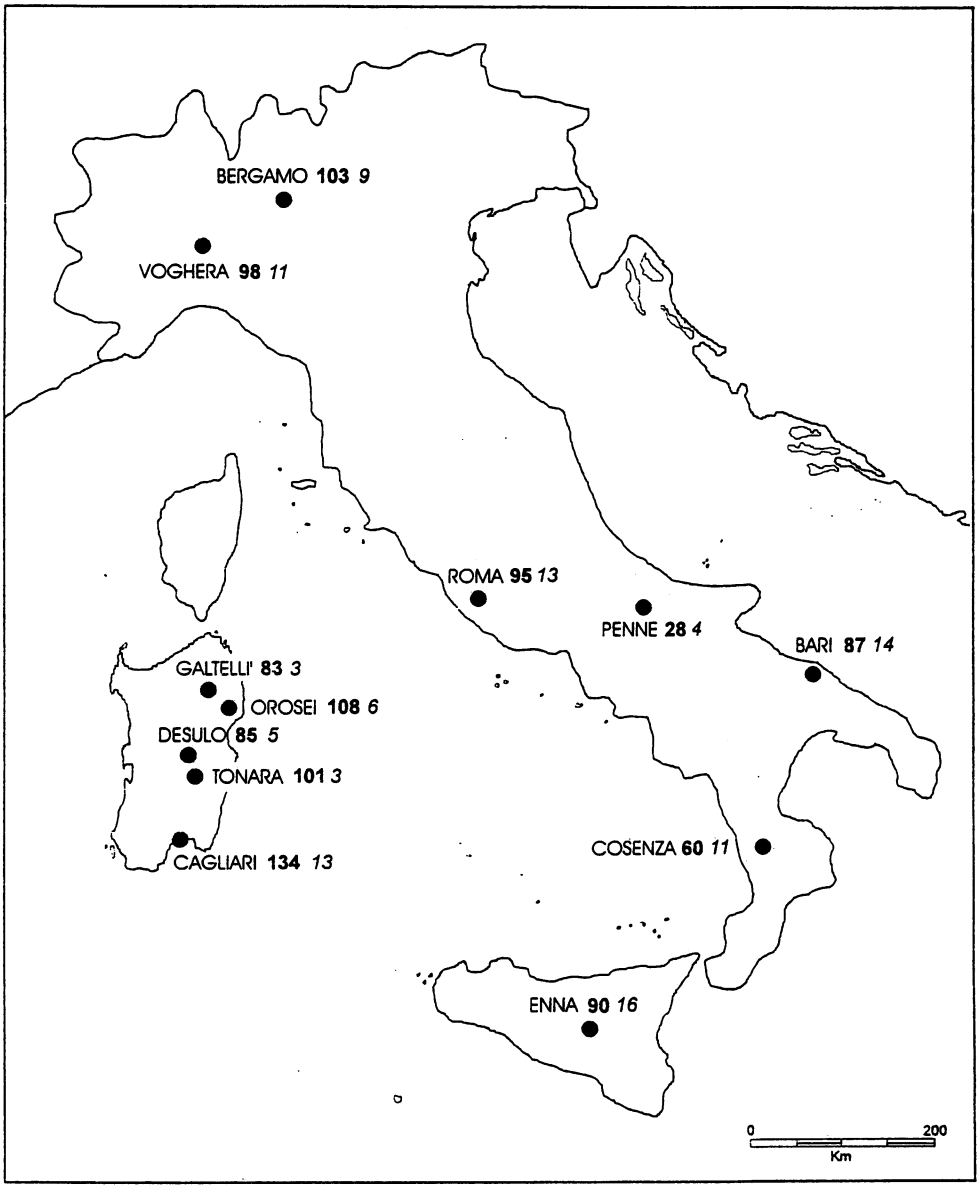

FIG. 1. Samples considered. Numbers in boldface are sizes, and numbers in italics are numbers of different mtDNA types in the samples.

Table 1. Relative frequency of mtDNA types in the 12 samples (the first 5 samples are Sardinian)

Type	Desulo	Tonara	Orosei	Galt.	Cagl.	Vogh.	Berg.	Roma	Enna	Bari	Cosen.	Penne
1	0.588	0.842	0.926	0.916	0.754	0.737	0.767	0.621	0.556	0.529	0.667	0.714
2									0.011	0.034		
6	0.059	0.020			0.045	0.006	0.117	0.084	0.033	0.080	0.067	0.179
11			0.009					0.032		0.057		
13					0.015							
18	0.306	0.138	0.028	0.036	0.075	0.070	0.039	0.137	0.100	0.092	0.067	0.071
21					0.015	0.030	0.019	0.032	0.122	0.046	0.033	
22					0.007	0.010		0.011		0.034	0.017	
23									0.011			
24										0.012	0.017	
28	0.024				0.007						0.017	
34									0.011			
38							0.019	0.021		0.012		
39					0.015							
42								0.010	0.011			
55					0.007			0.010	0.056	0.012	0.050	0.036
56			0.009				0.009	0.010	0.011			
57					0.030	0.030	0.009	0.010	0.022	0.034	0.017	
58–59								0.020				
60–62					0.029							
64			0.009									
66	0.024											
72–73									0.022			
74			0.018				0.009					
75				0.048			0.009					
76–77									0.022			
78–80										0.069		
81–85						0.050						
86						0.010					0.033	
87											0.017	
190									0.011			
J_k	0.444	0.728	0.859	0.843	0.578	0.551	0.605	0.415	0.340	0.306	0.459	0.548

Galt., Galtellì; Cagl., Cagliari; Vogh., Voghera; Berg., Bergamo; Cosen., Cosenza. J_k is Nei's gene identity (23).

Significant departures from spatial randomness were observed in the frequency distributions of seven morphs (Table 2). The positive autocorrelation in the first distance class was expected under the hypothesis of isolation by distance. However, only for the frequency of one morph (*Hae* II-1) was a pure isolation-by-distance pattern observed (14). In addition, there were two gradients (*Ava* II-2 and *Ava* II-3), two cases of long-distance differentiation (*Bam*HI-2 and *Ava* II-1), and two depressions (*Hae* II-2 and *Hae* II-3)—that is to say, two significant patterns where the largest genetic divergence was observed at intermediate distances (≈600 km). Different distance classes were also used (data not given), all of them confirming the patterns described in Table 2.

AIDAs were calculated twice, once based only on the seven samples from continental Italy and Sicily (hereafter referred to as the Italy samples) and once considering all 12 samples (Italy + Sardinia samples). All autocorrelation coefficients were comparatively low, with none of them exceeding 0.04 (Fig. 2). Both correlograms show positive and significant autocorrelation at distance 0 (i.e., when the comparison is between sequences sampled in the same population) and significant negative peaks at large distances. The correlogram calculated from the Italy samples shows significant similarity around 400 km (reflecting genetic resemblance between the Roma and Bari samples). In the Italy + Sardinia correlogram, significant negative autocorrelation also occurs between 600 and 700

Table 2. Autocorrelation coefficients (Moran's *I*) for the frequencies of 12 mtDNA morphs in 12 Italian populations including Sardinia

	Distance class, km				
Morph	1–261	262–514	515–614	615–1010	Correlogram probability
*Bam*HI-1	−0.35	0.01	0.11	−0.19	NS
*Bam*HI-2	0.03	0.03	0.12	−0.57**	**
*Bam*HI-3	−0.37	−0.02	0.06	−0.09	NS
Hae II-1	0.42**	0.02	−0.45	−0.36	*
Hae II-2	0.56**	0.08	−0.98***	−0.01	***
Hae II-3	0.02	−0.08	0.15*	−0.48	*
Msp I-1	−0.33	−0.02	−0.02	−0.05	NS
Msp I-4	−0.34	0.00	−0.05	−0.02	NS
Ava II-1	0.16	0.15*	−0.10	−0.59	*
Ava II-2	0.54**	−0.20	0.07	−0.77***	***
Ava II-3	0.56**	0.28**	−0.43*	−0.77***	***
Ava II-9	−0.32	−0.04	−0.01	−0.04	NS

Class limits defined so as to obtain four classes with the same amount of information. NS, not significant. Two-tailed significance: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

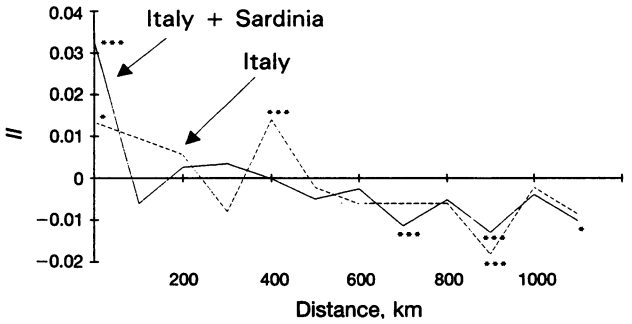

FIG. 2. Coefficients of spatial autocorrelation obtained by AIDA in Italy (dashed line) and Italy + Sardinia (solid line). *x* axis, distances between individuals (upper class limits); *y* axis, autocorrelation coefficient. *, $P < 0.05$; ***, $P < 0.005$ (two-tailed tests).

km—namely, in the distance class including most comparisons between Sardinian and non-Sardinian individuals.

To better understand the distribution of mtDNA haplotype variation, Nei's measures of genetic identity and diversity were calculated (23). Within the 12 samples, the gene identity J_k varied between 0.306 and 0.859 (Table 1), with an average value J_S of 0.556. Gene identity in the total population, J_T, was 0.529. The total gene diversity, H_T, was then 0.471, which can be partitioned in a within-sample component ($H_S = 0.444$) and a between-sample component ($D_{ST} = 0.027$). mtDNA diversity in Italian + Sardinian populations can therefore be attributed only for <6% ($G_{ST} = D_{ST} / H_T = 0.057$) to differences between localities. This value, which is 1/6th that ($G_{ST} = 0.351$) estimated for the entire human population (3), reflects essentially the fact that the most common haplotypes are the same and have similar frequencies in all the populations analyzed; the probability of two alleles to be different is only slightly larger when the alleles are sampled from different populations rather than from the same one.

DISCUSSION

The frequencies of several mtDNA morphs show nonrandom spatial patterns in Italy. North–South gradients and depressions are observed, the latter reflecting morph frequency differences between Sardinia and the mainland. When spatial variation is analyzed by AIDA—i.e., when the degree of sequence difference between alleles is considered—the patterns found are similar but less marked. Sardinian samples are still genetically differentiated; Northern and Southern populations seem to be separated by a step, rather than by a smooth cline, and their differences are significant beyond 800 km.

The establishment of geographical heterogeneity in the distribution of DNA sequences depends widely on the long-term process of mutation, which has a negligible effect on the frequencies of DNA or protein markers. This may explain why the spatial patterns inferred from morph frequencies and haplotype sequences are not identical, with the former resembling the clinal and depression patterns of allele frequencies observed in the same area (24). The time elapsed since separation of geographically close groups may have been too short for mutation to differentiate them sharply but sufficient for genetic drift and gene flow to produce some degree of spatial structuring. The low fraction of haplotype diversity due to differences between localities also supports the idea of a recent separation of Italian groups from a genetically heterogeneous ancestral pool of genes.

Although limited in its extent, genetic diversity does show a clinal component. Clines of selectively equivalent alleles are commonly attributed to gene flow and especially to processes of directional migration. By itself, however, gene flow does not result in wide and stable clines, unless it is accompanied by an increase in population size (22). Tajima (25) has demonstrated that such demographic expansions, among other phenomena, affect the relationship between two measures of DNA variation, the number of polymorphic sites and the average number of sequence differences between pairs of individuals. We applied Tajima's test to the data of this study and found a significant departure from equilibrium in the direction expected after rapid population growth (Sardinia: $D = 1.77, P < 0.05$; Italy: $D = -1.76, P < 0.05$; Italy and Sardinia: $D = -1.91, P < 0.05$). Tajima's tests, therefore, agree with the other results of this study and suggest that the spatial patterns we described may reflect a demographic expansion.

To locate such an expansion in time, we focused again on the pairwise differences between individuals, or mismatches. Rogers and Harpending (26) showed that, when a population grows rapidly, the mismatch distribution is smooth and has a single peak. Its mean and variance contain information on the population size and on the age of the expansion (26–28). As a preliminary step, we evaluated an ad hoc statistic called raggedness, which describes the shape of mismatch distributions, and has proved effective in discriminating expanding and nonexpanding populations (27, 28). Raggedness values in Sardinia and in peninsular Italy were equal to 0.355 and 0.075, respectively. The latter figure is well within the range obtained by Harpending (29) in simulations of expanding populations, but the former is not. In other words, the mismatch distribution in Sardinia corresponds to that expected in a demographically stable population. As a consequence, estimating the time since the expansion was justified only for the Italy data.

In principle, population expansions occurring 1, 2, ... S units of mutational time before the present (B.P.) are expected to result in mismatch distributions with a mean at 1, 2, S differences, respectively (26), minus a correction factor that depends on the variance (28). In this study, the observed mean is 1.16 for the Italy data. The mutation rate in the region we studied, u, can be estimated as $2\,k\mu$, where k quantifies the number of nucleotide sites surveyed by restriction analysis and μ is the mutation rate per site (30). We estimated k according to Nei and Tajima (30). Using two estimates of μ, 2.5 and 5×10^{-7} per generation (26), u falls between 0.000184 and 0.000369, and Rogers' method of moments (28) indicates that a population expansion compatible with our data might have occurred between 821 and 410 generations ago, respectively, or between 8200 and 20,525 years ago (assuming that a generation lasts 20 or 25 years). With the same method, the effective population size before the expansion can be estimated between 1160 and 2326 females. For the sake of completeness, we carried out analogous calculations on the Sardinian data and estimated that a population increase compatible with mtDNA data, if any, should have occurred 460–1100 years ago.

Unimodal (or smoothly decreasing) distributions of pairwise differences may also be due to processes other than population expansions. Selection (2) or high levels of homoplasy (31) may produce the same pattern, because both reduce the correlation between sequences. By the simulation approach described in Bertorelle and Slatkin (32), we tried to establish whether another observed measure of genetic variation, the number of polymorphic sites, S, is really compatible with an expansion in Italy but not in Sardinia. The choice of S seemed appropriate because the number of segregating sites and the mismatch distribution have a different dynamic during a population expansion (25). To that end, we used the demographic parameters estimated from the mismatch distribution to simulate 1000 genealogies of two samples of genes, of size 511 (corresponding to the five Sardinian samples) and 561 (corresponding to the seven Italian samples). Mutations were Poisson-distributed along the genealogical tree, and the mutation rate varied across sites (see ref. 33). S was calculated at the end of each cycle of the simulation, and its empirical frequency distribution was thus constructed. The number of polymorphic

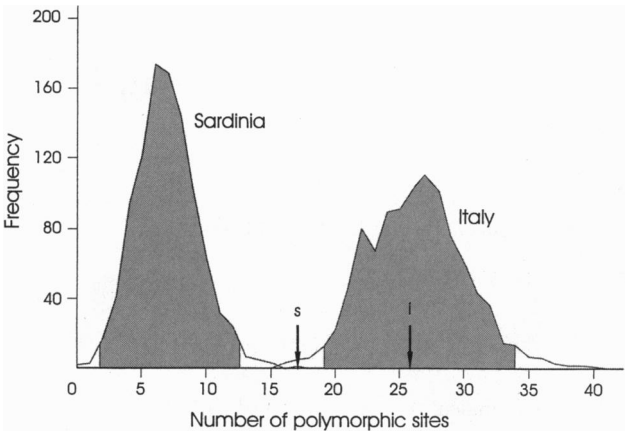

FIG. 3. Simulation results. *x* axis, number of polymorphic sites at the end of one simulation experiment; *y* axis, number of simulations. The 95% confidence regions are shaded. Arrows indicate observed number of polymorphic sites in Sardinia (s) and in Italy (i). The latter falls near the mode of the distribution of the values obtained by simulating an expanding population, but the former does not.

sites observed in the Sardinian samples is out of the 95% confidence region of that distribution (Fig. 3), whereas the value observed in the Italian samples falls well within that region. Although not a proof, this result corroborates other findings of this study in suggesting that mtDNA diversity in Italy and Sardinia reflects two different population processes, with only Italy showing good evidence for rapid growth.

Many European populations have expanded during the Neolithic with the diffusion of farming (22). Neolithic farmers, presumably spreading from the South to East, reached Italy between 7500 and 6000 years ago (22, 34), causing a large, perhaps 100-fold (35), population increase. The clinal patterns observed along the Italian peninsula may result from a population expansion; its age, estimated from the distribution of the pairwise differences, falls in a range whose lower limit is not much greater than 7500 years.

This is not the only archaeologically documented process that may account for our results. There is very little evidence of human presence in Italy between 25,000 and 30,000 B.P. (36). With the retreat of glaciers, the number of settlements increases, suggesting a steady population growth after the last glacial maximum, 20,000–16,000 B.P. (36). Also this Paleolithic date falls in the time interval estimated from the mtDNA mismatch distribution. A Paleolithic expansion would seem more plausible because of the better overlap of genetic and archaeological dates and because the mismatch distribution tends to reflect the oldest expansion event in a population's history (26). On the other hand, population growth was presumably slow in the Paleolithic (36), perhaps too slow to substantially affect the mtDNA mismatch distribution, and so a Neolithic expansion cannot be ruled out.

Inferences on demographic processes from DNA data are approximate, and some underlying assumptions are impossible to test. Nevertheless, this correspondence between genetic and archaeological results is intriguing. It may mean that patterns and levels of mtDNA variation in peninsular Italy reflect a phenomenon that archaeologists already described (36, 39)—that is, either of two phases of demographic growth occurring, respectively, after the last glacial maximum in the upper Paleolithic and at the beginning of the Neolithic.

This does not seem true of Sardinian populations. Their genetic divergence from peninsular and Sicilian populations and the shape of their mismatch distribution suggest they have been isolated and demographically stable or nearly so. Radiocarbon dating of human fossils indicates that Sardinia was colonized around 9000 years ago; the extreme physical characteristics of its pre-Neolithic population is interpreted as a sign of severe isolation (37). Evidence for farming activities is much more recent, and large population sizes may have been reached only 3000 years ago (38). Our findings do not contradict these views. They suggest that the population ancestral to Sardinians increased in numbers more slowly than other Italian groups and had a somewhat independent evolutionary history.

We thank Henry Harpending, Marta Mirazon Lahr, Andrea Pilastro, Alan Rogers, Ornella Semino, and Silvana Santachiara for fruitful discussion and for giving us access to unpublished material. This work was supported by Grant 94.2270 of the Italian Consiglio Nazionale delle Ricerche.

1. Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C. & Cavalli-Sforza, L. L. (1983) *J. Mol. Evol.* **19,** 255–271.
2. Excoffier, L. (1990) *J. Mol. Evol.* **30,** 125–139.
3. Merriwether, D. A., Clark, A. G., Ballinger, S. W., Schurr, T. G., Soodyall, H., Jenkins, T., Sherry, S. T. & Wallace, D. C. (1991) *J. Mol. Evol.* **33,** 543–555.
4. Stoneking, M. (1993) *Evol. Anthropol.* **2,** 60–73.
5. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) *Genetics* **131,** 479–491.
6. Bowcock, A. M., Kidd, J. R., Mountain, J. L., Hebert, J. M., Carotenuto, L., Kidd, K. K. & Cavalli-Sforza, L. L. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 839–843.
7. Templeton, A. R. (1993) *Am. Anthropol.* **95,** 51–72.
8. Sokal, R. R., Harding, R. M. & Oden, N. L. (1989) *Am. J. Phys. Anthropol.* **80,** 267–294.
9. Sokal, R. R. & Menozzi, P. (1982) *Am. Nat.* **119,** 1–17.
10. Barbujani, G., Pilastro, A., DeDomenico, S. & Renfrew, C. (1994) *Am. J. Phys. Anthropol.* **95,** 137–154.
11. Bertorelle, G. & Barbujani, G. (1994) *Genetics* **140,** 811–819.
12. Sokal, R. R. & Oden, N. L. (1978) *Biol. J. Linn. Soc.* **10,** 199–228.
13. Oden, N. L. (1984) *Geogr. Anal.* **16,** 1–16.
14. Barbujani, G. (1987) *Genetics* **117,** 777–782.
15. Brega, A., Scozzari, R., Maccioni, C., Iodice, C., Wallace, D. C., Bianco, I., Cao, A. & Santachiara Benerecetti, A. S. (1986) *Ann. Hum. Genet.* **50,** 327–338.
16. Semino, O., Torroni, A., Scozzari, R., Brega, A., De Benedictis, G. & Santachiara Benerecetti, S. (1989) *Ann. Hum. Genet.* **53,** 193–202.
17. De Benedictis, G., Rose, G., Passarino, G. & Quagliariello, C. (1989) *Ann. Hum. Genet.* **53,** 311–318.
18. De Benedictis, G., Rose, G., Caccio, S., Picardi, P. & Quagliariello, C. (1989) *Gene Geogr.* **3,** 33–40.
19. Sartoris, S., Varetto, O., Migone, N., Cappello, N., Piazza, A., Ferrara, G. B. & Ceppellini, R. (1988) *Ann. Hum. Genet.* **52,** 327–340.
20. Brega, A., Mura, G., Caccio, S., Semino, O., Brdlicka, R. & Santachiara-Benerecetti, S. (1994) *Gene Geogr.* **8,** 45–54.
21. Anderson, S., Bankier, A. T., Barrel, B. G., de Brujin, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1981) *Nature (London)* **290,** 457–465.
22. Ammerman, A. J. & Cavalli-Sforza, L. L. (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, Princeton).
23. Nei, M. (1987) *Molecular Evolutionary Genetics* (Columbia Univ. Press, New York).
24. Barbujani, G. & Sokal, R. R. (1991) *Hum. Biol.* **63,** 253–272.
25. Tajima, T. (1989) *Genetics* **123,** 597–601.
26. Rogers, A. R. & Harpending, H. (1992) *Mol. Biol. Evol.* **9,** 552–569.
27. Harpending, H., Sherry, S. T., Rogers, A. R. & Stoneking, M. (1993) *Curr. Anthropol.* **34,** 483–496.
28. Rogers, A. R. (1995) *Evolution*, in press.
29. Harpending, H. (1994) *Hum. Biol.* **66,** 591–600.
30. Nei, M. & Tajima, F. (1981) *Genetics* **97,** 145–163.
31. Lundstrom, R., Tavaré, S. & Ward, R. H. (1992) *Math. Biosci.* **112,** 319–335.
32. Bertorelle, G. & Slatkin, M. (1995) *Mol. Biol. Evol.*, in press.
33. Wakeley, J. (1993) *J. Mol. Evol.* **37,** 613–623.
34. Sokal, R. R., Oden, N. L. & Wilson, C. (1991) *Nature (London)* **351,** 143–145.
35. Rendine, S., Piazza, A. & Cavalli-Sforza, L. L. (1986) *Am. Nat.* **128,** 681–706.
36. Mussi, M. (1990) in *The World at 18,000 BP: High Latitudes*, eds. Soffer, O. & Gamble, C. (Unwin, London), pp. 126–147.
37. Spoor, C. F. & Sondaar, P. Y. (1986) *J. Hum. Evol.* **15,** 399–408.
38. Piazza, A. (1993) *Science* **260,** 1767–1769.
39. Renfrew, C. (1987) *Archaeology and Language: The Puzzle of Indo-European Origins* (Jonathan Cape, London).