Network Analyses of Y-Chromosomal Types in Europe, Northern Africa, and Western Asia Reveal Specific Patterns of Geographic Distribution

Patrizia Malaspina,[1] Fulvio Cruciani,[2] Bianca Maria Ciminelli,[1] Luciano Terrenato,[1] Piero Santolamazza,[2] Antonio Alonso,[33] Juraj Banyko,[4] Radim Brdicka,[5] Oscar García,[6] Carlo Gaudiano,[7] Ginevra Guanti,[8] Kenneth K. Kidd,[9] João Lavinha,[10] Madalena Avila,[10] Paola Mandich,[11] Pedro Moral,[12] Raheel Qamar,[13] Syed Q. Mehdi,[13] Angela Ragusa,[14] Gheorghe Stefanescu,[15] Maria Caraghin,[15] Chris Tyler-Smith,[16] Rosaria Scozzari,[2] Andrea Novelletto[1]

[1]Department of Biology, University "Tor Vergata," and [2]Department of Genetics and Molecular Biology, University "La Sapienza," Rome; [3]Instituto Nacional de Toxicologia, Madrid; [4]Department of Anthropology, University of P. J. Safarik, Kosice, Slovak Republic; [5]Institute for Haematology and Blood Transfusion, Prague; [6]Basque Country Police, Bilbao, Spain; [7]Azienda Sanitaria Locale 4, Centro Lotto alle Microcitemie, Matera, Italy; [8]University of Bari, Bari, Italy; [9]Department of Genetics, Yale University School of Medicine, New Haven; [10]Instituto Nacional de Saúde, Lisbon; [11]I. Bi. G., University of Genoa, Genoa; [12]Departmento de Biologia Animal, Universitat de Barcelona, Barcelona; [13]A. Q. Khan Research Laboratories, Islamabad; [14]Oasi Institute, Troina, Italy; [15]Institutul de Cercetari Biologice, Iasi, Romania; and [16]Department of Biochemistry, University of Oxford, Oxford

Summary

In a study of 908 males from Europe, northern Africa, and western Asia, the variation of four Y-linked dinucleotide microsatellites was analyzed within three "frames" that are defined by mutations that are nonrecurrent, or nearly so. The rapid generation and extinction of new dinucleotide length variants causes the haplotypes within each lineage to diverge from one another. We constructed networks of "adjacent" haplotypes within each frame, by assuming changes of a single dinucleotide unit. Two small and six large networks were obtained, the latter including 94.9% of the sampled Y chromosomes. We show that the phenetic relationships among haplotypes, represented as a network, result largely from common descent and subsequent molecular radiation. The grouping of haplotypes of the same network thus fits an evolutionarily relevant criterion. Notably, this method allows the total diversity within a sample to be partitioned. Networks can be considered optimal markers for population studies, because reliable frequency estimates can be obtained in small samples. We present synthetic maps describing the incidence of different Y-chromosomal lineages in the extant human populations of the surveyed areas. Dinucleotide diversity also was used to infer time intervals for the coalescence of each network.

Received December 5, 1997; accepted for publication June 22, 1998; electronically published August 21, 1998.

Address for correspondence and reprints: Dr. Andrea Novelletto, Department of Biology, University of Rome "Tor Vergata," via E. Carnevale, 00173 Rome, Italy. E-mail: novelletto@bio.uniroma2.it

0002-9297/98/6303-0027$02.000

Introduction

Markers of the genetic diversity of the human Y chromosome currently are considered to have the potential to provide information on male-specific patterns of migration in the past. The desirable characteristics of markers of this kind are a high level of polymorphism in the population and the lowest possible incidence of recurrent mutations. Polymorphic single-nucleotide substitutions and *Alu* insertions on the male-specific part of the chromosome have been described (Hammer 1994, 1995; Seielstad et al. 1994; Whitfield et al. 1995; Underhill et al. 1996, 1997; Bianchi et al. 1997; Zerjal et al. 1997), and, in view of the unlikely occurrence of convergent mutational events at each site, Y-chromosomal types defined by these markers are amenable to classic phylogenetic analysis. However, these markers are not very polymorphic. On the other hand, microsatellite loci show higher levels of polymorphism (Roewer et al. 1992, 1996; Mathias et al. 1994), but they cannot be treated by parsimony analysis (Cooper et al. 1996). For statistical treatment of microsatellite data, these articles have introduced network analysis. They have shown that, with some exceptions (see Hammer et al. 1997), a large share of length variation can be accounted for by putative mutational events that introduce a change of a single repeated unit. As opposed to trees, networks are reticulated graphs. They allow more than one pathway to describe convergent mutational events that may generate haplotypes identical in state (Bandelt et al. 1995).

In previous studies we have shown that Y-chromosomal dinucleotide polymorphisms have great power in revealing interpopulation differences (Ciminelli et al.

1995; Scozzari et al. 1997). To better understand the relationships among allelic states at microsatellite loci, we analyzed their variation on chromosomes characterized in terms of the YAP insertion (Hammer 1994) and the presence/absence of the *Hin*dIII site in alphoid units (Tyler-Smith and Brown 1987; Santos et al. 1995)—that is, two polymorphisms with a low incidence of recurrent mutations and in strong disequilibrium (Persichetti et al. 1992). In studying a large group of males from Europe, northern Africa, and western Asia, we observed that identical microsatellite haplotypes are seldom independently generated along different lineages. We also combined network analysis with analysis of the occurrence, frequency, and dispersal of different molecular types. Our results strengthen the idea that the phenetic similarity among haplotypes, represented as a network, is, to a large extent, the result of common descent from one or a few ancestral states and the subsequent molecular-radiation process. On this basis, we present synthetic maps that summarize the incidence of different Y-chromosomal lineages in the extant human populations of the "Caucasian" group. This methodology proves to be extremely powerful in revealing specific patterns in the distributions of molecular types related by descent.

Subjects and Methods

Subjects

We analyzed a total of 908 subjects from 33 populations corresponding to 33 locations (table 1). Each sample consisted of males collected in a specific location (fig. 1*a*) after ascertainment of grandparental origin. For some locations, males were assigned to the location on the basis of their grandparents' origin, independently from the site of collection.

Screen for Y-Chromosomal Variants

The presence of the YAP element was assayed by PCR followed by agarose electrophoresis, as described elsewhere (Hammer and Horai 1995). The presence of the alphoid *Hin*dIII site was tested by either Southern hybridization of digested genomic DNA (Tyler-Smith and Brown 1987) or PCR followed by digestion (Santos et al. 1995). The CAII polymorphic system was detected according to the method of Mathias et al. (1994). This system consists of two Y-specific loci (Kayser et al. 1997), each containing a $(CA)_n$ microsatellite, that are coamplified during PCR. The larger PCR fragment and the smaller PCR fragment generated at this system in each individual were assigned to the allelic classes CAIIa and CAIIb, respectively. We confirm here that, in both classes, the length of the alleles differs in the number of units in a perfect $(CA)_n$ repeat. Whenever a single band was observed, two fragments of the same size were assumed (Mathias et al. 1994). The DYS413 polymorphic system was amplified with a Y-specific primer pair developed by Malaspina et al. (1997). This system, too, consists of two loci, each containing a $(CA)_n$ microsatellite, and PCR fragments were assigned to allelic classes DYS413a and DYS413b, as described above for CAII.

Table 1

Population Samples in Present Study

Region	Sample	No. of Individuals (% of Total)
Northern Europe	1. Norwegian[a]	8 (.9)
	2. Lithuanian	14 (1.5)
	3. Danish[b]	35 (3.9)
Great Britain	4. Londoner[c]	19 (2.1)
Iberian Peninsula	5. Northern Portuguese	25 (2.8)
	6. Southern Portuguese	26 (2.9)
	7. Central Spaniard	20 (2.2)
	8. Basque	24 (2.6)
	9. Southern Spaniard[b]	48 (5.3)
Italian Peninsula	10. Ligurian	20 (2.2)
	11. Venetian[d]	21 (2.3)
	12. Latium[b]	66 (7.3)
	13. Apulian[d]	20 (2.2)
	14. Calabrian[d]	27 (3.0)
	15. Lucanian	24 (2.6)
Sicily	16. Sicilian	21 (2.3)
Sardinia	17. Northern Sardinian[d]	171 (18.8)
	18. Southern Sardinian [d]	29 (3.2)
Central Europe	19. Slovakian	24 (2.6)
	20. Northern Romanian	25 (2.8)
	21. Eastern Romanian	19 (2.1)
Southeastern Europe	22. Albanian	7 (.8)
	23. Continental Greek[d]	21 (2.3)
Crete	24. Cretan[d]	16 (1.8)
Western Asia	25. Turkish[b]	15 (1.7)
	26. Omani[d]	11 (1.2)
	27. United Arab Emirate[d]	21 (2.3)
	28. Iranian	6 (.7)
	29. Pathan (Pakistan)[b]	19 (2.1)
	30. Sindhi (Pakistan)	18 (2.0)
Northern Africa	31. Moroccan Arab	44 (4.8)
	32. Northern Egyptian[c]	24 (2.6)
	33. Southern Egyptian[d]	20 (2.2)
Total		908

[a] Source: Zerjal et al. (1997).
[b] Source: Scozzari et al. (1997).
[c] Source: Persichetti et al. (1992).
[d] Source: Ciminelli et al. (1995).

In the rest of this article, the term "haplotype" will be used to indicate the result of typing for the four microsatellite allelic classes CAIIa, CAIIb, DYS413a, and DYS413b, whereas the term "superhaplotype" will be used to indicate the combination of the microsatellite haplotype with the YAP and the alphoid *Hin*dIII results.

Statistical Analysis

Statistical calculations were performed with SPSS version 6.1.3. Principal component (PC) analysis was performed on the matrix of relative frequencies of the net-

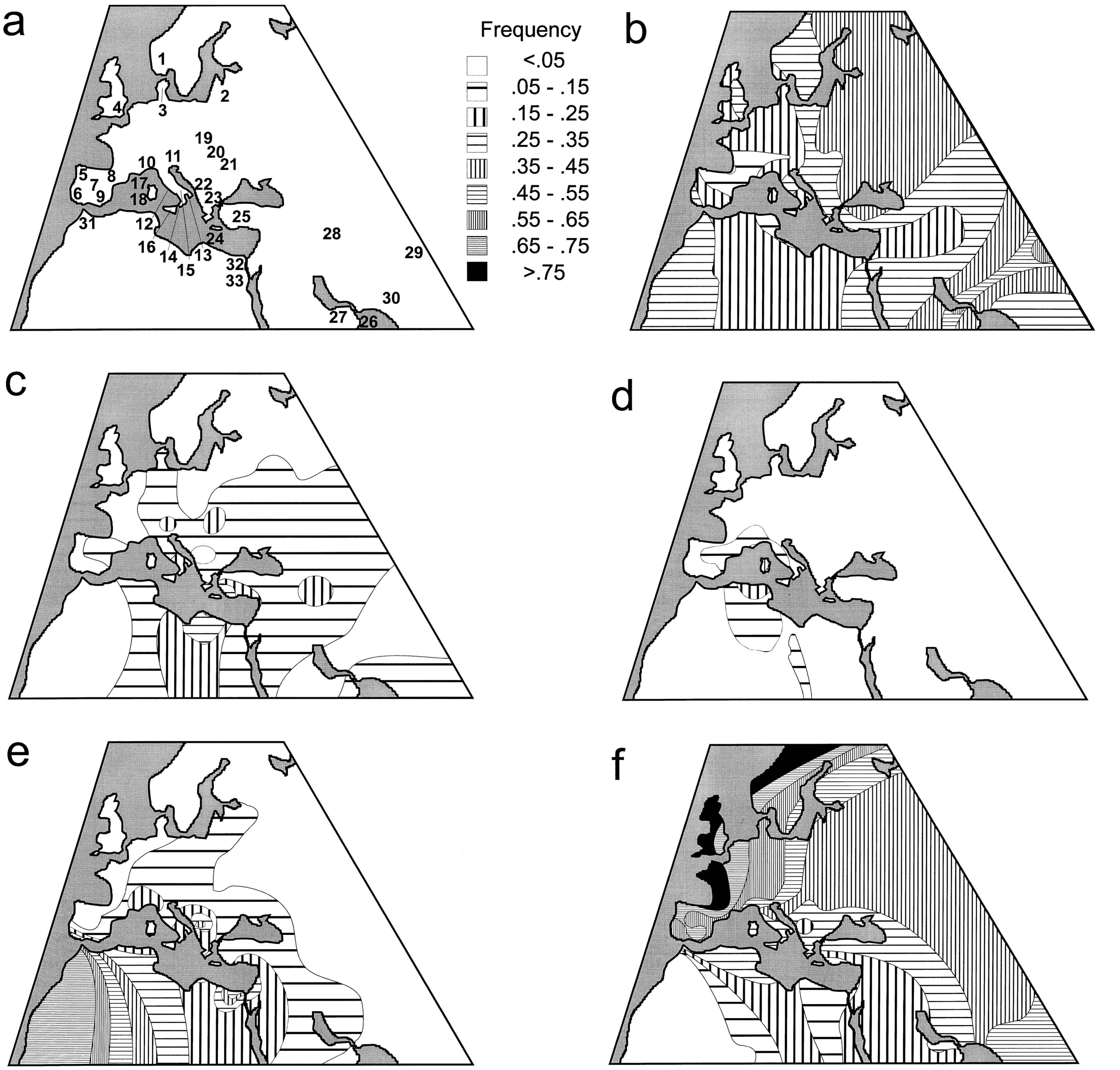

Figure 1 *a,* Surveyed area and locations assigned to 33 populations. Numbers identify populations as designated in table 1. *b–f,* Frequency maps for networks 1.1 (*b*), 1.2 (*c*), 1.3 (*d*), 2.1 (*e*), and 3.1 (*f*). Surfaces spanning frequency intervals of .10, starting from .05, are hatched with increasing density of lines.

works (see below) in the 33 populations, by weighting for sample size. PC values for the 33 populations were recalculated by regression and were used to construct maps as described below.

Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was used to calculate variances within and among populations and F_{ST} (the correlation between random superhaplotypes or networks within subsamples, relative to that in the total population). In order not to introduce a priori assumptions into haplotype relationships, AMOVA did not take into account length differences among alleles; it considered only allele frequencies. In two independent runs of the analysis, we considered either (i) the frequencies of the 223 superhaplotypes or (ii) the frequencies of seven classes of superhaplotypes, each corresponding to one of the networks as defined below. In this phase of the analysis, genetic variances were computed only within samples and between samples, without considering the higher hierarchical levels of population structure.

Estimates of coalescence time intervals, in generations (t), were calculated, according to equation (2) of Gold-

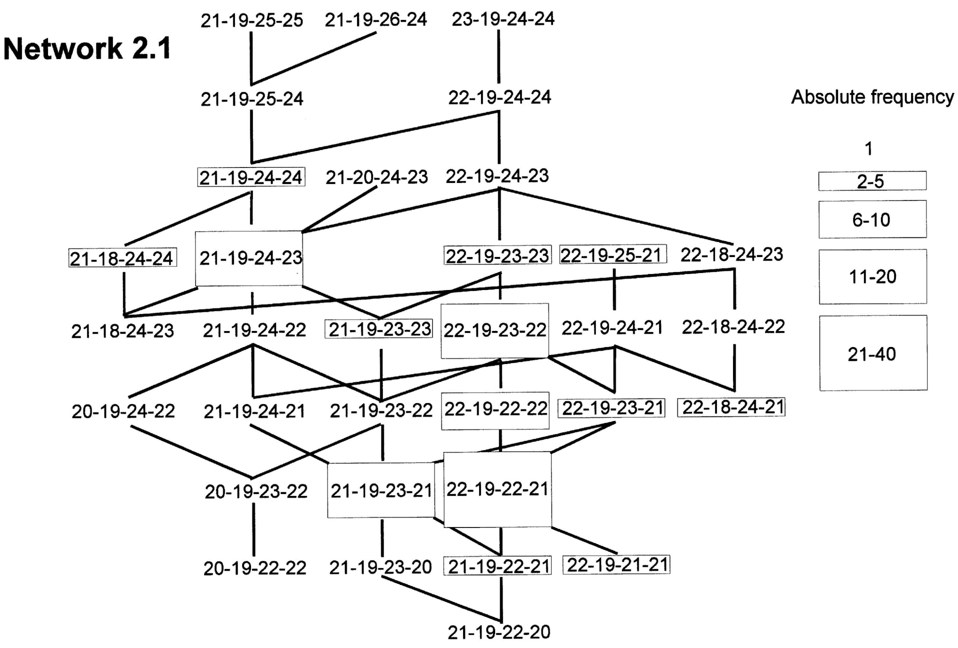

Figure 2 Network 2.1, constructed as described in Subjects and Methods. Haplotypes found in only one individual are not framed. Four discrete sizes of rectangles are used to indicate the abundance of each haplotype in the entire set of 908 subjects.

stein et al. (1996), for the six largest networks, as a function of a range of mutation rates. Initial variance of $(CA)_n$ was considered null. Observed variance of $(CA)_n$ was obtained by averaging the values of the four microsatellite allelic classes. Mutation rate (μ) was varied between 2×10^{-4} and 2×10^{-3}. Effective population size varied between 5,000 (Goldstein et al. 1996) and 1,000, to account for the relatively low number of Y chromosomes within each network.

Network Construction

The similarity between all possible pairs of haplotypes of the same frame was measured by $\Sigma\,|x_i - y_i|$, where x_i and y_i are the $(CA)_n$ at each allelic class in the two haplotypes and the summation is over the four allelic classes. When the result was 1, the two haplotypes were considered "adjacent," were assigned the same network, and were connected in the corresponding graph (see fig. 2). Graphs were constructed as described elsewhere (Cooper et al. 1996), by beginning with the most common haplotype and sequentially adding the adjacent haplotypes. All possible adjacent relationships were indicated by connecting lines. Arrangement of haplotypes within the graph followed two simple rules—(1) all haplotypes with the same combined $(CA)_n$ length were placed on the same horizontal level and (2) line crossovers were minimized—but otherwise was subjective. Since they include only observed haplotypes differing by one CA unit, these networks are, by necessity, "one-step networks" (Bandelt et al. 1995).

Map Construction

Maps were obtained with Surfer System version 4.15 (Golden Software), with the Kriging procedure (Delfiner 1976). Among the advantages of this method are the following: (i) for unsampled points it provides minimum variance estimators for the value of the variable (the network relative frequency, or PC value, in this context) at that point, on the basis of a linear combination of neighboring observed values (best linear unbiased estimators); and (ii) the surfaces of the estimated values of the variable coincide with the observed values at the sampled locations. This latter characteristic is not shared with other methods, such as the inverse-squared-distance method or the polynomial-interpolation method. We used a 69 × 39 grid, and estimates at each grid node were obtained by considering a maximum of 10 nearest points in each quadrant. To represent metric distances between locations, a transformation of actual longitude was adopted (longitude′ = longitude × cos[latitude]).

Results

Y-Chromosomal Haplotypes and Superhaplotypes

In this study we analyzed the joint variation of two Y-specific biallelic polymorphisms and four dinucleotide microsatellites. The two biallelic polymorphisms define four categories, which previous reports have termed "frames" (Persichetti et al. 1992): frame 1 = YAP−,*Hin*d+ ; frame 2 = YAP+,*Hin*d+ ; frame 3 = YAP−,*Hin*d−; and frame 4 = YAP+,*Hin*d−. The two biallelic polymorphisms were in almost complete disequilibrium (table 2) ($D' = 98.2\%$; two-tailed Fisher's test $P < 10^{-4}$), with only one subject (0.11%) having a frame 4 chromosome. These data are in agreement with those reported by Persichetti et al. (1992) and Santos et al. (1996).

In the total sample of 908 males, dinucleotide typing revealed 179 different haplotypes. These resulted in a total of 223 superhaplotypes (detailed data are available on request), when combined with the YAP and alphoid data. A large number of dinucleotide haplotypes were found in each of the three major frames (table 2), because of variation in both the CAII and the DYS413 systems. Frame 1 dinucleotide haplotypes showed the largest variation in size, a total of 49 CA units over the four-band series. Frames 2 and 3 dinucleotide haplotypes varied by 24 and 32 CA units, respectively. These data agree with the observation that frame 2 is derived (Hammer 1995), and they point toward a greater antiquity of frame 1, compared with frame 3. Interestingly, 143 of the dinucleotide haplotypes were each found on one frame only, whereas 29 dinucleotide haplotypes were found on two frames, and 7 were found on frames 1–3. Whenever a haplotype was shared by more than one frame, either it was found at a low frequency in both frames or its frequency on one frame greatly exceeded the frequency in the other(s). This is particularly true for the most common haplotypes of each frame, whose absolute frequencies can be as high as 147 (also see table 3, col. 6). These data indicate that the net result of the production of new length variants and of their extinction along each lineage represented by the frames was not convergent but, instead, generated a majority (143/179 [80%]) of unique haplotypes. In fact, under the hypothesis of a frequent origin and increase in frequency of dinucleotide lengths equal in state, haplotypes common on more than one frame would be expected. This result also shows that consideration of multiple loci reduces the extent of homoplasy due to recurrent mutation at each locus.

Network Analysis

To give a detailed description of the dinucleotide variation, we constructed networks of "adjacent" haplotypes within each frame (see Subjects and Methods). Six major and two minor networks could be constructed by considering the changes of a single dinucleotide unit. Their main characteristics are reported in table 3; one of them is exemplified in figure 2 (other detailed graphs are available on request). Multiple networks within the same frame result from the impossibility of linking haplotypes by fewer than two dinucleotide units. Five separate networks (1.1–1.5) could be formed by haplotypes in frame 1, two networks (2.1 and 2.2) by haplotypes in frame 2, and a single network (3.1) by haplotypes in frame 3. Network 1.1 includes the majority of frame 1 haplotypes. Network 1.2 is characterized by short DYS413a and DYS413b fragments, a peculiarity of European populations (Scozzari et al. 1997). Network 1.3 is identified by a very short CAIIb fragment described by Ciminelli et al. (1995) for the first time in Sardinia. Network 2.1 includes the majority of chromosomes with the YAP insertion. Network 3.1 groups chromosomes devoid of alphoid units with the *Hin*dIII site. Network 1.4 is again made of frame 1 haplotypes, characterized by large CAIIa and CAIIb but small DYS413a and DYS413b fragments. Two haplotypes of frame 1 and two haplotypes of frame 2 form networks 1.5 and 2.2, respectively (see the Appendix, table A1).

The different networks include a variable number of subjects not strictly related to the number of haplotypes within the network (e.g., network 1.1 vs. 3.1 [table 3, cols. 3 and 4]). Overall, the six largest networks represent 94.9% of our population sample. Of the haplo-

Table 2

Incidence and Dinucleotide Diversity of Y-Chromosomal Frames

Characteristic of Frame	Frame 1	Frame 2	Frame 3	Frame 4	Total No.
Defining allele:					
YAP element	−	+	−	+	
Alphoid *Hin*dIII site	+	+	−	−	
No. of subjects	417	136	354	1	908
Dinucleotide diversity:					
No. of superhaplotypes	115	40	67	1	223
Range of $(CA)_n$:					
CAIIa	11–25	19–23	17–25	21	
CAIIb	11–24	18–22	15–23	19	
DYS413a	17–26	21–26	17–26	23	
DYS413b	12–25	14–25	17–24	21	

Table 3

Primary Features of Eight Networks of Y-Chromosomal Haplotypes

					$(CA)_n$							
					Range[b]				Variance			
Frame and Network	No. of Superhaplotypes	No. (%) of Subjects	Major Haplotype[a] $(CA)_n$	No. of Carriers	CAIIa	CAIIb	DYS413a	DYS413b	CAIIa	CAIIb	DYS413a	DYS413b
1:												
1.1	69	233 (25.7)	22-22-22-21	33	−3 to +1	−4 to 0	−2 to +3	−2 to +2	.83	1.45	1.03	.68
1.2	13	73 (8.0)	22-19-17-17	49	−1 to +3	−1 to +1	0 to +1	−1 to +1	.34	.03	.13	.14
1.3	10	85 (9.4)	21-11-21-21	66	−2 to +1	0	0 to +2	−2 to +1	.15	.00	.11	.13
2:												
2.1	33	128 (14.1)	22-19-22-21	22	−2 to +1	−1 to +1	−1 to +4	−1 to +4	.31	.07	.93	1.12
3:												
3.1	50	333 (36.7)	23-19-23-23	147	−4 to +1	−2 to +2	−2 to +3	−3 to +1	.70	.15	.45	.55
1:												
1.4	6	9 (1.0)	24-23-20-20	4	−1 to +1	−1 to +1	0	0	.50	.36	.00	.00
1.5	2	2 (.2)										
2:												
2.2	2	2 (.2)										
Unclassified[c]	38	43 (4.7)										
Total Sample	223	908							1.50	6.92	3.24	2.98

[a] In the order CAIIa-CAIIb-DYS413a-DYS413b.

[b] Compared with major haplotype.

[c] Includes a single frame 4 chromosome.

types of frames 1, 2, and 3, 37 haplotypes (15, 5, and 17, respectively) could not be related by fewer than two CA changes, to either the networks of their frame or to each other. It is interesting to observe that 32/37 of these haplotypes were unique and that the other 5 haplotypes occurred twice in the sample; in 4 of these latter 5 cases, the two subjects were from the same location, suggesting that these haplotypes are strictly local.

Evolutionary Relationships within Networks

The unidimensional variation of dinucleotide length prompts caution in transferring a concept of phylogenetic relationship to a network that phenetically relates a set of haplotypes. Although massive recurrent production of haplotype states can be excluded on the basis of comparisons across frames (see above), this may not be the case within the same frame. We attempted to fit some of the features of the networks to a model in which each of them stems from a founder haplotype that undergoes successive mutational events represented as a network. Three lines of evidence in our data support this view. First, there is the structure of the networks themselves. In fact, in all networks the haplotype with the highest absolute frequency (i.e., the major haplotype; table 3, cols. 5 and 6) is central—that is, it is connected to haplotypes that represent either gains or losses of CA units in at least one of the allelic classes (table 3, cols. 7–10)—and these haplotypes are found at decreasing frequencies as the difference in CA units increases. The overwhelming frequency of the major haplotype is prominent in networks 1.2 and 1.3, in which a single central haplotype accounts for 67% and 78% of the subjects, respectively (table 3). At the center of networks 2.1 and 3.1, clusters of 3 and 7 haplotypes related by changes of a single unit are found, representing 37% and 79% of the subjects, respectively. Among all networks, of the 19 haplotypes representing the maximum or the minimum length of the dinucleotide repeats, 17 are found in one subject each, and none are found in more than three subjects.

Second, grouping of the haplotypes according to the network criterion enables detection of a higher level of population structure. When the 223 superhaplotypes were entered into the AMOVA analysis (Excoffier et al. 1992), the observed variances among and within the 33 samples were .033 and .448, respectively (F_{ST} = .069; when the 223 superhaplotypes were grouped in terms of the six largest networks, variances of .055 and .330 were obtained, respectively (F_{ST} = .143). This result cannot be attributed to the mere pooling of subjects, as long as the total variance is only slightly decreased (.481 vs. .385) whereas the variance among samples shows a net increase. The increase of F_{ST} is in line with the hypothesis that common descent determines that haplotypes similar by state—and, therefore, grouped within the same network—still are found clustered within the same population(s).

Third, there are specific relationships between network structures and the geographic region where their haplotypes were found. The incidence of three of the networks shows a significant covariation with the geographic coordinates of the sampling locations (table 4, cols. 2 and 3). Moreover, we tested the hypothesis that haplotypes derived from a network founder (here assumed to be the major haplotype) still are found in roughly the same home range as the founder. In this context, not only was the home range considered as the area where a superhaplotype is found, but the number of carriers of each superhaplotype also was taken into consideration. To this purpose, we attempted to estimate the average spatial mobility of haplotypes during the radiation process. For each network, we calculated all pairwise linear distances among the sampling locations of subjects carrying the major haplotype (table 4, col. 4), between them and subjects carrying other haplotypes within the same network (table 4, col. 5), and between them and subjects carrying any other haplotype (table 4, col. 6). Under the hypothesis of a dispersal center for the network, the first and second measures are expected to be similar and smaller than the third. This expectation is verified for networks 1.2, 1.3, and 2.1, and the same trend is found for network 1.4. It should be observed that the results for networks 1.1 and 3.1 are basically not informative. In fact, haplotypes of these networks are almost ubiquitous, and distances among them are

Table 4

Parameters of Geographic Distributions, for Haplotypes Pooled According to Network Criterion

	Partial Regression Coefficient		Average Pairwise Sampling Distance between Subjects (km)			
Network	Latitude	Longitude	*A*[a]	*B*[b]	*C*[c]	*D*[d]
1.1	.00295	.00412[e]	2,406	2,432	2,296	2,409
1.2	−.00096	.00015	1,452	1,422[f]	1,792	1,954
1.3	−.00127	−.00113	115	171[f]	1,395	1,664
2.1	−.00788[g]	−.00300	1,645	1,814[f]	2,167	2,285
3.1	.01074[g]	−.00082	1,307	2,047	1,714	2,008
1.4	−.00062	−.00004	1,405	1,239	1,691	1,856

[a] Between any two subjects, each with the major haplotype of the network.

[b] Between any subject with the major haplotype of the network and any subject with another haplotype of the same network.

[c] Between any subject with the major haplotype of the network and any subject with a haplotype of other networks.

[d] Between any subject with the major haplotype of the network and any subject with an unclassified haplotype.

[e] Significant at $P < .001$.

[f] Significantly lower than distance *C* (Mann-Whitney $P < .001$).

[g] Significant at $P < .05$.

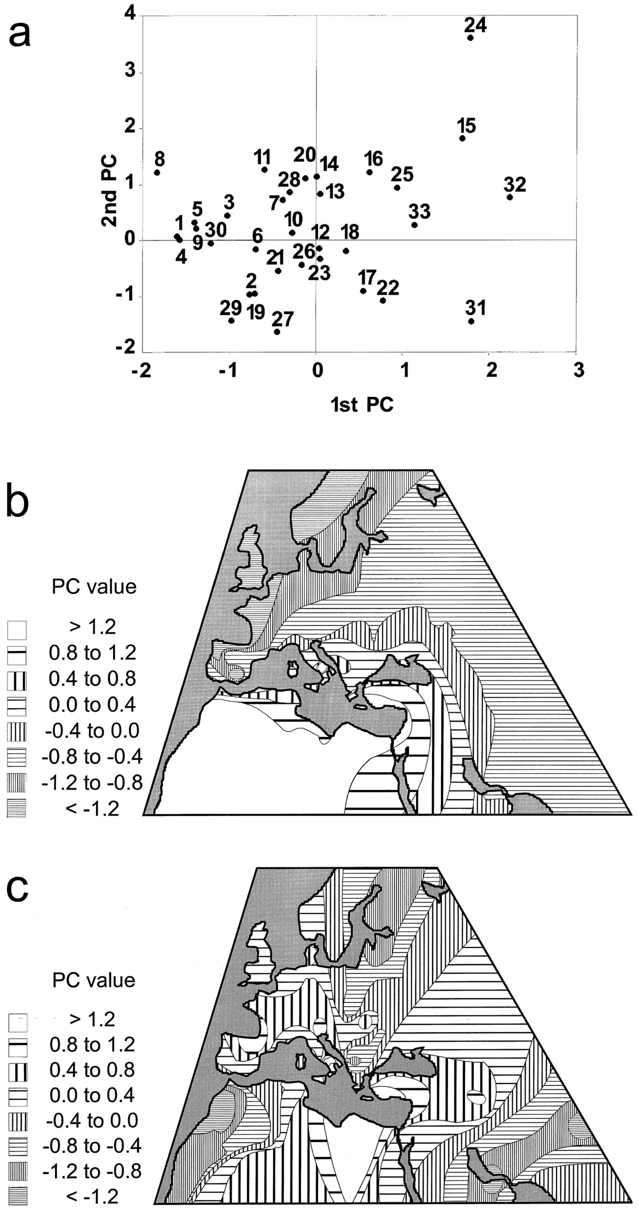

Figure 3 *a,* Display of values of the two PCs of network frequencies for 33 populations. Numbers identify populations as designated in table 1. *b* and *c,* Maps of the first PC (*b*) and second PC (*c*) of network frequencies.

comparable to the average distance (1,923 km) among all 908 subjects. In addition, we cannot exclude the possibility that the lack of evidence for a dispersal center is attributable to the absence of clearly derived characteristics in chromosomes not necessarily closely related to one another. Indeed, data obtained with the marker p12f2, first described by Casanova et al. (1985), suggest that it is able to dissect further network 1.1 (R. Scozzari, unpublished results). It is worth noting that sampling distances between major haplotypes and haplotypes excluded on the basis of classification into networks are often the largest (table 4, col. 7). This reinforces the view that this latter group of haplotypes is geographically marginal and likely to have origins different than those of the other haplotypes.

The results reported above establish definite correlations among molecular variations, incidence of different types within populations, and patterns of geographic clustering or confinement of certain types. Although the evidence is stronger for some networks than for others, these observations support the idea that the phenetic relationships between haplotypes within networks represent, to a good extent, lineages of haplotypes produced by consecutive mutational events.

Geographic Distributions

We used the networks of haplotypes in the 33 sampling locations to construct frequency maps (fig. 1*b–f*). Network 1.1 shows a significant east-to-west gradient (fig. l*b;* also see table 4, col. 3). Superhaplotypes of this network reach the lowest frequency in Basques (<.05). Network 1.2 (fig. 1*c*) has a prominent center of high frequency in Crete, and it also is present in continental Greece and southern continental Italy. Network 1.3 (fig. 1*d*) is essentially confined to Sardinia, with the exception of few instances in continental Italy and central Spain. Elsewhere, we already have described both the peculiar CAII haplotype that characterizes this network and the abundant variation of tetranucleotide alleles associated with it at DYS19 (Ciminelli et al. 1995) and at loci DYS389, DYS390, and DYS393 (Caglià et al. 1997). The frequency distribution of network 2.1 follows a marked south-to-north gradient (fig. 1*e;* also see table 4, col. 2), in agreement with data reported elsewhere (Hammer et al. 1997). The gradient is generally steeper on the African-European boundary, particularly at the Strait of Gibraltar (.73 vs. <.10, in Morocco vs. the southern Iberian peninsula). It is worth noting that different haplotypes within this network contribute differently to this trend. In fact, four of the five most frequent haplotypes are found in at least five locations each, whereas the fifth (21-19-23-21) is found in three locations only, one of which has a high frequency of this haplotype (16/18 subjects are Moroccan Arabs). Network 3.1 largely overlaps with a set of chromosomes identified by markers 92R7 and M911 (Mathias et al. 1994; R. Scozzari and A. Novelletto, unpublished results) and shows a northwest-to-southeast gradient (fig. 1*f;* also see table 4, col. 2). Haplotypes of this network reach notably high frequencies in Europe (as high as .92, in Basques), north and west toward the Alps. A single major haplotype of this network is found in 147 subjects (table 4, col. 6) and in 67% of Basques. Finally, network

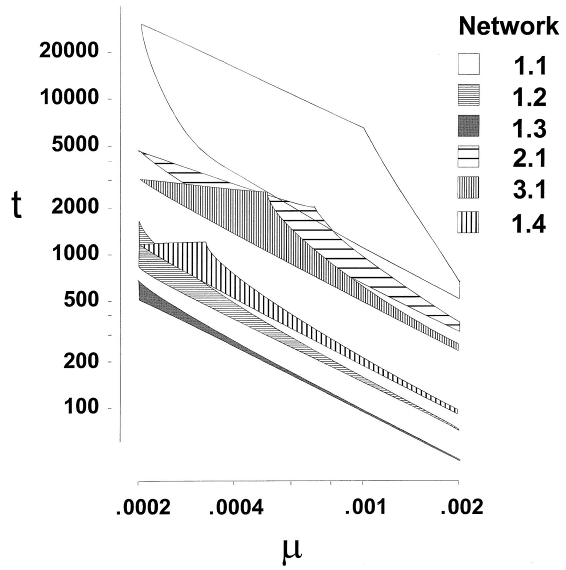

Figure 4 Estimates of *t*, as a function of μ, for the six largest Y-chromosomal networks (Goldstein et al. 1996). Effective population size was 1,000–5,000. These two figures determined the lower-right and the upper-right boundary of each area, respectively. For networks 1.1, 2.1, 3.1, and 1.4, calculations were limited by inapplicability of the formula (*upper straight lines of each area*).

1.4 is confined to the eastern Mediterranean sea and to southern continental Italy. Overall, the heterogeneity of frequencies of the six largest networks in the 33 locations is highly significant (contingency $\chi^2 = 716$; 160 df; $P < 10^{-5}$).

We synthesized the above-discussed data by PC analysis. The first PC explained 29.6% of total variance, not much different from the values found for different series of autosomal data (for a review see Cavalli-Sforza and Minch 1997). This factor proved to be highly correlated with networks 2.1, 3.1, and 1.4 (r = .69, −.89, and .58, respectively), whereas the three remaining networks displayed absolute values of $r < .3$. The second PC explained an additional 23.6% of total variance and was correlated mainly with networks 1.1 and 1.3 (r = −.53 and .73, respectively). Figure 3*a* shows a plot of the 33 populations in the space of the first two PC's. With the exception of Moroccan Arabs, European and northern African populations are arranged in a west-to-east fashion, according to the first PC, with the Basques and other western European populations on the left side and with the Turks, Cretans, and Egyptians on the right side. The western Asian populations do not fit this trend and display negative values for the first PC. Thus, the first PC reaches maximum values in the eastern Mediterranean, decreasing both westward and eastward. The second PC isolates the Cretan population as an effect of the focal distribution of network 1.3. The Basques are identified by a high value for this PC. Figure 3*b* and *c* show maps of the first two PCs. The first PC shows a cline from northern Africa, northward and eastward, that is particularly steep between Morocco and Iberia. With regard to Europe, the PC values decrease in the southeast-to-northwest direction. Interestingly, in addition to a focal distribution of high values in the central Mediterranean, the second PC highlights a clustering of low values in the Balkans.

Inference of Network Antiquity

A notable result of network analysis is its ability to enable partitioning of the total diversity. The variance of $(CA)_n$ in the entire sample approaches the saturation value for the variance of $(CA)_n$—2.52 (Goldstein et al. 1996)—at the Y-specific mutation/drift equilibrium (for CAIIb, the estimate is inflated by the overrepresentation of Sardinians carrying an unusually short fragment). On the other hand, variances within each network (table 3, cols. 11–14) are far from equilibrium, thus making it feasible to use dinucleotide diversity to estimate network antiquity. We used equation (2) of Goldstein et al. (1996) to evaluate the space of possible values of *t* (the time, in generations, for the coalescence of haplotypes within each network) for a range of mutation rates μ and effective population sizes (N_e) (fig. 4). For $\mu = 5.6 \times 10^{-4}$ (Weber and Wong 1993), the large $(CA)_n$ variance for network 1.1 resulted in an estimate of $t > 3{,}000$ generations, or 60,000–75,000 years. The two largest networks with derived characteristics—that is, networks 2.1 and 3.1—both showed much lower values, t = 1,000–3,000 generations. Finally, the three smaller networks—1.2, 1.3 and 1.4—gave estimates of t =~ 300, ~200, and ~450 generations, respectively. These latter estimates are fairly insensitive to different values of N_e. Use of the lower value of $\mu = 2.5 \times 10^{-4}$ (Gyapay et al. 1994) resulted in nearly a doubling of these estimates but, for networks 1.1, 2.1, and 3.1, led to a vast area of inapplicability of the equation, because of the small N_e values used here.

Discussion

Grouping of Haplotypes into Networks

In this study, the overall variability of Y-linked dinucleotide microsatellites was analyzed, within three frames defined by mutations—the YAP insertion and the loss of alphoid units carrying the *Hin*dIII site—that either are in or approach the condition of nonrecurrence. The first of these mutations is considered a unique event; however, for the second one, Santos et al. (1996) pointed out the need to postulate a non-null, albeit low, recurrence. In these conditions we could detect a set of dinucleotide haplotypes that most likely arose by recurrent mutation, and we assessed their frequency as being low. This low identity by state is in contrast with data obtained with tetranucleotides (Ciminelli et al. 1995; Cooper et al. 1996; Caglià et al. 1997; Hammer et al. 1997; Zerjal et al. 1997) and can be explained by a lower mutation rate of the $(CA)_n$ dinucleotides (Weber and Wong 1993; Gyapay et al. 1994; Heyer et al. 1997). The net result is a divergent accumulation of variants across frames in our data set. Our network analyses, performed within frames, considered only changes of a single dinucleotide unit (Cooper et al. 1996; Heyer et al. 1997) and, indeed, led to the grouping of the majority of haplotypes. This continuity between haplotypes was broken in frames 1 and 2, where more than one network could be constructed. Some of these networks could be linked if it is assumed that one or a few intermediates are missing because of incomplete sampling and/or recent extinction. However, this explanation can hardly be accepted for the large intervals between networks 1.1 and 1.3 and between networks 1.2 and 1.3 (8 and 12 CA units, respectively). On the whole, the data currently available favor the view that, in order to explain the entirety of dinucleotide variation, a pure stepwise-mutation model needs integration with rare events involving changes of larger numbers of CA units, as already postulated has been for di- and trinucleotides (Di Rienzo et al. 1994; Deka et al. 1995; Watkins et al. 1995). Some of these events can be used as landmarks in the process of accumulation of variation within a frame. A conclusion of the present work is that at least two of the rare events consisting of large changes in repeat number can complement the information of single-site or insertional mutations in the definition of subsets of chromosomes with a common origin. For example, the two relevant groups of chromosomes of networks 1.2 and 1.3 could be identified within frame 1 only by virtue of microsatellite data.

The grouping of haplotypes of the same network is validated by our analyses and fits an evolutionarily relevant criterion. The greatest part of the total haplotype diversity proves to be condensed in no fewer than four common networks (1.1, 1.2, 2.1, and 3.1), each of which is represented in ⩾10% of the total sample. These networks can be considered optimal for population studies, not least because reliable frequency estimates can be obtained in small samples (e.g., 20–30 males).

Geographic Distributions

We have constructed maps describing the extant frequencies of networks over the sampled area. These are necessarily preliminary, because they are influenced strongly by the small size of some samples and also because large areas lack sampling. We deliberately used a fitting method that preserves the experimental result at each point, in order not to hide the effect of data suffering from large sampling errors. The maps represent useful models for the distribution of Y-chromosomal types and allow immediate comparisons with previous autosomal, Y-chromosomal, and mtDNA data presented in this form (Cavalli-Sforza et al. 1994; Cavalli-Sforza and Minch 1997). Emphasis should be put on the caution with which the maps must be interpreted. A specific frequency pattern is the result of both the migration and admixture of people, possibly associated with demic expansions, and of local expansions of types, because of drift and/or founder effects. The aforementioned factors could, in principle, be discriminated against when a large collection of autosomal data is used (Cavalli-Sforza et al. 1994), whereas such discrimination is not always possible in the case of nonrecombining Y-linked markers. Indeed, an enhanced effect of drift has been postulated and demonstrated for this chromosome worldwide (Torroni et al. 1990; Spurdle et al. 1994; Jobling and Tyler-Smith 1995; Scozzari et al. 1997; Underhill et al. 1997). The confinement and high frequency of network 2.1 haplotype 21-19-23-19 in Morocco (see above) suggests a strong drift that is able to affect markedly the shaping of the corresponding map (fig. 1*e*).

Many features replicate previous results—for example, the peculiarity of the Basque population (Cavalli-Sforza et al. 1994; Santachiara-Benerecetti et al. 1994; Lucotte and Hazout 1996), the sharp genetic changes at the African boundaries (Cavalli-Sforza et al. 1994; Ruiz-Linares et al. 1996; Hammer et al. 1997), and some of the genetic changes associated with linguistic and geographic boundaries (Barbujani and Sokal 1990, 1991). Novel features are also emerging, which seem more peculiar to the Y chromosome. These features include the high incidence of network 1.1 haplotypes (fig. 1*b*) in northeastern Europe. Chromosomes carrying the Tat-C mutation (Zerjal et al. 1997) are included within this set (A. Novelletto, unpublished results) and may label affinities with central Asians. Network 1.2 haplotypes reach high frequencies in south-central and southeastern Europe (fig. 1*c*). Since samples from the Middle East are missing, a better definition of the area with high frequencies of network 1.2 cannot be attained at present. Also emerging from our data is the accumulation of network 3.1 types in western Europe, with a focus in the entire Iberian peninsula, including the Basques (fig. 1*f*). This pattern is contributed mainly by a single haplotype of this network—that is, 23-19-23-23. This haplotype largely overlaps with chromosomes carrying haplotype XV, detected by probe 49a/49f (Persichetti et al. 1992; R. Scozzari, unpublished results). Semino et al. (1996) have described the high incidence of this latter haplotype in the same geographic area.

PC analysis (fig. 3) reveals that the distribution of network 3.1 types and chromosomes carrying the YAP insertion are the main determinants of the overall Y-chromosomal picture in Europe. Since this analysis took into account sample size, the lines of equal value of the first PC display less pronounced local variations, but they still retain most of the features discussed above. The map shows a clear southeast-to-northwest gradient all over Europe, a main feature of maps obtained with autosomal, Y-chromosomal, and mtDNA data (Cavalli-Sforza and Minch 1997, fig. 1*b–d*). The poor coverage of areas east of the Mediterranean gives less support to such a gradient over western Asia, in our map. A strong African influence also is evident, but it cannot be compared with that in previous data (Cavalli-Sforza and Minch 1997, fig. 1*c*).

Implications for the Peopling of Europe

Important hints on the processes that have led to the observed distributions may be obtained by a dating of the network antiquity. In two instances our dating results can be evaluated against available data. The estimate for network 2.1 is in agreement with the data reported by Hammer (1995) and Hammer et al. (1997). They showed that chromosomes carrying the YAP insertion, found in Europe and North Africa, belong primarily to YAP haplotype 4, whose origin is estimated as 90,000 years ago. Also, for network 1.3, confined to Sardinia, our estimate is compatible with an origin not earlier than the first human settlements of the island (9,000 years ago [Cappello et al. 1996]).

Network 1.1 appears to be the oldest network, and thus is a good candidate to represent the remainder of an ancestral set of background haplotypes on which subsequent variation was generated by aboriginal mutations or immigration. Barbujani et al. (1998) and Richards and Sykes (1998) warned against use of the age of molecules to infer the dating of splitting of the populations that carry them. In our data, too, the events that have led to the attainment of the observed frequencies of networks with derived characteristics may have occurred much more recently than the origin of the different types. Hammer et al. (1997) pointed out that the observed clines of YAP+ chromosomes (network 2.1) over Europe are compatible with many population movements out of Africa, which occurred 40,000–10,000 years ago. As for network 3.1, its preneolithic origin is supported. The geographic distribution of the major haplotype of this network parallels the late-paleolithic expansion from Iberia northward, recently demonstrated on the basis of the mtDNA data reported by Torroni et al. (1998). Our data also reveal the contribution of recent lineages (networks 1.2, 1.3 and 1.4) that emerged from an ancient background. In particular, network 1.2 haplotypes might represent a novel characteristic of chromosomes involved in the neolithic gene flow into mainland Europe from the southeast.

Controversy exists over the relevance, in the present European autosomal and mtDNA gene pool, of gene genealogies that coalesce in the neolithic period versus the preneolithic period (Richards et al. 1996, 1997; Cavalli-Sforza and Minch 1997). The three largest networks here reported appear to originate during the preneolithic period. On the other hand, our PC analysis is compatible with the demic diffusion associated with the demographic changes promoted by the spread of agriculture (Cavalli-Sforza et al. 1994), a neolithic process. In this context, the fact that our three older networks account for 76% of extant Y chromosomes may well hide the fact that a subset of haplotypes of these networks experienced numerical expansion due to demographic changes. In fact, it is likely that migrants' and preexisting populations' gene pools were not completely differentiated. Such an event leaves space for a wide range of values for the proportion of chromosomes that reached the present frequency by virtue of neolithic (or more recent) phenomena. The main conclusions of the present study can be summarized as follows: (1) there is a low level of homoplasy among dinucleotide microsatellite haplotypes; (2) there is high structuring of populations, with regard to Y-chromosomal network frequencies; and (3) networks are optimal markers for population studies addressing the radiation and dispersal processes associated with the preneolithic/neolithic transition.

Acknowledgments

We thank Mrs. M. Radikene, Kraujas (Vilnius), for help in collection of samples from Lithuania. This work was supported by Consiglio Nazionale delle Ricerche grants 96.01182.PF36 (to R.S.), 96.01192.PF36 (to L.T.), and 96.03130.CT4 (to A.N.) and by funds from the Ministry of the University (MURST 40%) (to R.S. and L.T.).

Appendix

Table A1

Absolute Frequency of Haplotypes

	No. of Subjects											
	Classified into Network								Unclassified			
	Frame 1			Frame 2	Frame 3	Frame 1		Frame 2				
Sample	1.1	1.2	1.3	2.1	3.1	1.4	1.5	2.2	Frame 1	Frame 2	Frame 3	Frame 4
Norwegian	2				6							
Lithuanian	6			1	6				1			
Danish	8	3		1	21				1		1	
Londoner	5				14							
Northern Portuguese	4			1	18						2	
Southern Portuguese	7	1		2	12				3		1	
Central Spaniard	3	3	2	1	9				1		1	
Basque	1	1			22							
Southern Spaniard	9			2	34				1	1	1	
Ligurian	3	1	1	5	10							
Venetian	3	4		2	12							
Latium	24	6	2	8	22	1					3	
Apulian	4	4		4	8							
Calabrian	6	7		3	10							1
Lucanian	4	3	1	6	7	2					1	
Sicilian	2	1		6	10	1		1				
Northern Sardinian	40	8	67	19	31	1			2	2	1	
Southern Sardinian	5	4	12	1	7							
Slovakian	10			2	10						2	
Northern Rumanian	7	7		1	9						1	
Eastern Rumanian	7	1		2	7				2			
Albanian	2			3	1						1	
Central Greek	7	2		5	7							
Cretan	3	8			2	1					2	
Turkish	6	2		1	4	1			1			
Omani	4	1			2		1		2		1	
United Arab Emirate	12			2	6				1			
Iranian	1	1			2		1	1				
Pathan	11				8							
Sindhi	6	1			11							
Moroccan Arab	12			32								
Northern Egyptian	3	3		14	2	1					1	
Southern Egyptian	6	1		4	3	1				3	2	

References

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Barbujani G, Bertorelle G, Chikhi L (1998) Evidence for paleolithic and neolithic gene flow in Europe. Am J Hum Genet 62:488–491

Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc Natl Acad Sci USA 87:1816–1819

——— (1991) Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. Am J Hum Genet 48: 398–411

Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martinez-Marignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am J Phys Anthropol 102:79–89

Caglià A, Novelletto A, Dobosz M, Malaspina P, Ciminelli BM, Pascali V (1997) Y-chromosome STR loci in Sardinia and continental Italy reveal islander-specific haplotypes. Eur J Hum Genet 5:288–292

Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, Piazza A (1996) Genetic analysis of Sardinia. I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125–141

Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, et al (1985) A human Y-linked

DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230:1403–1406

Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and geography of human genes. Princeton University Press, Princeton

Cavalli-Sforza LL, Minch E (1997) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 61:247–251

Ciminelli BM, Pompei F, Malaspina P, Hammer M, Persichetti F, Pignatti PF, Palena A, et al (1995) Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. J Mol Evol 41:966–973

Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759–1766

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, et al (1995) Population genetics of dinucleotide $(dC\text{-}dA)_n$ $(dG\text{-}dT)_n$ polymorphisms in world populations. Am J Hum Genet 56:461–474

Delfiner P (1976) Linear estimation of non-stationary spatial phenomena. In: M Guarasio, M David, C Haijbegts (eds) Advanced geostatistics in the mining industry. Reidel, Dordrecht, the Netherlands, pp 49–68

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Goldstein DB, Zhivotovsky LA, Nayar K, Ruiz-Linares A, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol Biol Evol 13:1213–1218

Gyapay G, Morissette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, et al (1994) The 1993–94 Généthon human genetic linkage map. Nat Genet 7:246–249

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

——— (1995) A recent common ancestry for human Y chromosomes. Nature 378:376–378

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, et al (1997) The geographic distribution of human Y chromosome variation. Genetics 145: 787–805

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11: 449–456

Kayser M, de Knijff P, Dialtjes P, Krawczak M, Nagy M, Zerjal T, Pandya A, et al (1997) Applications of microsatellite-based Y chromosome haplotyping. Electrophoresis 18: 1602–1607

Lucotte G, Hazout S (1996) Y-chromosome DNA haplotypes in Basques. J Mol Evol 42:472–475

Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellitto D, et al (1997) Characterization of a small family (CAIII) of microsatellite-containing sequences with X-Y homology. J Mol Evol 44:652–659

Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115–123

Persichetti F, Blasi P, Hammer M, Malaspina P, Jodice C, Terrenato L, Novelletto A (1992) Disequilibrium of multiple DNA markers on the human Y chromosome. Ann Hum Genet 56:303–310

Richards M, Côrte-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 M, Macaulay V, Sykes B, Pettitt P, Hedges R, Forstar P, Bandelt H-J (1997) Reply to Cavalli-Sforza and Minch. Am J Hum Genet 61:251–254

Richards M, Sykes B (1998) Reply to Barbujani et al. Am J Hum Genet 62:491–492

Roewer L, Arnemann J, Spurr NK, Grzeschik K-H, Epplen JT (1992) Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. Hum Genet 89:389–394

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y chromosome specific microsatellites in two closely related human populations. Hum Mol Genet 5: 1029–1033

Ruiz-Linares A, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, et al (1996) Geographic clustering of human Y-chromosome haplotypes. Ann Hum Genet 60:401–408

Santachiara-Benerecetti AS, Semino O, Passarino G, Bertranpetit J, Fellous M (1994) The genetic peculiarity of the Basques shown by some Y-specific polymorphisms. Am J Hum Genet Suppl 55:A164

Santos FR, Bianchi NO, Pena SDJ (1996) Worldwide distribution of human Y-chromosome haplotypes. Genome Res 6:601–611

Santos FR, Pena SDJ, Tyler-Smith C (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. Gene 165: 191–198

Scozzari R, Cruciani F, Malaspina P, Santolamazza P, Ciminelli BM, Torroni A, Modiano D, et al (1997) Differential structuring of human populations for homologous X and Y microsatellite loci. Am J Hum Genet 61:719–733

Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3:2159–2161

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

Spurdle AB, Hammer MF, Jenkins T (1994) The Y *Alu* polymorphism in southern African populations and its relationship to other Y-specific polymorphisms. Am J Hum Genet 54:319–330

Torroni A, Bandelt H-J, 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, Semino O, Scozzari R, Sirugo G, Spedini G, Abbas N, Fellous M, et al (1990) Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. Ann Hum Genet 54:287–296

Tyler-Smith CT, Brown WRA (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. J Mol Biol 195:457–470

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, et al (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. Proc Natl Acad Sci USA 93:196–200

Watkins WS, Bamshad M, Jorde LB (1995) Population genetics of trinucleotide repeat polymorphisms. Hum Mol Genet 4: 1485–1491

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378:379–380

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, et al (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183