ELSEVIER

Human Y-Chromosome Variation in the Western Mediterranean Area: Implications for the Peopling of the Region

Rosaria Scozzari, Fulvio Cruciani, Alessandra Pangrazio, Piero Santolamazza, Giuseppe Vona, Pedro Moral, Veronica Latini, Laurent Varesi, Marc M. Memmi, Valentino Romano, Giacomo De Leo, Massimo Gennarelli, Jadwiga Jaruzelska, Richard Villems, Jüri Parik, Vincent Macaulay, and Antonio Torroni

ABSTRACT: Y-chromosome variation was analyzed in a sample of 1127 males from the Western Mediterranean area by surveying 16 biallelic and 4 multiallelic sites. Some populations from Northeastern Europe and the Middle East were also studied for comparison. All Y-chromosome haplotypes were included in a parsimonious genealogic tree consisting of 17 haplogroups, several of which displayed distinct geographic specificities. One of the haplogroups, HG9.2, has some features that are compatible with a spread into Europe from the Near East during the Neolithic period. However, the current distribution of this haplogroup would suggest that the Neolithic gene pool had a major impact in the eastern and central part of the Mediterranean basin, but very limited consequences in Iberia and Northwestern Europe. Two other haplogroups, HG25.2 and HG2.2, were found to have much more restricted geographic distributions. The first most likely originated in the Berbers within the last few thousand years, and allows the detection of gene flow to Iberia and Southern Europe. The latter haplogroup is common only in Sardinia, which confirms the genetic peculiarity and isolation of the Sardinians. Overall, this study demonstrates that the dissection of Y-chromosome variation into haplogroups with a more restricted geographic distribution can reveal important differences even between populations that live at short distances, and provides new clues to their past interactions. *Human Immunology 62, 871–884 (2001).* © American Society for Histocompatibility and Immunogenetics, 2001. Published by Elsevier Science Inc.

KEYWORDS: Y-chromosome polymorphisms; European populations; West Mediterranean basin; Y-chromosome haplogroups

INTRODUCTION

Analyses of Y-chromosome sequence variation have provided major insights in the analysis of human origins, evolution, and dispersals [1–7]. The nonrecombining region (NRPY) of the Y chromosome is uniparentally transmitted and escapes recombination. Thus, its variation arises only by the sequential accumulation of new mutations along radiating paternal lineages. The sequence differentiation of human Y chromosomes has

From the Department of Genetics and Molecular Biology (R.S., F.C., A.P., P.S., A.T.), University of Rome "La Sapienza," Rome, Italy; Department of Experimental Biology (G.V., V.L.), University of Cagliari, Cagliari, Italy; Departament de Biologia Animal (P.M.), Universitat de Barcelona, Barcelona, Spain; Faculté des Sciences et Techniques (L.V., M.M.M.), Université de Corse, Corte, France; Department of Biopathology and Biomedical Methodology (V.R., G.D.L.), University of Palermo, Palermo, Italy; IRCCS (M.G.), Fatebenefratelli, Brescia, Italy; Polish Academy of Sciences (J.J.), Poznan, Poland; Department of Evolutionary Biology (R.V., J.P.), Tartu University and Estonian Biocentre, Tartu, Estonia; Department of Statistics (V.M.), University of Oxford, Oxford, United Kingdom; and Department of Genetics and Microbiology (A.T.), University of Pavia, Pavia, Italy.

Address reprint requests to: Dr. R. Scozzari, Department of Genetics and Molecular Biology, Faculty of Sciences, University of Rome "La Sapienza," Piazza A. Moro 5, 00185 Rome, Italy; Tel: +39 06 49912826; Fax: +39 06 4456866; E-mail: rosaria.scozzari@uniroma1.it.

Received May 16, 2001; accepted June 15, 2001.

0198-8859/01/$–see front matter
S0198-8859(01)00286-5

occurred during and after the process of colonization and diffusion into the different geographic regions and continents, and its dissection is a useful tool for the investigation of range expansions, migrations and other forms of gene flow in prehistoric and historic times. In other words, the sequence variation of modern Y chromosomes represents a unique record of the history of human males, and of the relationships between past populations.

A number of recent studies carried out in numerous populations have shown that most Y chromosomes can be classified into monophyletic units (haplogroups), which tend to be specific to each continent and major ethnic group [7–9]. This study takes advantage of this ethnic/geographic specificity to define better the origins and relationships of the populations living in the Western Mediterranean basin.

MATERIALS AND METHODS

Subjects

A sample of 1382 unrelated males from 36 regions of Europe, North Africa, and the Middle East was studied. Among these, 1127 were from the following locations in the Western Mediterranean area: 171 from Spain (three different locations), 73 from France, 141 from Corsica (four locations), 155 from continental Italy (six locations), 331 from Sardinia (ten locations corresponding to the linguistic domains reported by Cappello *et al.* [10]), 131 from Sicily (three locations), and 125 from Morocco (two population groups). Among the Spanish populations, a small sample of 19 subjects from an isolated population living in a restricted area (Pas valleys) of the community of Cantabria is of particular interest. The origin of this population is not clearly defined [11], although some historical information traces the peopling of the region back to the 11th century as a result of a repopulating from different sources, including Moorish slaves [12]. In addition, the following samples were analyzed for comparison: 171 from Northeastern Europe (35 from Denmark, 36 from Poland, 74 Estonian, and 26 Russians from Estonia), and 84 from the Middle East (27 Bedouin, 28 Druze, and 29 Palestinians from Israel) (Table 1). Among the 1382 Y chromosomes, 711 (see legend of Table 1) had their haplotypes partially typed by Malaspina *et al.* [13], and were further studied here by using 13 additional markers (SRY_{4064}, PN2, DYS271, M9, M13, PN3, p12f2, M12, SRY_{-2627}, LLY22g, M20, TAT, and XY275Y).

Typing

Y chromosomes were typed using 15 biallelic and four microsatellite polymorphisms of NRPY, and one biallelic polymorphism just distal to the Yp/Xp pseudoautosomal boundary (XY275Y). All 1382 individuals were tested for four dinucleotide repeat markers (YCAIIa, YCAIIb [14], DYS413a, and DYS413b [14, 15]), and for the indel binary polymorphism YAP [16]. For the remaining 15 markers, the following hierarchical typing scheme was used:

1. SRY_{4064} [17, 18], PN2 [1, 3], and DYS271 [19] were tested in all YAP(+) and some YAP(−);
2. SRY_{10831} [17, 18] and M9 [8] were tested in all YAP(−) and some YAP(+);
3. M13 [8] was tested in all samples that carried the ancestral allele at both SRY_{10831} and M9, and a few of the other samples;
4. PN3 [1, 3] was tested in all YAP(−)/SRY_{10831}(A)/M9(C) and many of the others;
5. p12f2 polymorphism [7] was analyzed in all YAP(−)/SRY_{10831}(G)/M9(C);
6. M12 [8] was tested in the samples that were p12f2(8kb) and some of the p12f2(10kb);
7. DYS257 [4] was tested in all M9(G) and some M9(C);
8. SRY_{-2627} [20] was tested in all DYS257(A) and some DYS257(G);
9. LLY22g (E. Righetti and C. Tyler-Smith, unpublished data) and M20 [8] were tested in all M9(G)/DYS257(G), and some DYS257(A);
10. TAT [21] was typed in all LLY22g(A);
11. XY275Y [22, 23] was tested in all YAP(+) and many of the others.

Statistical Analysis

Indices of Y chromosome diversity ($\hat{H}$) were obtained by using the Arlequin package, version 1.1 (University of Geneva, Geneva, Switzerland [24]). Y-chromosome diversity and its sampling variance were calculated as:

$$\hat{H} = \frac{n}{n-1}\left(1 - \sum_{i=1}^{k} p_i^2\right)$$

$$V(\hat{H}) = \frac{2}{n(n-1)}\left\{2(n-2)\left[\sum_{i=1}^{k} p_i^3 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right] + \sum_{i=1}^{k} p_i^2 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right\}$$

where p_i is the relative frequency of the i-th microsatellite haplotype and k is the number of haplotypes.

Arlequin was also used to calculate ϕ_{ST} values [25]. Molecular distances among biallelic haplogroups were computed by counting the number of mutations that separate two haplogroups on a most-parsimonious tree. Whether ϕ_{ST} was significantly different from zero was assessed by examining the distribution of ϕ_{ST} under

TABLE 1 Frequencies (%) of Y-chromosome haplogroups in the 36 regions examined

Population	Code	N	Haplogroup															
			1	2.1	2.2	3	7.1	7.2	8	9.1	9.2	9.3	16	21	25.1	25.2	22	26.1	26.2
Spanish		171																	
Southern Spaniard[a]	1	62	58.1	12.9	1.6	1.6	–	–	–	3.2	–	4.8	–	–	4.8	1.6	8.1	1.6	1.6
Asturias	2	90	58.9	4.4	6.7	4.4	–	–	–	3.3	5.6	1.1	–	–	11.1	2.2	1.1	1.1	–
Pasiego	3	19	31.6	5.3	–	21.1	–	–	–	–	–	–	–	–	–	42.1	–	–	–
French[b]	4	73	54.8	19.2	2.7	2.7	–	–	–	2.7	2.7	1.4	–	–	4.1	4.1	4.1	1.4	–
Corsican		141																	
Balagna	5	24	45.8	4.2	–	–	–	–	–	–	37.5	–	–	–	–	–	–	12.5	–
Corte[c]	6	62	66.1	12.9	1.6	–	–	–	–	1.6	–	3.2	–	–	6.5	–	–	8.1	–
Ajaccio[c]	7	28	42.9	32.1	3.6	–	–	–	–	–	7.1	7.1	–	–	3.6	–	–	3.6	–
Bonifacio	8	27	18.5	40.7	–	–	–	–	–	3.7	18.5	3.7	–	–	14.8	–	–	–	–
Italian		155																	
Ligurian	9	17	41.2	35.3	–	–	–	–	–	11.8	5.9	–	–	–	5.9	–	–	–	–
Lombard	10	18	61.1	11.1	–	–	–	–	–	–	11.1	–	–	–	5.6	5.6	–	5.6	–
Venetian	11	20	30.0	30.0	–	10.0	–	–	–	–	5.0	–	–	–	25.0	–	–	–	–
Latium[d]	12	66	33.3	24.2	1.5	3.0	–	–	–	9.1	12.1	4.5	–	–	9.1	1.5	–	1.5	–
Campania	13	15	20.0	26.7	–	–	–	–	–	6.7	40.0	6.7	–	–	–	–	–	–	–
Calabrian	14	19	21.1	26.3	5.3	10.5	–	–	–	–	10.5	5.3	–	–	21.1	–	–	–	–
Sardinian		331																	
Sassarese[e]	15	43	20.9	11.6	27.9	–	–	2.3	–	2.3	14.0	2.3	–	2.3	11.6	–	–	4.7	–
Gallurese	16	46	37.0	19.6	17.4	2.2	–	–	–	2.2	8.7	2.2	–	–	8.7	–	–	2.2	–
Logudoro South	17	21	14.3	28.6	33.3	–	–	–	–	4.8	14.3	–	–	–	4.8	–	–	–	–
Bitti[e]	18	37	8.1	10.8	27.0	–	–	–	–	16.2	–	–	–	8.1	29.7	–	–	–	–
Nuoro area[e]	19	26	26.9	23.1	42.3	3.8	–	–	–	–	3.8	–	–	–	–	–	–	–	–
Orosei and Siniscola[e]	20	36	22.2	8.3	50.0	2.8	–	–	–	8.3	5.6	–	–	–	–	–	2.8	–	–
Fonni and Barbagia of Ollolai[e]	21	36	25.0	25.0	36.1	–	–	–	–	8.3	2.8	–	–	–	2.8	–	–	–	–
Monte Ferru	22	18	16.7	11.1	55.6	–	–	–	–	–	–	–	–	5.6	11.1	–	–	–	–
Trexenta	23	47	10.6	25.5	36.2	2.1	–	–	–	2.1	6.4	6.4	–	2.1	4.3	2.1	–	2.1	–
Campidano of Cagliari	24	21	14.3	14.3	52.4	4.8	–	–	–	9.5	4.8	–	–	–	–	–	–	–	–
Sicilian		131																	
Trapani[f]	25	43	34.9	7.0	–	4.7	–	–	–	2.3	23.3	–	–	–	14.0	–	7.0	7.0	–
Sciacca	26	43	30.2	16.3	–	2.3	–	–	–	2.3	18.6	2.3	–	–	16.3	2.3	–	9.3	–
Troina[g]	27	45	13.3	15.6	–	4.4	–	–	–	4.4	28.9	4.4	–	–	28.9	–	–	–	–
Moroccan		125																	
Arab[a]	28	56	1.8	3.6	–	–	–	–	1.8	23.2	–	–	–	–	41.1	28.6	–	–	–
Berber	29	69	–	4.3	–	–	2.9	–	4.3	5.8	–	–	–	1.4	10.1	71.0	–	–	–
Danish[a]	30	35	57.1	22.9	–	5.7	–	–	–	–	8.6	–	2.9	–	2.9	–	–	–	–
Northeastern European		136																	
Polish[a]	31	36	19.4	27.8	–	41.7	–	–	–	–	5.6	–	2.8	–	2.8	–	–	–	–
Estonian[a]	32	74	1.4	20.3	–	36.5	–	–	–	1.4	1.4	–	32.4	–	5.4	–	–	–	1.4
Russian	33	26	–	7.7	–	26.9	–	–	–	–	3.8	3.8	50.0	–	–	–	7.7	–	–
Middle Eastern		84																	
Bedouin	34	27	–	7.4	–	11.1	–	–	–	66.7	–	–	–	3.7	11.1	–	–	–	–
Druze	35	28	28.6	10.7	–	3.6	–	–	–	3.6	14.3	–	–	–	14.3	–	–	–	25.0
Palestinian	36	29	–	62.1	–	–	–	–	10.3	3.4	3.4	3.4	–	–	13.8	–	–	3.4	–

[a] Sample corresponding to that previously reported [13].

[b] 26 of the 73 subjects were previously reported [13].

[c] A subset of the mixed sample of 90 Corsican previously reported [13].

[d] A subset of the 76 subjects from Latium previously reported [13].

[e] A subset of the mixed sample of 189 Northern Sardinian previously reported [13].

[f] A subset of the 65 Western Sicilian previously reported [13].

[g] A subset of the North-Eastern Sicilian previously reported [13].

10,000 permutations of individuals between populations.

Correspondence analysis was performed with the program Correspondence, version 1.0, included in the SPSS, version 8.0, package (SPSS Inc., Chicago, IL, USA). This method allows the examination of the relationships between two nominal variables in a multidimensional space and is the method of choice for crosstabulations where the cells contain frequency counts. This study used the matrix of the biallelic haplogroup counts with row principal normalization, because we were primarily interested in the differences and similarities among row categories (populations).

For each biallelic haplogroup, networks of "adjacent" (one repeat difference over the four loci) microsatellite haplotypes were constructed. Microsatellite haplotypes with the same combined repeat length were placed at the same horizontal level and the network was constructed by sequentially adding haplotypes differing by a single repeat unit [26, 27].

The genealogic depth of haplogroups was estimated under a stepwise microsatellite mutation model by using the average squared distance (ASD), a measure linearly related to coalescence time [28, 29]. This was obtained by calculating the squared difference in length (in CA units) for the alleles of each microsatellite between each individual's value and the value found in the ancestral haplotype, which was assumed to be the haplotype carrying the most frequent allele at each microsatellite. The average values for chromosomes belonging to the same haplogroup were then averaged over the four microsatellites and divided by the mutation rate. Values of 5.6×10^{-4} [30], and 25 were used for the mutation rate per microsatellite per generation and number of years per generation, respectively. As to HG2.2 (Figure 1), ASD was estimated by only using YCAIIa, DYS413a, and DYS413b. In order to calculate confidence intervals on ASD, the method of Thomas *et al.* [31] was followed. In brief, it was assumed that the paternal genealogy of the haplogroup in question was perfectly starlike, with a time depth as estimated from ASD. Mutations on this genealogy were then simulated using a Poisson process with rate equal to branch length multiplied by mutation rate, and choosing whether each mutation increased or decreased allele length by one step (each with probability 0.5). ASD was then evaluated for the simulated data, and the whole process repeated 1000 times. From the simulated values of ASD, an interval that covered the central 95% of values was quoted. It should be noted that uncertainties in the mutation rate, in the shape of the genealogy, and in the mutation process would increase the confidence intervals.

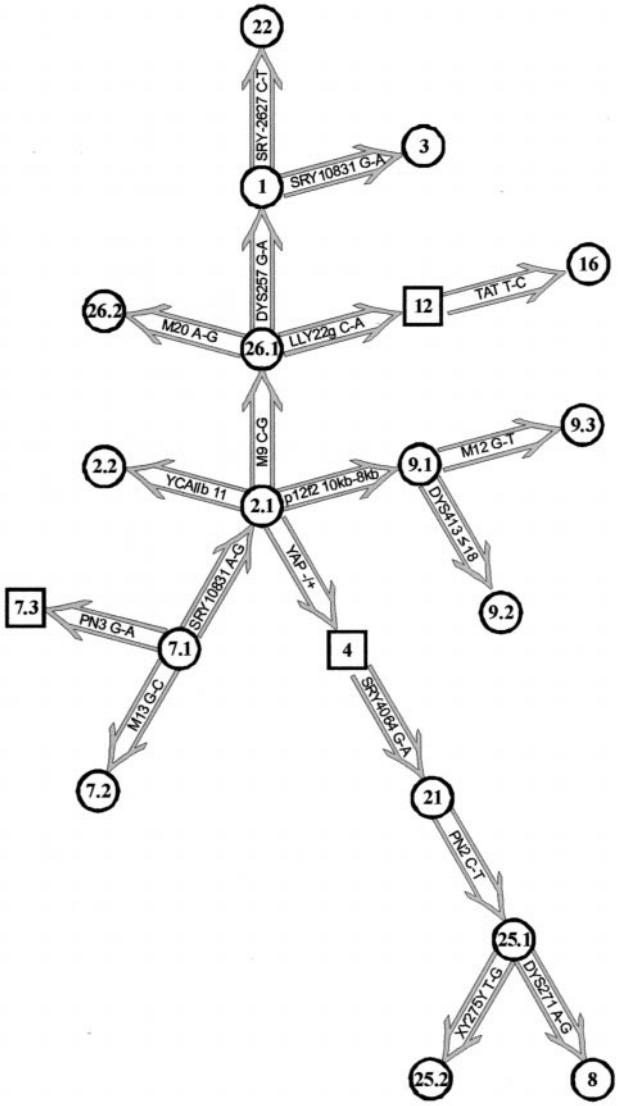

FIGURE 1 Maximum parsimony unrooted tree of Y-chromosomal HGs. Numbers within circles and squares are assigned HG names, and arrows between them represent the defining mutations. The 17 HGs observed in this study and encompassing our 1382 Y chromosomes are indicated by circles, the three HGs not found in this study are indicated by squares. Overall, the HG nomenclature follows that reported by Rosser *et al.* [7] and Bouzekri *et al.* [32] with some additional subtyping.

RESULTS

Defining Haplogroups

The 15 NRPY-specific biallelic polymorphisms were found to define major haplogroups (HGs) whose phylogenetic relationships had either been previously described [4, 7, 8, 32] or were refined in this study by using the hierarchical typing scheme (HGs 7.2, 7.3, 9.3, and 26.2 in Figure 1). For each haplogroup, a network of adjacent microsatellite haplotypes was constructed and

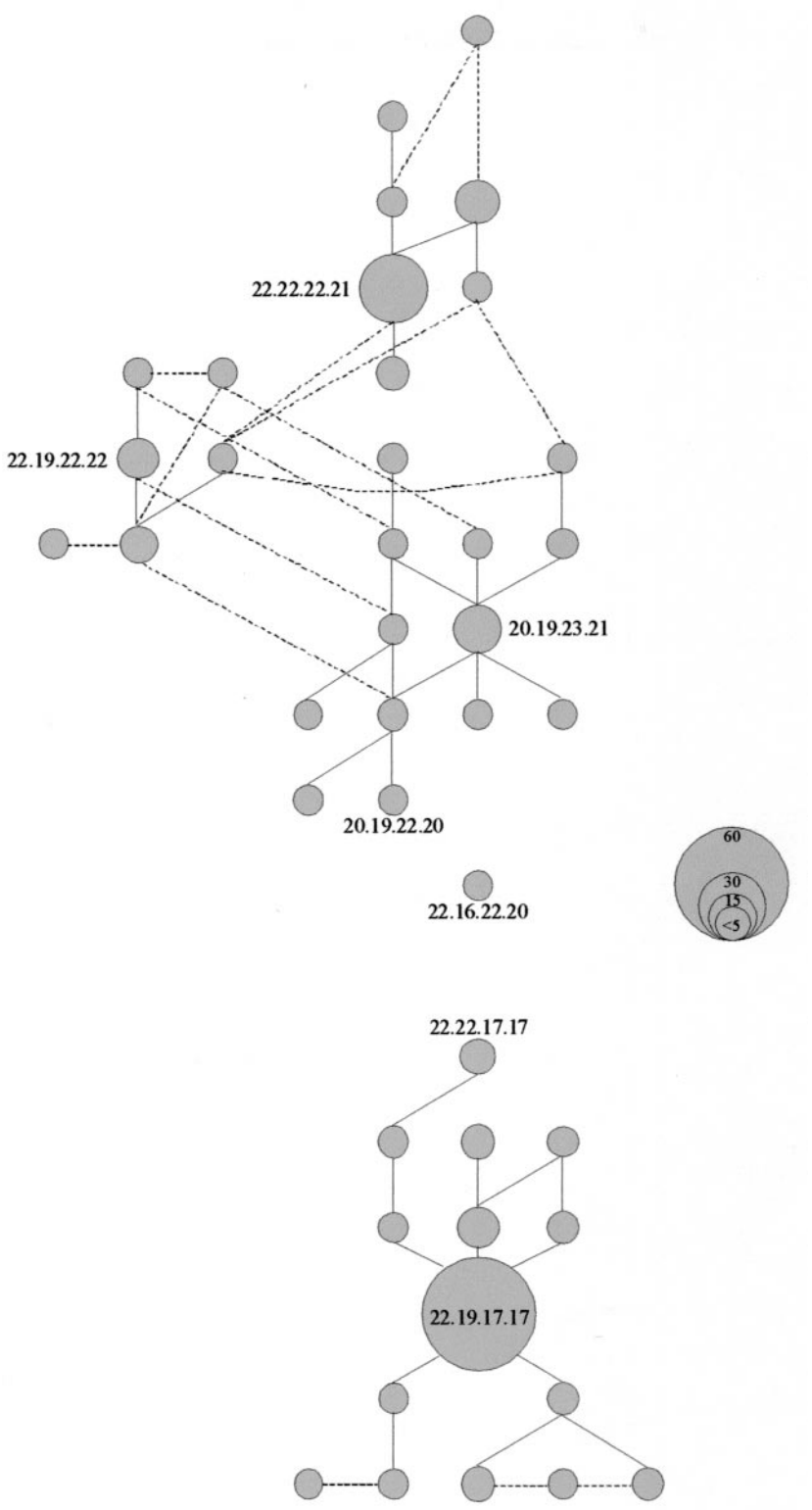

FIGURE 2 Network of adjacent microsatellite haplotypes within the haplogroup YAP(−)/p12f2(8kb) (HG9 in Rosser *et al.* [7]). All possible adjacent relationships are indicated by unbroken connecting lines, while dashed lines denote a two-repeat difference. Each circle area is proportional to the absolute frequency of the sampled haplotype. CA numbers are indicated in the order YCAIIa, YCAIIb, DYS413a, DYS413b. Allele size (number of CA repeats) is reported for selected haplotypes.

those of the haplogroups YAP(−)/p12f2(8kb) (HG9 in Rosser *et al.* [7]) and SRY_{10831}(G)/YAP(−)/p12f2(10kb)/M9(C) (HG2 in Rosser *et al.* [7]) are illustrated in Figures 2 and 3, respectively. Consistent with previous results [13, 27], a jump of more than two repeats at both DYS413a and DYS413b subdivides the

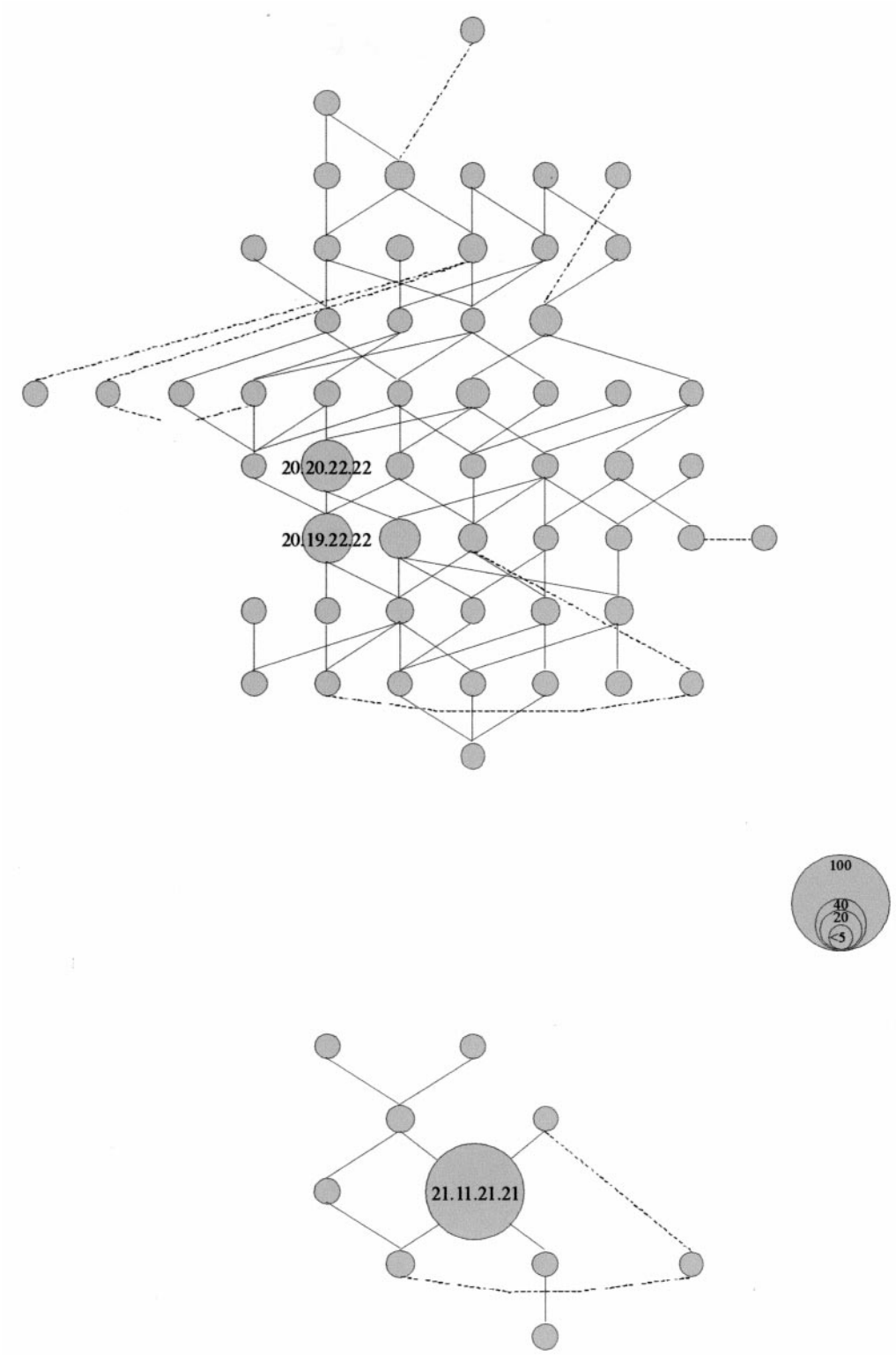

FIGURE 3 Network of adjacent microsatellite haplotypes within the haplogroup SRY_{10831}(G)/YAP(−)/p12f2(10kb)/M9(C) (HG2 in Rosser *et al.* [7]). Three haplotypes (not shown) could not be placed within either of the two subnetworks (for additional information see the legend of Figure 2).

YAP(−)/p12f2(8kb) haplogroup into two distinct subnetworks, one of which (HG9.2 in Figure 1, and lower part of Figure 2) is characterized by short-length alleles (18 repeats or less at both loci). These alleles were not

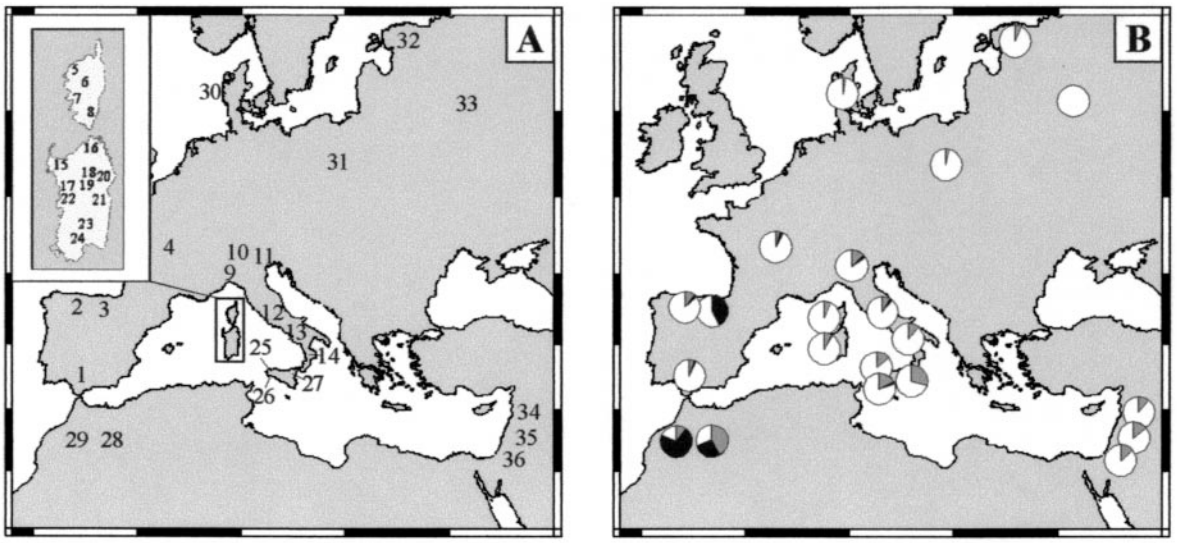

FIGURE 4 Maps showing the 36 sampled locations (panel A) and the frequency distribution of HGs 25.1 and 25.2 (grey and black sectors, respectively) (panel B). Code numbers for each sampled population are given in Table 1. The area of each sector is proportional to the frequency of the corresponding haplogroup.

seen in other haplogroups and were always associated with the ancestral G allele at the M12 site. This indicates that most likely a single deletion event involving multiple microsatellite units occurred on a p12f2(8kb)/M12(G) background and gave rise to HG9.2. Moreover, it suggests that indel events of multiple repeat units are rare enough to allow the dissection of haplogroups into phylogenetically meaningful subhaplogroups. This hypothesis was confirmed by the analysis of HG2. The lower part of Figure 3 depicts a subset of HG2 characterized by the allele with 11 repeats at YCAIIb, an allele already known to be at high frequency in Sardinia [13, 27, 33, 34]. In this study, 130 such alleles were observed and none of them was found in other haplogroups, indicating, also in this case, that most likely a single deletion event on a HG2 background gave rise to all members of HG2.2 (Figure 1).

Most of the Y chromosomes were also typed for the XY275Y T/G polymorphism [22, 23] that is located in the short-arm pseudoautosomal region, 275 bp distal from the boundary, and is shared by the X and Y chromosomes. Strong linkage disequilibrium between XY275 and the boundary has previously been observed [22, 23, 35], with the T allele occurring on both the X and Y chromosomes at varying frequencies, and the G allele occurring almost exclusively on the X chromosome. Because the association is not absolute, it is feasible that some chromosomes are generated by crossover between the X and the Y, as previously suggested [35]. In the present study, all XY275Y(G) chromosomes (a total of 83) appeared to be members of the haplogroup YAP(+)/PN2(T)/DYS271(A). This haplogroup includes only 16% of the 1382 sampled Y chromosomes, and only 25% of the 902 analyzed for DXY275Y (663 YAP[−] and all YAP[+]). These observations appear to indicate that, despite the possibility of recurrence, crossover events are rare, and the chromosomes with XY275Y(G) found in this survey are a monophyletic group (HG25.2 in Figure 1), which originated on a YAP(+)/PN2(T)/DYS271(A) background. If this is true, we would have expected the YAP(+)/PN2(T)/DYS271(A) chromosomes on the whole to have a higher microsatellite diversity than the XY275Y(G) chromosomes. Indeed, diversity values measured in terms of $\hat{H}$ (see the methods and materials section) were 0.846 ± 0.019 and 0.161 ± 0.055, respectively, supporting the hypothesis.

By using both biallelic and microsatellite information all individuals were assigned to 17 haplogroups, some of which are new (Figure 1). Regarding the nomenclature, this study followed that of Rosser *et al.* [7] and Bouzekri *et al.* [32] with some subtyping for the new haplogroups. HG9 is now subdivided into 9.1, 9.2, and 9.3 by the DYS413a,b and M12 mutations; HG2 is subdivided into HGs 2.1 and 2.2 by the YCAIIb variation; and HG25 is split into 25.1 and 25.2 by the XY275Y mutation. Finally, the analysis of M13 and PN3 (R. Scozzari, unpublished data) resulted in the HGs 7.1, 7.2, and 7.3, and that of M20 in the HGs 26.1 and 26.2 (the latter also termed HG28; C. Tyler-Smith, personal communication) (Figure 1).

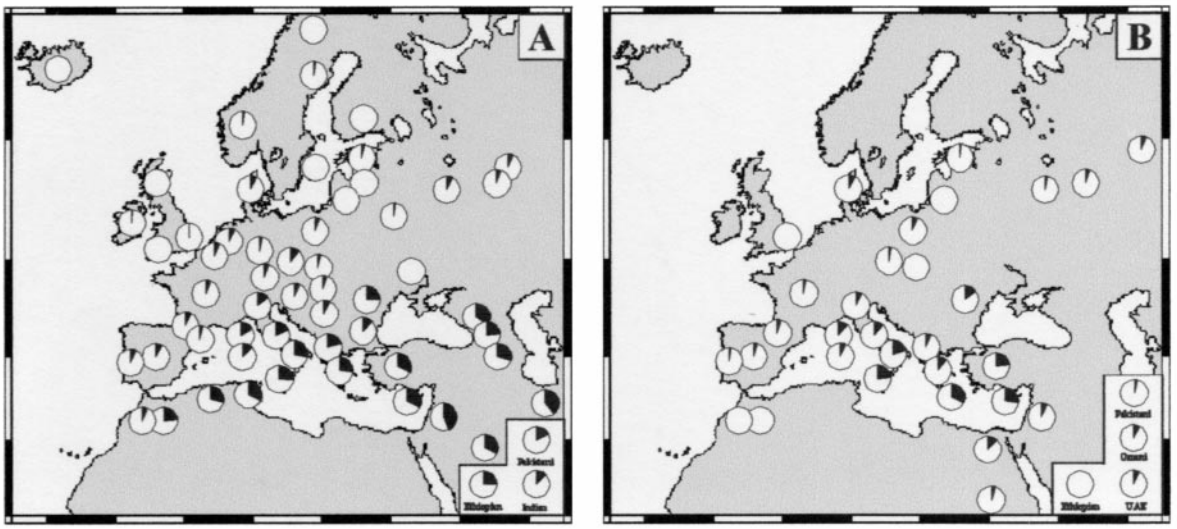

FIGURE 5 Maps showing the frequency distributions of HG9 (haplogroups 9.1 + 9.2 + 9.3) (panel A) and HG9.2 (panel B). Panel A also includes the p12f2 data from Semino *et al.* [36] and Hammer *et al.* [37], and the HG9 data from Rosser *et al.* [7] and Quintana-Murci *et al.* [38]. Panel B includes additional data from Malaspina *et al.* [13].

Geographical Distribution of Haplogroups

Haplogroup frequencies in the populations analyzed (Figure 4A) are illustrated in Table 1. The haplogroup most commonly observed in this study was HG1, which is known to be the most common in Western Europe [7], and has been recently found at appreciable frequencies also in some populations from sub-Saharan Africa [18]. We observed frequencies of HG1 higher than 50% in populations from Southern Spain, Asturias, France, and Denmark. In contrast, this haplogroup was rare in populations from Northeastern Europe, where HGs 3 and 16 reached the highest frequencies. The second most frequent haplogroup was HG2.1, found in all populations at varying frequencies, with no specific geographic distribution.

Haplogroup 25.1 was found at its highest frequency (41%) in the Arabs from Morocco, but it was also very common in the Sicilians, the Sardinians from Bitti, and in some groups from continental Italy. The Berbers from Morocco could be clearly differentiated from the Arabs of the same region by the frequency of HG25.2 (71% and 29%, respectively), a haplogroup distinguished from its ancestor (HG25.1) by the G allele at XY275Y. Interestingly, HG25.2 in Europe was found only in western populations, with a particularly high incidence (42%) in the Pasiego from the Pas valleys (Figure 4B).

Haplogroup 9 (which includes HGs 9.1, 9.2, and 9.3) encompassed more than 65% of the Bedouins from Israel, and revealed declining frequency westwards, a distribution consistent with previous observations [7, 36]. The frequencies of HG9 observed in the present study are illustrated in Figure 5A, together with those previously reported in other populations [7, 36–38]. The dissection of HG9 revealed that its derivative HG9.2 had frequencies higher than 10% in some populations of the West Mediterranean islands and Italy, but it reached its highest values in Sicily and Southern Italy (Table 1). This finding agrees well with previous results [13] and confirms that HG9.2 is a good indicator of the genetic links between Southern Italians/Sicilians and East Mediterranean populations (Figure 5B).

As previously reported [13, 27, 33, 34], haplogroup 2.2 is essentially confined to Sardinia, where it was found at high frequencies (17%–56%) in all population samples (Table 1). Very few instances of this haplogroup were observed in the surrounding regions, including nearby Corsica. Consistent with previous analyses, HG22 was observed in Spain and surrounding regions. However, the postulated recent Iberian origin of this haplogroup [39] appears to be in contrast with its finding in our Russian sample (8%). Finally, the sub-Saharan-specific haplogroups 7.1, 7.2, 21, and 8 [4, 18] were very uncommon and found only in six Berbers, one Arab, one Bedouin, three Palestinians, and seven Sardinians, which indicates that male-mediated gene flow across the Sahara has been very low both in historic and prehistoric times.

Correspondence Analysis

Figure 6 displays two correspondence analysis maps of Y-chromosome variation. Figure 6A includes all populations and illustrates that the first dimension (21.4% of the inertia) separates the populations of Northeastern Europe from the remaining populations because of the high frequencies of HG3 and 16 in the former. The

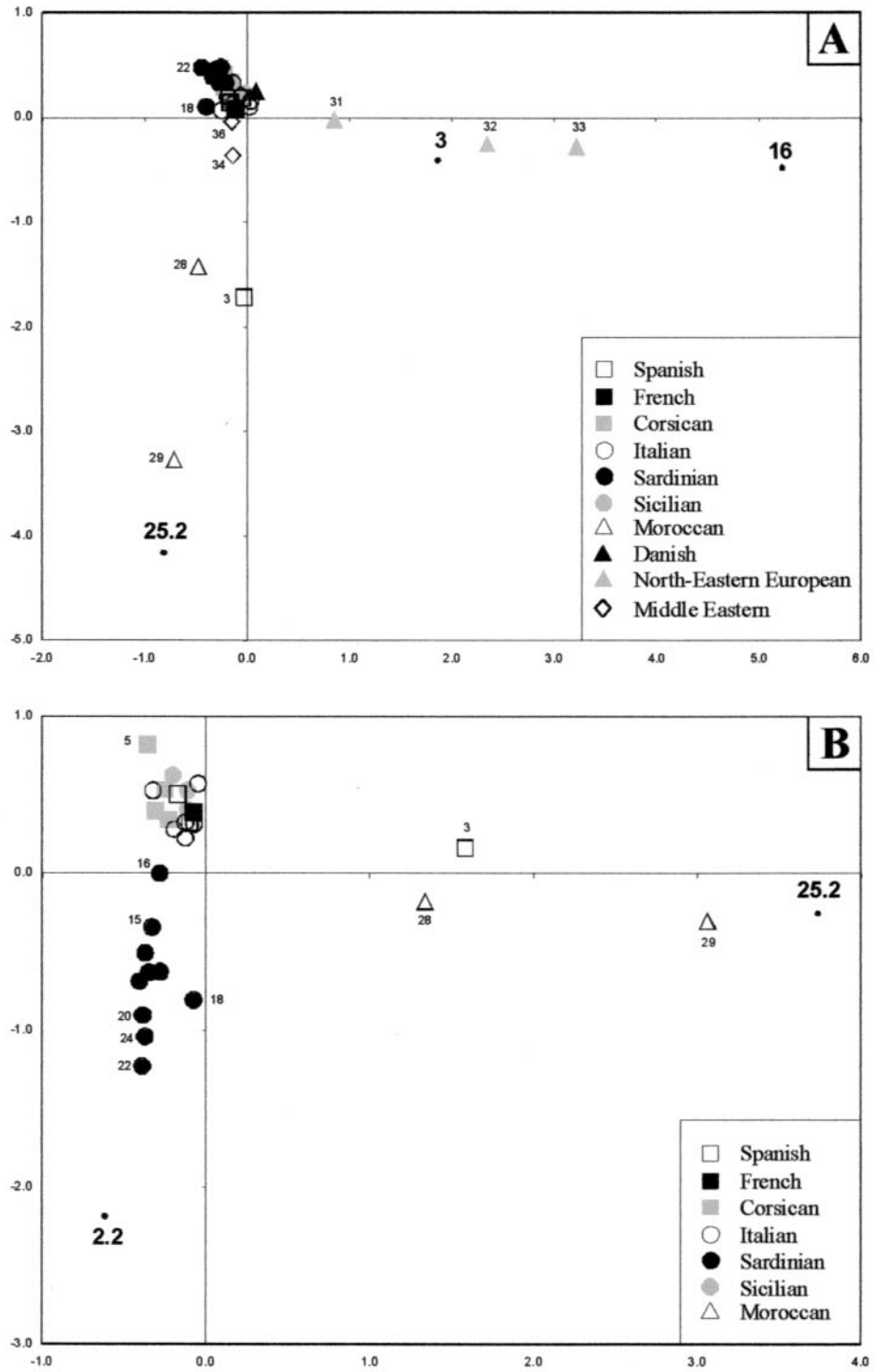

FIGURE 6 Plots of the correspondence analysis scores. (A) Plot of all populations examined. (B) Plot of Western Mediterranean populations. First dimension: X-axis; second dimension: Y-axis. Code numbers for populations are given in Table 1. Haplogroups are indicated by a dot. For the haplogroup nomenclature see text and legend for Figure 1.

second dimension (20.9% of the inertia) separates the two Moroccan groups from the rest because of their high frequency of HG25.2. The high frequency of this haplogroup also separates the Pasiego from the other Spanish, putting them close to the Moroccans.

To better define the relationships among Western Mediterranean populations, a second correspondence analysis was performed by excluding all populations not belonging to that geographic area (Figure 6B). The first dimension (36.1% of the inertia) again separates Moroccan and Pasiego from the other populations, whereas the second dimension (21.9% of the inertia) distinguishes the Sardinian populations from the rest. The Sardinian sample from Gallura occupies an intermediate position. Regarding the Corsican populations, their genetic heterogeneity ($\phi_{ST} = 0.138$; $p < 10^{-4}$) is clearly identified by the third dimension (data not shown), thus supporting previous analyses with classical genetic markers [40]. The high level of heterogeneity among the Corsican populations contrasts with the relatively low genetic diversity observed in the other two major islands of the West Mediterranean basin ($\phi_{ST} = 0.045$; $p < 10^{-4}$ for Sardinia, and $\phi_{ST} = 0.036$; $p = 0.02$ for Sicily).

DISCUSSION

The NRPY is characterized by a wealth of different polymorphic systems with different mutational mechanisms and rates. Biallelic markers with relatively low mutation rates allow the subdivision of a set of chromosomes into stable monophyletic clusters, or haplogroups. Multiallelic markers with higher mutation rates, such as microsatellites, can be used to define haplotypes within haplogroups, leading to a better-characterized diversity. In an initial attempt to use Y-chromosome variation to investigate the origin of European Y chromosomes, networks of microsatellites differing by one repeat were constructed for haplogroups defined by only two stable markers [13, 27]. That analysis provided new information about the mutational process of microsatellites, and suggested that the occurrence of rare multistep changes in the repeat number could be used to define more geographically coherent lineages within widely distributed haplogroups. In the current study, validation of the utility of microsatellites for that purpose has been made by linking their variation to haplogroups defined by a much larger number of biallelic markers. By using this approach this study demonstrates that two new Y-chromosomal haplogroups, termed HG9.2 and HG2.2, can be identified. These haplogroups are both defined by an abrupt change in the microsatellite repeat number. In contrast, a third haplogroup, HG25.2, was identified by using a marker from the pseudoautosomal region, lying very close to the boundary with the NRPY. Thus, in our sample of 1382 Y chromosomes we were able to detect a total of 17 haplogroups.

The geographic distribution of these haplogroups confirmed some previous observations, but also provided new data on the genetic relationships among Mediterranean populations. As in previous analyses [7, 36], our data showed an east-to-west distribution for HG9. It has been suggested that this pattern was generated by the expansions in Europe of farmers from the Near East during the Neolithic period [36]. Figure 5A illustrates that the spread of HG9 seems to have mainly affected the Eastern and Central Mediterranean coasts rather than Western, Central, and Northern Europe, a result that fits the distribution of Y-chromosome haplotypes defined by a different set of markers [41]. To refine the analysis of HG9, a network of the microsatellite haplotypes in our HG9 sample was constructed (Figure 2). This network is manifestly not starlike and can be differentiated into sublineages. This topology might reflect founder events and different expansion episodes not necessarily associated with the demic diffusion of agriculture. One well differentiated sublineage is HG9.2, identified by the ($\leqslant$18 repeats) allele at DYS413. This haplogroup reveals a decreasing-frequency cline in the same direction as that of HG9 (Figure 5B), but is more regionally localized. Indeed, our data suggest that if HG9.2 was involved in the expansion of Neolithic farmers, a very low proportion of their genes arrived in the westernmost part of the Mediterranean area. We estimated the time to the most recent common ancestor of HG9.2 chromosomes at ~14,000 YBP (95% confidence interval [CI]: 11,200–17,100 YBP). Because the CI of this estimate was derived under several assumptions, which might not be completely fulfilled, its width could be underestimated. Nevertheless, this range is congruent with previous estimates [13], and with the possibility that HG9.2 may have spread along the northern coast of the Mediterranean as far as Italy, possibly in association with the first Neolithic expansion, or as a result of a more recent westward pulse. A similar distribution has been observed for the fourth PC component in the analysis of classical genetic markers, and was associated with the Greek colonization process in the first millennium BC [42]. However, given the uncertainty in the demographic impact of the latter process [43], other explanations may need to be sought.

We also observed HG2.2, which is rare or absent in most of the populations examined, but it is extremely common in Sardinia. Within this island, HG2.2 is present at high frequencies in all locations, with a focus in a central region that archaeology has characterized as the "archaic" area because it harbors the earliest settlements [10]. Because of its frequency distribution across Europe, HG2.2 may be equivalent to the haplotype Eu8, whose origin was traced back to a population already present in Southwestern Europe during the Paleolithic period, presumably in one of the refugia occupied by humans during the last glaciation [41]. Although the finding of some HG2.2 chromosomes in some areas of

Spain, Italy, and France (Table 1) agrees with such a scenario, the coalescence age of HG2.2 (~5700 YBP, CI = 3900–7600 YBP), would rather favor a more recent origin in Sardinia. The correspondence analysis (Figure 6B) further confirms that Sardinia is clearly differentiated from all surrounding regions [42], including Corsica, and suggests almost no male-mediated gene flow from Sardinia to Corsica and continental Italy in historic and prehistoric times. As displayed in Figure 6B, the Gallurese sample, geographically located at the northern tip of Sardinia, occupies an intermediate position between the two islands, most likely a reflection of the geography and/or the language spoken by its inhabitants ([10] and citations therein). A similar result was obtained by mtDNA analysis and interpreted as being due to a recent peopling event from continental Italy involving both Corsica and Gallura [44].

The newly defined HG25.2 originated on a HG25.1 background. In Africa, HG25.2 is observed in 29% of Arabs and 71% of Berbers from Morocco, but is not found in those Ethiopian populations in which a high frequency of the ancestral HG25.1 is observed (R. Scozzari and associates, unpublished results [18]). Outside Northern Africa, HG25.2 was seen at generally low frequencies in Spain, France, and Italy, although no traces could be detected in the Near East. However, a particularly high frequency of this haplogroup (42%) was found in the Pasiego of the Pas valleys. In the correspondence analysis (Figure 6), the Pasiego do not cluster with the other Spanish populations, but rather with the Arabs and Berbers from Morocco, supporting historic and demographic records that would trace back the origin of this population to a heterogeneous resettlement, including also Moslem slaves [12]. The microsatellite diversity associated with HG25.2 provided coalescence age estimates of ~1400 YBP (CI = 540–2200 YBP). Although it is not possible at present to determine where HG25.2 originated, the simplest interpretation of our data is that HG25.2 diverged from the ancestor HG25.1 somewhere in North Africa a few thousand years ago. A founder effect led first to its expansion among the Berber populations, followed, in historical time, by its spread into the Iberian peninsula. Interestingly, the distribution of YAP(+)/DYS271(A) chromosomes was recently demonstrated to be strongly clinal in Portugal, with the highest frequencies in the south, and interpreted as a reflection of the Moorish invasions from North Africa in the Middle Ages [45]. A dissection of the Portuguese YAP(+)/DYS271(A) chromosomes by PN2 and XY275Y would determine whether they indeed belong to HG25.2, as could be inferred from an early report, which unfortunately did not provide haplotype information [46].

In conclusion, this study illustrates that the dissection of Y-chromosome variation into haplogroups with a more restricted geographic distribution allows the detection of important differences even between populations living short distances from each other, and can provide new clues to the processes by which the different areas of the Mediterranean basin were peopled. At the macrogeographic level, this study demonstrates that HG9.2 has a frequency pattern and an estimated coalescence age that are both compatible with a spread into Europe from the Near East during the Neolithic. However, if this scenario is correct, the data also indicate that the Neolithic wave had a major genetic impact in the eastern and central part of the Mediterranean basin, but only negligible consequences in Iberia and Northwestern Europe. This is in agreement with autosomal [47], mtDNA [48, 49], and other Y-chromosome data [7, 41]. At the microgeographic level two haplogroups, HG25.2 and HG2.2, appear to be very informative. Haplogroup 25.2 is very common only in Northern Africa, and most likely originated in the Berbers of Northwestern Africa within the last few thousand years. Its recent origin and high frequency in North Africa make it an excellent marker to detect recent gene flow to Iberia and Southern Europe. Similarly, HG2.2 is common only in Sardinia and could have arrived 9000 years ago ([10] and citations therein) from Western Europe at the time of the first human settlement of the island. Alternatively, it could have originated *in situ* after the peopling of the island. In any case, its rarity outside the island confirms the genetic peculiarity and the isolation of the Sardinians in the framework of European variation [42, 50, 51].

ACKNOWLEDGMENTS

We would like to express our gratitude to all blood donors for their helpful collaboration that made this study possible. We thank Chris Tyler-Smith and Mark A. Jobling for helpful comments. We gratefully acknowledge Jean-Paul Moisan and Damian Labuda for contributing French DNA samples, and Kenneth K. Kidd and Judith R. Kidd for providing the Danish DNA samples. We also thank the National Laboratory of Israeli Populations for the Bedouin, Druze, and Palestinian DNA samples. This work has been supported by funds from the Italian Ministry of the University, Progetti Ricerca Interesse Nazionale 1999, P.F. Beni Culturali CNR grants 97.00702.36 and 99.03852.36 (all to R.S.), by the CNR grant 99.02620.CT04, and F.A.R. of the University of Pavia (both to A.T.). V.M. is supported by a Research Career Development Fellowship from The Wellcome Trust.

REFERENCES

1. Hammer MF: A recent common ancestry for human Y chromosomes. Nature 378:376, 1995.
2. Jobling MA, Tyler-Smith C: Fathers and sons: the Y

chromosome and human evolution. Trends Genet 11:449, 1995.

3. Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL: The geographic distribution of human Y chromosome variation. Genetics 145:787, 1997.
4. Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL: Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15:427, 1998.
5. Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF: Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. Am J Hum Genet 64:817, 1999.
6. Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ: The central Siberian origin for native American Y chromosomes. Am J Hum Genet 64:619, 1999.
7. Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kučinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar A-V, Mitchell RJ, Nafa K, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA: Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526, 2000.
8. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996, 1997.
9. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ: Y chromosome sequence variation and the history of human populations. Nat Genet 26:358, 2000.
10. Cappello N, Rendine S, Griffo R, Mameli GE, Succa V, Vona G, Piazza A: Genetic analysis of Sardinia: I. Data on 12 polymorphisms in 21 linguistic domains. Ann Hum Genet 60:125, 1996.
11. Sánchez-Velasco P, Escribano de Diego J, Paz-Miguel JE, Ocejo-Vinyals G, Leyva-Cobián F: HLA-DR, DQ nucleotide sequence polymorphisms in the Pasiegos (Pas valleys, Northern Spain) and comparison of the allelic and haplotypic frequencies with those of other European populations. Tissue Antigens 53:65, 1999.
12. Esteban E, Dugoujon JM, Guitard E, Sénégas MT, Manzano C, de la Rúa C, Valveny N, Moral P: Genetic diversity in Northern Spain (Basque Country and Cantabria): GM and KM variation related to demographic histories. Eur J Hum Genet 6:315, 1998.
13. Malaspina P, Cruciani F, Santolamazza P, Torroni A, Pangrazio A, Akar N, Bakalli V, Brdicka R, Jaruzelska J, Kozlov A, Malyarchuk B, Mehdi SQ, Michalodimitrakis E, Varesi L, Memmi MM, Vona G, Villems R, Parik J, Romano V, Stefan M, Stenico M, Terrenato L, Novelletto A, Scozzari R: Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. Ann Hum Genet 64:395, 2000.
14. Mathias N, Bayés M, Tyler-Smith C: Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115, 1994.
15. Malaspina P, Ciminelli BM, Viggiano L, Jodice C, Cruciani F, Santolamazza P, Sellitto D, Scozzari R, Terrenato L, Rocchi M, Novelletto A: Characterization of a small family (CAIII) of microsatellite-containing sequences with X-Y homology. J Mol Evol 44:652, 1997.
16. Hammer MF, Horai S: Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951, 1995.
17. Whitfield LS, Sulston JE, Goodfellow PN: Sequence variation of the human Y chromosome. Nature 378:379, 1995.
18. Scozzari R, Cruciani F, Santolamazza P, Malaspina P, Torroni A, Sellitto D, Arredi B, Destro-Bisol G, De Stefano G, Rickards O, Martinez-Labarga C, Modiano D, Biondi G, Moral P, Olckers A, Wallace DC, Novelletto A: Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am J Hum Genet 65:829, 1999 (erratum appears in Am J Hum Genet 66:346, 2000).
19. Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL: Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3:2159, 1994.
20. Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K: Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype. Hum Genet 99:648, 1997.
21. Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C: Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174, 1997.

22. Ellis N, Kidd J, Goodfellow PJ, Kidd K, Goodfellow PN: Strong linkage disequilibrium between the XY274 polymorphism and the pseudoautosomal boundary. Am J Hum Genet 46:950, 1990 (erratum appears in Am J Hum Genet 49:908, 1991).
23. Ellis N, Taylor A, Bengtsson BO, Kidd J, Rogers J, Goodfellow P: Population structure of the human pseudoautosomal boundary. Nature 344:663, 1990.
24. Schneider S, Kueffer J-M, Roessli D, Excoffier L: Arlequin ver. 1.1: A software for population genetic data analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva, 1997.
25. Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479, 1992.
26. Cooper G, Amos W, Hoffman D, Rubinsztein DC: Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759, 1996.
27. Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, García O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Stefanescu G, Caraghin M, Tyler-Smith C, Scozzari R, Novelletto A: Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographic distribution. Am J Hum Genet 63:847, 1998.
28. Goldstein DB, Ruiz Linares A, Cavalli-Sforza LL, Feldman MW: An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463, 1995.
29. Slatkin M: A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457, 1995.
30. Weber JL, Wong C: Mutation of human short tandem repeats. Hum Mol Genet 2:1123, 1993.
31. Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB: Origins of Old Testament priests. Nature 394:138, 1998.
32. Bouzekri N, Taylor PG, Hammer MF, Jobling MA: Novel mutation processes in the evolution of a haploid minisatellite, MSY1: array homogenization without homogenization. Hum Mol Genet 7:655, 1998.
33. Caglià A, Novelletto A, Dobosz M, Malaspina P, Ciminelli BM, Pascali VL: Y-chromosome STR loci in Sardinia and Continental Italy reveal islander-specific haplotypes. Eur J Hum Genet 5:288, 1997.
34. Quintana-Murci L, Semino O, Poloni ES, Liu A, Van Gijn M, Passarino G, Brega A, Nasidze IS, Maccioni L, Cossu G, Al-Zahery N, Kidd JR, Kidd KK, Santachiara-Benerecetti AS: Y-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: a comparative study. Ann Hum Genet 63:153, 1999.
35. Spurdle A, Ramsay M, Jenkins T: The Y-associated XY275 low allele is not restricted to indigenous African peoples. Am J Hum Genet 50:1301, 1992.
36. Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS: A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am J Hum Genet 59:964, 1996.
37. Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonné-Tamir B: Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc Natl Acad Sci USA 97:6769, 2000.
38. Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, McElreavey K: Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. Am J Hum Genet 68: 537, 2001.
39. Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA: Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65:1437, 1999.
40. Memmi M, Moral P, Calò CM, Autuori L, Mameli GE, Succa V, Varesi L, Vona G: Genetic structure of southwestern Corsica (France). Am J Hum Biol 10:567, 1998.
41. Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiæ M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA: The genetic legacy of paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290:1155, 2000.
42. Cavalli-Sforza LL, Menozzi P, Piazza A: The History and Geography of Human Genes. Princeton NJ: Princeton University Press, 1994.
43. Renfrew C: At the edge of knowability: towards a prehistory of languages. Cambridge Archaeological J 10:7, 2000.
44. Morelli L, Grosso MG, Vona G, Varesi L, Torroni A, Francalacci P: Frequency distribution of mitochondrial DNA haplogroups in Corsica and Sardinia. Hum Biol 72:585, 2000.
45. Pereira L, Prata MJ, Brion M, Jobling MA, Carracedo A, Amorim A: Clinal variation of YAP+ Y chromosome frequencies in Western Iberia. Hum Biol 72:937, 2000.
46. Gonçalves J, Lavinha J: The Y-associated XY275G (low) allele is common among the Portuguese. Am J Hum Genet 55:583, 1994.
47. Arnaiz-Villena A, Martínez-Laso J, Alonso-García J: Iberia: population genetics, anthropology, and linguistics. Hum Biol 71:725, 1999.
48. Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonné-

Tamir B, Scozzari R: mtDNA analysis reveals a major late paleolithic population expansion from Southwestern to Northeastern Europe. Am J Hum Genet 62:1137, 1998.

49. Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, Bandelt H-J: Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251, 2000.
50. Modiano G, Terrenato L, Scozzari R, Santachiara-Benerecetti SA, Ulizzi L, Santolamazza C, Petrucci R, Santolamazza P: Population genetics in Sardinia (with a historical account on the birth of the Haldane "malaria hypothesis"). Atti Acc Lincei Mem Fis 18:257, 1986.
51. Vona G: The peopling of Sardinia (Italy): history and effects. Int J Anthropol 12:71, 1997.