ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution 28 (2003) 387–395

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects☆

F. Di Giacomo,[a] F. Luca,[b] N. Anagnou,[c,d] G. Ciavarella,[e] R.M. Corbo,[f] M. Cresta,[f] F. Cucci,[g] L. Di Stasi,[h] V. Agostiano,[i] M. Giparaki,[j] A. Loutradis,[j] C. Mammi',[k] E.N. Michalodimitrakis,[c] F. Papola,[l] G. Pedicini,[m] E. Plata,[j] L. Terrenato,[a] S. Tofanelli,[n] P. Malaspina,[a] and A. Novelletto[b,*]

[a] *Tor Vergata University, Rome, Italy*
[b] *Department of Cell Biology, University of Calabria, Via P. Bucci, 87030 Arcavacata di Rende, Italy*
[c] *University of Crete, Heraklion, Greece*
[d] *University of Athens, and IMBB, Greece*
[e] *IRCCS S. Giovanni Rotondo, Italy*
[f] *La Sapienza University, Rome, Italy*
[g] *Az. Osped. Perrino, Brindisi, Italy*
[h] *ASL 1, Cosenza, Italy*
[i] *Lab. Tipizzazione Tissutale, Matera, Italy*
[j] *Laikon Hospital, Athens, Greece*
[k] *Az. Osped. B.M.M., Reggio Calabria, Italy*
[l] *Lab. Tipizzazione Tissutale, CNR, L'Aquila, Italy*
[m] *Az. Osped. Rummo, Benevento, Italy*
[n] *University of Pisa, Pisa, Italy*

Received 10 July 2002; revised 1 November 2002

Abstract

We explored the spatial distribution of human Y chromosomal diversity on a microgeographic scale, by typing 30 population samples from closely spaced locations in Italy and Greece for 9 haplogroups and their internal microsatellite variation. We confirm a significant difference in the composition of the Y chromosomal gene pools of the two countries. However, within each country, heterogeneity is not organized along the lines of clinal variation deduced from studies on larger spatial scales. Microsatellite data indicate that local increases of haplogroup frequencies can be often explained by a limited number of founders. We conclude that local founder or drift effects are the main determinants in shaping the microgeographic Y chromosomal diversity.

1. Introduction

The non-recombining portion of the human Y chromosome (NRY) is widely used as a marker of inter-population divergence. Since the NRY does not undergo recombination, it formally behaves as a single locus with many alleles. These are called haplotypes or haplogroups (Hg) depending on the type of variation used to define them.

For single nucleotide polymorphism (SNP) variation, the phylogenetic relationships among Hg's can be reconstructed unequivocally, assuming a monophyletic origin of the derived state at each variant position. Experimental data have not contradicted this assumption and a comprehensive tree has been recently presented (YCC, 2002). A notable feature of this tree is the presence of 18 major clades defined by a limited number of

☆ Presented at the 2002 annual meeting of the Society for Molecular Biology and Evolution (SMBE), Sorrenty, Italy, June 13–16, 2002.

* Corresponding author. Fax: +39-0984-492911.
E-mail address: novelletto@bio.uniroma2.it (A. Novelletto).

1055-7903/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1055-7903(03)00016-2

mutations. Chromosomes belonging to some clades are confined to specific continents. Thus, geographic distribution of Hg's can be used as a powerful tool to shed light on population affinities and male-mediated gene flow and admixture.

An additional type of variation which has been used to address these questions is represented by microsatellite markers (STR), which can be arrayed into haplotypes. The most promising approach combines the analysis of these two types of polymorphisms to describe the variation of fast-evolving STRs within each SNP-defined lineage (de Knijff, 2000). STR allele size distributions in fact convey information on each lineage antiquity and the population demography underlying its dispersal (King et al., 2000 and references therein).

Several studies have shown that the NRY markers display the highest quotas of inter-population divergence ever reported for human populations (Hammer and Zegura, 2002; Romualdi et al., 2002). The causes of this are currently debated, and may not be necessarily the same when considering different geographic scales (i.e., continental vs. local; Hammer et al., 2001). In a survey of European data spanning over 1000 km, NRY markers displayed a higher rate of spatial divergence among populations than the one observed for mitochondrial and autosomal markers (Seielstad et al., 1998). These authors attribute their observation to a higher female migration rate. These findings suggest that NRY diversity may be used to reveal population structuring also on a micro-geographical scale, allowing the analysis of genetic consequences of peopling events, population spread, gene flow, and admixture for local areas (Lell and Wallace, 2001; Renfrew, 2001).

Previous studies have showed that, using a limited number of NRY markers and the appropriate sampling scheme, it is generally possible to recognize distinct geographical patterns across Europe (Malaspina et al., 1998, 2000; Rosser et al., 2000; Semino et al., 2000; Stefan et al., 2001).

In this paper we describe micro-geographic variation of nine NRY Hg's from thirty male groups sampled in Italy and Greece. The aim of this study is to test alternative expectations on micro-geographic Hg distributions, i.e., clinal distributions reflecting those reported throughout Europe, abrupt changes in Hg frequencies reflecting sharp genetic boundaries, absence of relevant geographic patterning.

2. Materials and methods

2.1. The subjects

We analyzed blood samples from a total of 890 male subjects with known parental and granparental origins: 524 Italians (locations 1–17 in Table 1 and Fig. 1), 154 individuals from continental Greece (locations 18–24), 212 subjects from Crete (locations 25–28), and the Aegean islands of Lesvos and Chios (locations 29–30).

Informed consent was obtained in all cases. Sampling was anonymous in order to prevent link to the original donor.

2.2. The markers

We used the SNPs at DYS257, SRY_{10831}, $DYS221_{136}$, M170, and M172, the YAP element insertion/deletion polymorphism, the rearrangement detected by probe p12f2 and the dinucleotide STRs YCAII and DYS413. These define the 9 Hg's: A, DE, G2, I-M170, J* (xJ2), J2* (xDYS413 $\leqslant$ 18), J2-(DYS413 $\leqslant$ 18), P* (xR1a), and R1a (YCC, 2002). Hg's J2-(DYS413 $\leqslant$ 18) and J2* (xDYS413 $\leqslant$ 18) are based on a multirepeat deletion at DYS413 (Malaspina et al., 2001). Chromosomes that cannot be assigned to any of the above Hg are classified as Y* (xA, DE, G2, I, J, P).

We used the following sequential typing scheme to determine Hg frequencies. YAP (Hammer and Horai, 1995) and DYS257 (Hammer et al., 1998) were typed in all subjects. SRY_{10831} (Kwok et al., 1996) was first typed on all DYS257(A). p12f2 (Rosser et al., 2000) was typed on all YAP(−)/DYS257(G). M172 (Malaspina et al., 2001; Underhill et al., 2001) was typed on all p12f2(−). In all YAP(−)/DYS257 (G)/p12f2(+), M170 (Underhill et al., 2001), and $DYS221_{136}$ (Hammer et al., 2001) were tested sequentially. All the remaining subjects were tested for SRY_{10831} to detect haplogroup A.

DYS413 was typed on all but 7 subjects: 4 DYS257(A) and 3 YAP(+). YCAII was not typed in locations PAO, CIL, ALT, BRI, and BEN.

For M170, ASO probes and washing temperature were 5′-CATTGTTCATTTTTTC-3′ and 5′-CATTGTTCCTTTTTTTC-3′ (43 °C), and for $DYS221_{136}$ 5′-TGAATCTTACGCCTGAA-3′ and 5′-TGAATCTTATGCCTGAA-3′ (50 °C). DYS413 and YCAII were typed as described (Malaspina et al., 1997; Mathias et al., 1994).

2.3. Data analysis

Analysis of molecular variance (AMOVA, Excoffier et al., 1992), was used to calculate three fixation indexes: F_{st}, representing the correlation of random Hg's within samples relative to that of random pairs of Hg's drawn from the entire study; F_{ct}, representing the correlation of random Hg's within a group of samples relative to that of random pairs of Hg's drawn from the entire study, and F_{sc}, representing the correlation of random Hg's within a sample relative to that of random pairs of Hg's drawn from the group of samples. The significance of the fixation indexes was tested using 5000 permutations to generate null distributions (Schneider et al., 1997).

Table 1
Haplogroup relative frequencies (%), sample size and diversity indexes

	Haplogroup											
	Pop. sample[a] (code)	P* (xR1a)	R1a	DE	G2	I-M170	J2- (DYS413 ⩽ 18)	J2*- (xDYS413 ⩽ 18)	J* (xJ2)	A	Y* (xA, DE, G2, I, J, P)	Sample size (*n*)	Diversity index ± SE
	Italy												
1	*Val di Non* (VAL)	**73.3**	6.7				3.3	6.7			10.0	30	
2	Verona[b] (VER)	45.5	9.1	9.1		4.5	22.7	4.5			4.5	22	
3	*Garfagnana* (GAF)	**76.2**	4.8	2.4	4.8		2.4	7.1			2.4	42	
4	Genoa[b] (GEN)	**48.3**		24.1	10.3	6.9	6.9		3.4			29	
5	L'Aquila[b] (LAQ)	25.7	5.7	11.4	5.7	8.6	25.7	5.7	2.9		8.6	35	
6	Pescara[b] (PES)	45.0		15.0			10.0	5.0	15.0		10.0	20	
7	Avezzano[b] (AVE)	41.4	6.9	3.4	6.9	10.3	13.8	3.4	3.4		10.3	29	
8	Benevento (BEN)	26.1	2.2	17.4	10.9	8.7	19.6		6.5		8.7	46	
9	*Cilento* (CIL)	29.2	2.1	12.5	14.6	6.3	16.7	4.2	6.3		8.3	48	
10	Foggia[b] (FOG)	11.1		11.1	14.8	**18.5**	**40.7**	3.7				27	
11	*North Gargano*[b] (GAR)	27.6	3.4	24.1		3.4	10.3	10.3	17.2		3.4	29	
12	Casarano[b] (CAS)	30.0	10.0	20.0	10.0	5.0	20.0	5.0				20	
13	Brindisi (BRI)	18.4	5.3	**26.3**	7.9	13.2	13.2	7.9	2.6		5.3	38	
14	Altamura (ALT)	40.0		**36.0**	4.0	4.0	4.0	8.0			4.0	25	
15	Matera[b] (MAT)	33.3		25.0	4.2	12.5	12.5				12.5	24	
16	Paola (PAO)	25.9	3.7	11.1		7.4	**29.6**	3.7	11.1		7.4	27	
17	Reggio Calabria[b] (REG)	24.2		**27.3**	3.0		21.2	6.1	9.1		9.1	33	
	Total	36.4	3.4	15.8	6.3	6.5	15.8	4.8	4.6		6.3	524	.801 ± .011
	Cont. Greece												
18	Agrinion[b] (AGR)	19.0	4.8	9.5	4.8	23.8	9.5	14.3	4.8		9.5	21	
19	Ioannina[b] (IOA)	16.7	8.3	29.2	4.2	8.3	4.2	12.5	4.2		12.5	24	
20	Patrai[b] (ACA)	11.1	5.6	**44.4**		11.1	16.7				11.1	18	
21	Kardhitsa[b] (KAR)	8.0	**20.0**	28.0	12.0	12.0	4.0	8.0	4.0		4.0	25	
22	Serrai[b] (SER)	12.0	8.0	24.0	4.0	**36.0**	16.0					25	
23	Thessaloniki[b] (TES)	5.0	**25.0**	20.0	5.0	20.0	5.0	10.0			10.0	20	
24	Larisa[b] (LAR)	19.0	9.5	14.3	4.8	14.3	23.8	4.8			9.5	21	
	Total	13.0	11.7	24.0	5.2	18.2	11.0	7.1	1.9		7.8	154	.858 ± .011
	Crete												
25	Rethimnon[b] (RET)	4.5	**22.7**	9.1		22.7	**36.4**		4.5			22	
26	Khania[b] (CHA)	3.4	3.4	24.1	3.4	10.3	**44.8**		6.9		3.4	29	
27	Iraklion[b] (HER)	7.1		26.2	9.5	14.3	**35.7**	2.4	2.4		2.4	42	
28	*Lasithi*[b] (LAS)	18.0	10.0	4.0	10.0	12.0	22.0	4.0	2.0		**18.0**	50	
	Total	9.8	7.7	15.4	7.0	14.0	32.9	2.1	3.5		7.7	143	.830 ± .018
	Aegean islands												
29	Mitilini[b] (MIT)	7.4	11.1	18.5	7.4	18.5	18.5	7.4	3.7	3.7[c]	3.7	27	
30	Khios (CHI)	**26.2**	9.5	23.8	9.5	2.4	11.9	4.8	2.4		9.5	42	
	Total (Greece)	12.8	9.8	20.2	6.6	14.8	20.2	4.9	2.7		7.7	366	.859 ± .006

[a] For locations not identified by major towns, the geographical name of the local region is reported in italics.
[b] Partially included in Malaspina et al. (2001).
[c] Pooled with unclassified chromosomes in χ^2 and correspondence analysis calculations.

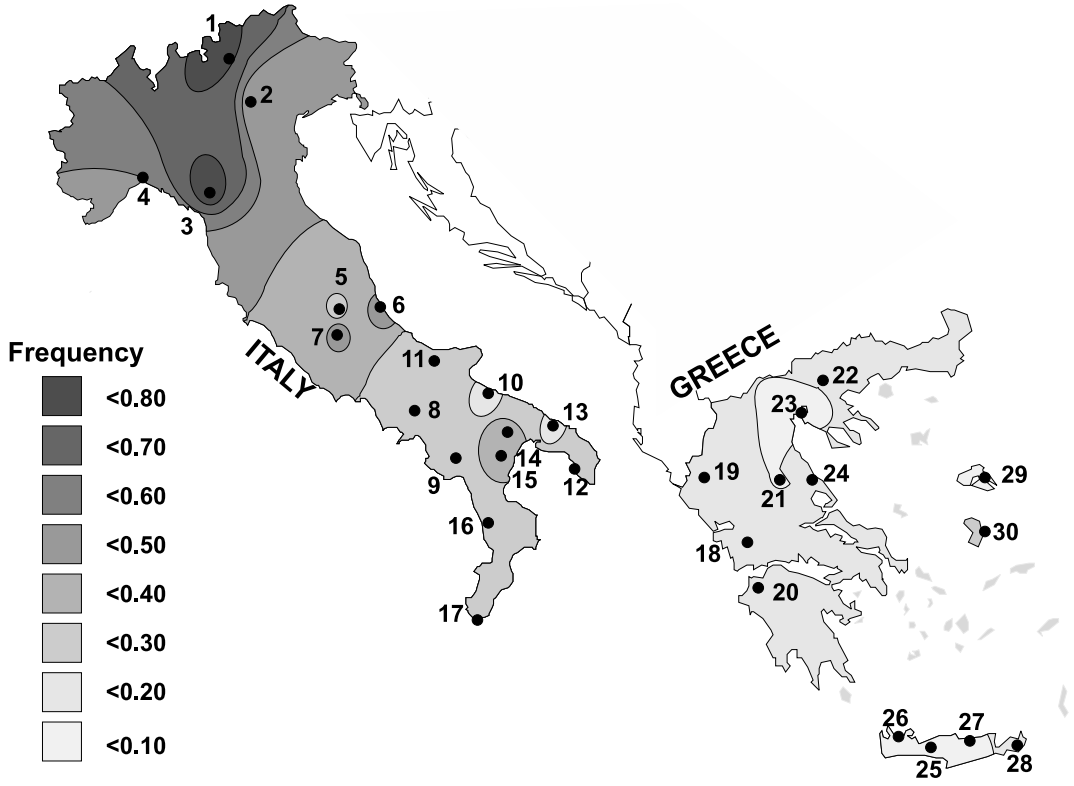

Fig. 1. Position of the 30 sampling locations, overimposed on the interpolated frequency surfaces for haplogroup P* (xR1a). Increasing intensity of the gray shading represents increasing frequency of Hg P* (xR1a) as shown (bottom left). Equal frequency lines were obtained as described in Section 2.

In order to assay the gradient of haplogroup differentiation, we carried out spatial autocorrelation analysis (Sokal and Oden, 1978), using the program AIDA (Bertorelle and Barbujani, 1995). We performed several runs using different numbers of distance classes to faithfully represent the geographical distances among samples, yet retaining a meaningful number of comparisons in each class. The option for obtaining distance classes of equal numerosity was used in all cases. Class-specific confidence intervals were obtained with 1000 simulations.

To evaluate the overall similarity among the thirty samples, we carried out correspondence analysis (program ANACOR, in SPSS v. 6.1, 1991), using Hg frequencies as columns and the 30 population samples as rows, with row principal normalization.

Multiple linear regression of Hg frequencies on latitude and longitude was performed according to standard methods (SPSS v. 6.1).

Frequency surfaces were drawn with the Surfer System v. 4.15 (Golden Software) using the Kriging procedure (Delfiner, 1976). The peculiarity of this method is that the estimated values of the variable coincide with the observed values at the sampled locations. This is therefore the best method to group samples with similar haplogroup frequencies into a definite number of belts.

3. Results and discussion

3.1. Hg frequency distribution in Italy and Greece

Within each country we found a highly significant heterogeneity of Hg frequencies (Table 1; $\chi^2 = 210$, 128 d.f., $p = .0001$ for Italy and $\chi^2 = 134$, 96 d.f., $p = .005$ for Greece).

The 9 NRY markers allow the identification of Hg's in the vast majority of subjects, since only 6.3% (Italy) and 7.7% (Greece) of the chromosomes were unassigned. Interestingly, in Greece more than one fourth of these chromosomes are found in a single location from Crete. Only a single Hg A individual was found, a finding which is in agreement with its rare occurrence outside Africa (Scozzari et al., 2001; Underhill et al., 2000).

The most common Hg among the Italian samples is P* (xR1a). Its overall frequency is more than twice that of the second one. However, its frequency ranges widely (11–76%). The second most common Hg's are DE and J2-(DYS413 $\leqslant$ 18), with frequencies varying between 0 and 36%, and 2 and 41%, respectively.

In Greece, the most common Hg's are DE and J2-(DYS413 $\leqslant$ 18), with an average frequency of 20.2% each, ranging from 4 to 44% and from 4 to 24% for DE and J2-(DYS413 $\leqslant$ 18), respectively.

Among the Italian samples, only three Hg's have frequencies >10%, accounting for more than two thirds of chromosomes. On the other hand, the three most common Hg's found among the Greek samples account for only 55% of chromosomes. Accordingly, Hg diversity is higher in Greece than in Italy (Table 1).

3.2. Patterns of genetic differentiation among populations

In order to have a synthetic view of gene pool similarities among population samples, we used correspondence analysis (Fig. 2a). The first two dimensions explain 40 and 17% of the total inertia, respectively, summarizing more than half of the total variation. Dimension 1 mainly reflects the frequency of Hg P* (xR1a) (61% of the dimension), whereas dimension 2 reflects mainly the frequency of Hg R1a (47%). Hg's I-M170 and J2-(DYS413 $\leqslant$ 18) contribute almost equally to both dimensions (16 and 15% to dimension 1 and 10 and 18% to dimension 2, respectively).

In the space defined by the first two dimensions, samples from the two countries show little overlap. Two Italian populations (GAF, VAL) are clearly separated from the rest on the first dimension. The Greek samples span the entire range of the second dimension. Within each country, there is no clear correspondence between the positioning of the samples in the first two dimensions and their geographical location. Two northern Italian samples (GEN, VER) map close to the southern samples. Three samples from central Italy (LAQ, PES, and AVE)

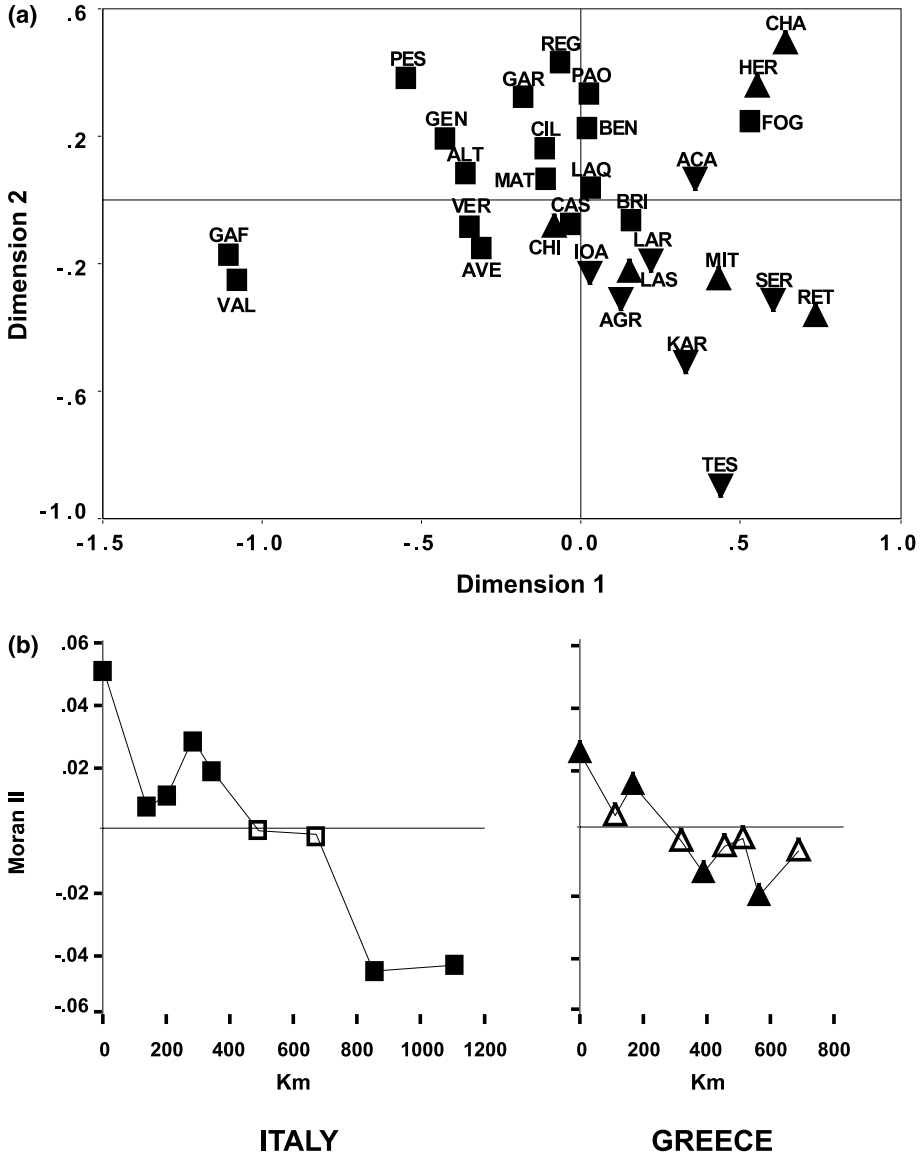

Fig. 2. (a) Plot of sample points in the space defined by the first two dimensions of correspondence analysis. Squares, Italian sample; triangles, Greek samples. Sample codes as in Table 1. (b) Moran II autocorrelation index obtained for 9 equally-represented distance classes for Italian and Greek population samples. Filled symbols, values significantly different from 0; open symbols, values not significantly different from 0.

occupy positions at opposite edges of the Italian cluster. The FOG sample, geographically very close to GAR and BRI, maps far apart from the latter ones, closer to two Cretean samples. As to Greece, the two Cretean samples CHA and HER are clearly separated from both RET and LAS. The AGR and ACA samples (geographically divided only by the gulf of Patras) are widely separated as well as the north-eastern TES and SER samples. Interestingly, the small Aegean islands of Chios and Lesvos are similar to the Greek average, suggesting a continuous immigration or repopulation events. On the other hand it is likely that in the island of Crete, distance from the Continent and a predominantly mountainous environment were the main factors in setting the conditions for internal differentiation.

In order to quantify differentiation among populations, we used AMOVA by considering all Hg's equidistant. We argued that all the Hg's detected in this study arose over a long time-span and not necessarily within the area here considered, having appeared in this area only at a later time through migrations (Chikhi et al., 1998). Therefore, it is likely that the demographic processes which have led to the observed pattern of genetic diversity are more recent than the accumulation of mutations on the chromosomes. As a consequence, the differences among populations are more faithfully described by differences in allele frequencies than by measures of molecular distance between Hg's (Excoffier and Smouse, 1994).

Table 2 shows the F_{st}, F_{sc}, and F_{ct} values for all the geographic groupings analyzed in this study. We found high levels of genetic differentiation when the set of Italian and Greek samples are compared ($F_{st} = 0.077$, $p < 0.01$; $F_{ct} = 0.037$, $p < 0.01$). This corroborates the results of the correspondence analysis, and is in agreement with previous studies (Barbujani and Sokal, 1990) which showed that the Adriatic sea represents a line of sharp change in the composition of European gene pool. However, both the Italian and Greek sets are not homogenous ($F_{sc} = 0.042$, $p < 0.01$), a finding that is also apparent from the results of the correspondence analysis.

The results of the hierarchical AMOVA analysis within the Italian (a non-significant F_{ct} coupled with a significant F_{sc}) suggests that the overall heterogeneity among Italian samples cannot be solely attributed to the peculiarity of the GAF and VAL samples, as they belong to two different and distant regions (Tuscany and Trentino, respectively). On the contrary, this portion of heterogeneity seems to be contributed by the central and southern regions. This result is also confirmed by re-running the analysis after grouping the samples in northen and southern, in order to test the boundary detected by Barbujani and Sokal (1990). In this case a significant F_{ct} is still accompanied by a significant F_{sc}.

As far as Greece is concerned, we tested the hypothesis that the Aegean Sea might represent a relevant barrier to gene flow and compared the continental and island samples. We found a small but significant F_{ct} value, together with a larger and significant F_{sc} value. It is clear that heterogeneity within the island of Crete contributes substantially to the within-groups component of variance. On the other hand, when the continental samples were regrouped into northeastern vs. southwestern, we obtained non-significant fixation indexes, in disagreement with the results by Barbujani and Sokal (1990).

Table 2
Analysis of molecular variance in 17 Italian and 13 Greek population samples

Variance component	Variance (%)	Fixation indexes
All samples		
Among groups (Italy and Greece)	3.67	$F_{ct} = 0.037^{**}$
Among samples within groups	4.03	$F_{sc} = 0.042^{**}$
Within samples	92.30	$F_{st} = 0.077^{**}$
Continental Italy		
Among groups (administrative regions)	3.75	$F_{ct} = 0.038$
Among samples within groups	1.86	$F_{sc} = 0.019^{*}$
Within samples	94.38	$F_{st} = 0.056^{**}$
Continental Italy		
Among groups (north vs. south)	8.15	$F_{ct} = 0.081^{**}$
Among samples within groups	1.88	$F_{sc} = 0.020^{**}$
Within samples	89.97	$F_{st} = 0.100^{**}$
Greece		
Among groups (continent vs. islands)	1.22	$F_{ct} = 0.012^{*}$
Among samples within groups	2.03	$F_{sc} = 0.021^{**}$
Within samples	96.75	$F_{st} = 0.032^{**}$
Continental Greece		
Among groups (north-east vs. south-west)	−0.08	$F_{ct} = -0.0008$
Among samples within groups	0.58	$F_{sc} = 0.0058$
Within samples	99.51	$F_{st} = 0.0049$

$^{*}P < 0.05$.
$^{**}P < 0.01$.

We used autocorrelation analysis (Bertorelle and Barbujani, 1995) to test for decrease in genetic similarity between samples with increasing geographic distance. In Italy, the autocorrelation index (Moran II, Fig. 2b) decreases from significantly positive to significantly negative values. However, this decline is not regular, and shows a sharp decrease of genetic similarity for samples separated by >800 km., i.e., northern vs. southern samples. When all the Greek samples are analyzed (Fig. 2b), a very irregular trend from positive to negative values is observed. When only the continental Greek samples are analyzed, none of the Moran II values differ significantly from zero (not shown).

In order to discriminate the contribution of each Hg to this pattern, we used multiple linear regression of Hg

frequencies as a function of latitude and longitude. This method was applied only to the most frequent Hg's (DE, I-M170, J2-(DYS413 $\leqslant$ 18) and P* (xR1a) if >10% in each national sample). Among the seven regressions, only P* (xR1a) in Italy shows a significant geographic dependence ($F = 12.1$; $p < 0.001$), with an average increase of 3.1% per degree of latitude and a decrease of 2.8% per degree of longitude. Interestingly, the map of this Hg frequencies (Fig. 1) shows equal-frequency lines crossing the Italian peninsula, with a limited number of outlying locations. Note that the method used here, which preserves the observed data points, also produces a central Italian belt of Hg P* (xR1a) even if none of the three locations sampled within this belt (LAQ, AVE, and PES) fall within this range.

3.3. Molecular diversity within Hg's

Given the highly variable spatial pattern of the overall Y chromosomal diversity, we identified the instances of the largest variation in Hg frequencies in a given location by comparison to the corresponding national average. The dinucleotide STRs associated with each Hg could provide hints on possible recent founder or drift effects that raised the frequency of the same Hg in certain locations. The criterion used was an increase of more than 10% in the frequency of a Hg (>1 standard error in all cases) in a local sample compared to the frequency of the same Hg in the corresponding overall national sample (if >5%). We identified (boldface in Table 1) 19 such Hg-location pairs (only 9.6 expected, based on the average sample size and random sampling fluctuation around the national average). When the allele size frequency distributions at YCAII and DYS413 in the above local samples were compared with the rest of samples from the same nation, we found significant departures in 7/19 cases (3 in Italy, 2 in continental Greece and 2 in Crete). In these cases the observed increase in the frequency of the Hg is accompanied by an increase of the frequency of at least one of the alleles at the STRs. For P* (xR1a) in GAF and J2-(DYS413 $\leqslant$ 18) in PAO, a complete loss of variation at the STRs was observed. All R1a chromosomes in RET carried the (CA)25 allele at YCAIIa, which we found in only 20 out of more than 2000 chromosomes sampled in the entire western Eurasia.

3.4. Hg diversity and inferences on the peopling processes

Defining the genetic features of human populations in Italy and Greece is crucial to test hypotheses on the peopling of the northern Mediterranean coast and southern Europe (Ammerman and Cavalli-Sforza, 1984). Groups of males representing the Italian and Greek gene pools contributed to the recognition of continent-wide patterns of geographic distributions of Y chromosomal Hg's. However, different authors attributed these patterns to alternative scenarios for the peopling of Europe (Chikhi et al., 2002; Rosser et al., 2000; Semino et al., 2000; Underhill et al., 2001). Using a large sample size within a microgeographic sampling scheme, the present work provides data to be compared with the expectations under each of these scenarios.

Italy and Greece lie in the central portion of the continent-wide clines across the Mediterranean basin and Europe. Thus, only a subset of the overall Hg frequency variation was to be expected, with the internal variation weakening to some extent the general pattern. Nevertheless, within each country we observe a degree of heterogeneity higher than the one expected as a result of sampling fluctuations. An excess of unusually high Hg frequencies and of significant F_{sc} is observed. This heterogeneity is not structured along the lines of clinal variation deduced from studies on larger spatial scales, since in only one case (P* (xR1a) in Italy) we found a positive regression with geography, which coincides with a sharp rather than a clinal change (Barbujani and Sokal, 1991).

The present work shows a highly complex local distribution of Y chromosomal diversity. Similar results were also obtained by Scozzari et al. (2001) in a survey of Sardinia, Sicily, and 6 Italian regions. Furthermore, Semino et al. (2000) and Rosser et al. (2000), by examining independent Italian and Greek samples, obtained remarkable variations in the frequencies of largely overlapping haplogroups; this discrepancy might result from sampling in different areas of each country.

We explored in detail the instances of local peaks of Hg frequencies. With a few loci it was possible to show that a single STR allele often marks these increases, suggesting a limited number of founders. This suggests that local founder or drift effects seem to play a key role in shaping the microgeographic Y-chromosomal diversity in these populations.

Seielstad et al. (1998) have shown that, at equal geographic distances, population pairs display higher average F_{st} values for the Y chromosome than for both mitochondrial and autosomal markers, as a result of the predominant practice of patrilocality. However, their Y-specific F_{st}-on-distance regression is the result of data points with large fluctuations around the regression line, also consistent with the reduced effective size of the NRY. Our study parallels the previous results both in the magnitude of pairwise F_{st} values and in their large fluctuations as a function of distance, and it also confirms a similar pattern for comparisons between samples less than 100 km apart.

Genetic isolation due to physical barriers can also affect NRY spatial heterogeneity. Among the main features of both Italy and Greece is the prevalence of mountainous and hilly areas. This may have favored isolation, with mountain ranges impairing gene flow and ultimately resulting in zones of rapid genetic change. In a previous study (Stefan et al., 2001), we observed a

systematic change of Hg frequencies among sampling locations spanning a drastic change in the altitudinal features of the environment, from the Carpathians mountains to the steppic plains of eastern Europe in Romania and Moldova. Similarly, six of our Greek samples are island populations. These are likely to have experienced some degree of reproductive isolation, possibly leading to random Hg frequencies fluctuations rather than clinal variations.

Major peopling events may also leave their signature. Only Hg P* (xR1a) in Italy displays a significant decrease in frequencies, from the north-west to the south-east. Many authors agree in considering this Hg as the signature of the Paleolithic inhabitants of the entire European continent. Wilson et al. (2001) have identified a particular STR haplotype within this Hg as the characteristic shared by Celtic-speaking populations and the Basques by common descent from a relatively homogeneous pre-agricultural gene pool. In this context, the most frequent YCAII and DYS413 STR alleles observed in Hg P* (xR1a) from the GAF and VAL samples are identical to the ones observed in 73% of Basques (Malaspina et al., 2000). The Hg homogeneity of GAF and VAL may thus represent a remnant of the pre-agricultural gene pool which now extends to some locations in northern Italy. The pattern reported here is compatible with the introduction of other lineages (DE, G, I, J, and R1a) on a P* (xR1a) background in southern Italy (Underhill et al., 2001). Implicit in this model is that the opposite clines for the alternative Hg's are masked by the local frequencies peaks mentioned above. In Greece, the corresponding P* (xR1a) cline would be lost, due to the overwhelming incidence of the pooled DE, G, I, J, and R1a. In summary, the single instance of a detectable cline would be the result of the incomplete colonization of the Italian territories by newly arrived lineages, a contingent situation seldom replicated in other countries. Piazza et al. (1988) attributed a similar pattern of the first principal component of genetic variation at 34 independent alleles to the Greek colonization in the southern-most region of Italy between 1000 and 400 B.C. We cannot exclude that additional local heterogeneity was contributed by immigrant groups already genetically differentiated as they came from different city-states of ancient Greece.

4. Conclusions

Previous studies describing clinal variation of Y chromosome diversity were generally based on samples supposedly representative of entire national communities or ethnic groups. Even in cases where representativeness of the sample is taken into account, this method implicitly reduces the complexity of spatial variation since the "clinal" description of the variation is largely the result of interpolation. This effect is further enhanced when the description of trends is pursued with smoothing methods such as those discussed by Sokal et al. (1999a,b) and Rendine et al. (1999).

In conclusion, local haplogroup frequencies cannot be simply predicted from the apparent pattern of clinal variation of the Y chromosome. The complexity here described then prompts caution in equating similarity in the frequencies of one or more Y chromosomal haplogroups among populations and common descent. This is particularly true for haplogroups whose origin long predates population splitting. In fact, random fluctuations may erase frequency divergence accumulated during isolation. It then becomes imperative to collect additional molecular information (i.e., a finer characterization of haplogroups with additional, more recent mutations) to confirm or dismiss a recent common ancestry.

Acknowledgments

We thank Prof. G. Barbujani for the VAL samples and for helpful comments on a first draft of this paper. We are grateful to Dr. M. Hammer for providing details of the $DYS221_{136}$ assay and to Dr. M. Jobling for control samples. The comments of three anonymous reviewers are gratefully acknowledged. Work supported by grants Agenzia 2000 and PRIN 2002 to A.N. and P.M. and CNR Grant 01.00646.PF36 to L.T.

References

Ammerman, A.J., Cavalli-Sforza, L.L., 1984. The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton.

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

Barbujani, G., Sokal, R.R., 1991. Genetic population structure of Italy. II. Physical and cultural barriers to gene flow. Am. J. Hum. Genet. 48, 398–411.

Bertorelle, G., Barbujani, G., 1995. Analysis of DNA diversity by spatial autocorrelation. Genetics 140, 811–819.

Chikhi, L., Destro-Bisol, G., Bertorelle, G., Pascali, V., Barbujani, G., 1998. Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. Proc. Natl. Acad. Sci. USA 95, 9053–9058.

Chikhi, L., Nichols, R.A., Barbujani, G., Beamont, M.A., 2002. Y genetic data support the Neolithic demic diffusion model. Proc. Natl. Acad. Sci. USA 99, 11008–11013.

de Knijff, P., 2000. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosomes. Am. J. Hum. Genet. 67, 1055–1061.

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

Excoffier, L., Smouse, P.E., Quattro, J.M., 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes:

application to human mitochondrial DNA data. Genetics 131, 479–491.

Excoffier, L., Smouse, P.E., 1994. Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics 136, 343–359.

Hammer, M.F., Zegura, S.L., 2002. The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Annu. Rev. Anthropol. 31, 303–321.

Hammer, M.F., Horai, S., 1995. Y chromosomal DNA variation and the peopling of Japan. Am. J. Hum. Genet. 56, 951–962.

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., Zegura, S.L., 1998. Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol. Biol. Evol. 15, 427–441.

Hammer, M.F., Karafet, T.M., Redd, A.J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., Zegura, S.L., 2001. Hierarchical patterns of global Y-chromosome diversity. Mol. Biol. Evol. 18, 1189–1203.

King, J.P., Kimmel, M., Chakraborty, R., 2000. A population analysis of microsatellite-based statistics for inferring past population growth. Mol. Biol. Evol. 17, 1859–1868.

Kwok, C., Tyler-Smith, C., Mendonca, B.B., Hughes, I., Berkovitz, G.D., Goodfellow, P.N., Hawkins, J.R., 1996. Mutation analysis of the 2 kb 5′ to SRY in XY females and XY intersex subjects. J. Med. Genet. 33, 465–468.

Lell, J.T., Wallace, B.C., 2001. The peopling of Europe from the maternal and paternal perspectives. Am. J. Hum. Genet. 67, 1376–1381.

Malaspina, P., Ciminelli, B., Viggiano, L., Jodice, C., Cruciani, F., Santolamazza, P., Sellitto, D., Scozzari, R., Terrenato, L., Rocchi, M., Novelletto, A., 1997. Characterization of a small family (CAIII) of microsatellite-containing sequences with X–Y homology. J. Mol. Evol. 44, 652–659.

Malaspina, P., Cruciani, F., Ciminelli, B.M., Terrenato, L., Santolamazza, P., Alonso, A., Banyko, J., Brdicka, R., Garcia, O., Gaudiano, C., Guanti, G., Kidd, K.K., Lavinha, J., Avila, M., Mandich, P., Moral, P., Qamar, R., Mehdi, S.Q., Ragusa, A., Stefanescu, G., Caraghin, M., Tyler-Smith, C., Scozzari, R., Novelletto, A., 1998. Network analyses of Y-chromosomal types in Europe, North Africa and West Asia reveal specific patterns of geographical distribution. Am. J. Hum. Genet. 63, 847–860.

Malaspina, P., Cruciani, F., Santolamazza, P., Torroni, A., Pangrazio, A., Akar, N., Bakalli, V., Brdicka, R., Jaruzelska, J., Kozlov, A., Malyarchuk, B., Mehdi, S.Q., Michalodimitrakis, E., Varesi, L., Memmi, M.M., Vona, G., Villems, R., Parik, J., Romano, V., Stefan, M., Stenico, M., Terrenato, L., Novelletto, A., Scozzari, R., 2000. Patterns of male-specific inter-population divergence in Europe, West Asia and North Africa. Ann. Hum. Genet. 64, 395–412.

Malaspina, P., Tsopanomichalou, M., Duman, T., Stefan, M., Silvestri, A., Rinaldi, B., Garcia, O., Giparaki, M., Plata, E., Kozlov, A.I., Barbujani, G., Vernesi, C., Papola, F., Ciavarella, G., Kovatchev, D., Kerimova, M.G., Anagnou, N., Gavrila, L., Veneziano, L., Akar, N., Loutradis, A., Michalodimitrakis, E.N., Terrenato, L., Novelletto, A., 2001. A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. Ann. Hum. Genet. 65, 339–349.

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

Piazza, A., Cappello, N., Olivetti, E., Rendine, S., 1988. A genetic history of Italy. Ann. Hum. Genet. 52, 203–213.

Rendine, S., Piazza, A., Menozzi, P., Cavalli-Sforza, L.L., 1999. A problem with synthetic maps: reply to Sokal et al. Hum. Biol. 71, 15–25.

Renfrew, C., 2001. From molecular genetics to archaeogenetics. Proc. Natl. Acad. Sci. USA 98, 4830–4832.

Romualdi, C., Balding, D., Nasidze, E.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., Barbujani, G., 2002. Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res. 12, 602–612.

Rosser, Z.H., Zerjal, T., Hurles, M., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Côrte-Real, H.B.S.M., De Knijff, P., Decorte, R., Dubrova, Y.E., Grafo, O., Gilissen, A., Glisic, S., Gölge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Lavinha, J., Livshits, L.A., Malaspina, P., Syrrou, M., Mcelreavey, K., Meitinger, T.A., Melegh, B., Mitchell, R.J., Nicholson, J., Nørby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., João Prata, M., Previderé, C., Rajczy, K., Roewer, L., Rootsi, S., Rubinsztein, D.C.M, Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Olun, A., Villems, R., Tyler-Smith, C., Jobling, M.A., 2000. Y-chromosomal diversity within Europe is clinal and influenced primarily by geography rather than language. Am. J. Hum. Genet. 66, 1526–1543.

Schneider, S., Kueffer, J.-M., Roessli, D., Excoffier, L., 1997. Arlequin ver.1.1, a software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Scozzari, R., Cruciani, F., Pangrazio, A., Santolamazza, P., Vona, G., Moral, P., Latini, V., Varesi, L., Memmi, M.M., Romano, V., De Leo, G., Gennarelli, M., Jaruzelska, J., Villems, R., Parik, J., Macaulay, V., Torroni, A., 2001. Human Y-chromosome variation in Western Mediterranean area: implications for the peopling of the region. Hum. Immunol. 62, 871–884.

Seielstad, M.T., Minch, E., Cavalli-Sforza, L.L., 1998. Genetic evidence for a higher female migration rate in humans. Nature Genet. 20, 278–280.

Semino, O., Passarino, G., Oefner, P., Lin, A.A., Arbuzova, S., Beckman, L.E., de Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhill, P.A., 2000. The genetic legacy of Paleolithic Homo sapiens in extant Europeans: a Y chromosome perspective. Science 290, 1155–1159.

Sokal, R.R., Oden, N.L., 1978. Spatial autocorrelation analysis in biology. I. Methodology. Biol. J. Linn. Soc. 10, 199–228.

Sokal, R.R., Oden, N.L., Thomson, B.A., 1999a. A problem with synthetic maps. Hum. Biol. 71, 1–13.

Sokal, R.R., Oden, N.L., Thomson, B.A., 1999b. Problems with syntetic maps remain: reply to Rendine et al. Hum. Biol. 71, 447–453.

SPSS Inc., 1991. SPSS statistical algorithms, second ed., SPSS Inc., Chicago.

Stefan, M., Stefanescu, G., Gavrila, L., Terrenato, L., Jobling, M.A., Malaspina, P., Novelletto, A., 2001. Y chromosome analysis reveals a sharp genetic boundary in the Carpathian region. Eur. J. Hum. Genet. 9, 27–33.

Underhill, P.A., Shen, P., Lin, A., Jin, A.A., Passarino, G., Yang, W.H., Kauffman, E., Bonné-Tamir, B., Bertranpetit, J., Francalacci, P., Ibrahim, M., Jenkins, T., Kidd, J.R., Mehdi, S.Q., Seielstad, M., Wells, S., Piazza, A., Davis, R.W., Feldman, M.F., Cavalli-Sforza, L.L., Oefner, P.J., 2000. Y chromosome sequence variation and the history of human populations. Nat. Genet. 26, 358–361.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Mirazon-Lahr, M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L., 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann. Hum. Genet. 65, 43–62.

Wilson, J.F., Weiss, D.A., Richards, M., Thomas, M.G., Bradman, N., Goldstein, D.B., 2001. Genetic evidence for different male and female roles during cultural transition in the British Isles. Proc. Natl. Acad. Sci. USA 98, 5078–5083.

Y Chromosome Consortium, 2002. A nomenclature system for the tree of human Y chromosomal binary haplogroups. Genome Res. 12, 339–348.