Ann. Hum. Genet. (2001), **65**, 339–349
Printed in Great Britain

A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area

P. MALASPINA[1], M. TSOPANOMICHALOU[2], T. DUMAN[3], M. STEFAN[4], A. SILVESTRI[1], B. RINALDI[1], O. GARCIA[5], M. GIPARAKI[6], E. PLATA[6], A. I. KOZLOV[7], G. BARBUJANI[8], C. VERNESI[8], F. PAPOLA[9], G. CIAVARELLA[10], D. KOVATCHEV[11], M. G. KERIMOVA[12], N. ANAGNOU[13], L. GAVRILA[4], L. VENEZIANO[14], N. AKAR[3], A. LOUTRADIS[6], E. N. MICHALODIMITRAKIS[2], L. TERRENATO[1] AND A. NOVELLETTO[15]

[1] *Department of Biology, University 'Tor Vergata', Rome, Italy*
[2] *Department of Forensic Sciences, University of Crete, Heraklion, Greece*
[3] *Pediatrics Department, Ankara University, Turkey*
[4] *Genetics Department, University of Bucharest, Rumania*
[5] *Basque Country Police, Bilbao, Spain*
[6] *Ministry of Health Center for Thalassemia, Athens, Greece*
[7] *Arct-An C Innovative Laboratory, Moscow, Russian Federation*
[8] *Department of Biology, University of Ferrara, Italy*
[9] *Centro Regionale di Immunoematologia e Tipizzazione Tissutale, L'Aquila, Italy*
[10] *IRCCS Casa Sollievo della Sofferenza, S. Giovanni Rotondo, Italy*
[11] *Department of Biology, Medical University, Varna, Bulgaria*
[12] *Department of Hygiene, Azerbaijan Medical University, Baku, Azerbaijan*
[13] *Institute of Molecular Biology and Biotechnology, Heraklion and University of Athens, Greece*
[14] *Institute of Experimental Medicine, CNR, Rome, Italy*
[15] *Department of Cell Biology, University of Calabria, Rende, Italy*

(Received 13.2.01. Accepted 3.5.01)

SUMMARY

In this work we focus on a microsatellite-defined Y-chromosomal lineage (network 1.2) identified by us and reported in previous studies, whose geographic distribution and antiquity appear to be compatible with the Neolithic spread of farmers. Here, we set network 1.2 in the Y-chromosomal phylogenetic tree, date it with respect to other lineages associated with the same movements by other authors, examine its diversity by means of tri- and tetranucleotide loci and discuss the implications in reconstructing the spread of this group of chromosomes in the Mediterranean area. Our results define a tripartite phylogeny within HG 9 (Rosser *et al.* 2000), with the deepest branching defined by alleles T (Haplogroup Eu10) or G (Haplogroup Eu9) at M172 (Semino *et al.* 2000), and a subsequent branching within Eu9 defined by network 1.2. Population distributions of HG 9 and network 1.2 show that their occurrence in the surveyed area is not due to the spread of people from a single parental population but, rather, to a process punctuated by at least two phases. Our data identify the wide area of the Balkans, Aegean and Anatolia as the possible homeland harbouring the largest variation within network 1.2. The use of recently proposed tests based on the stepwise mutation model suggests that its spread was associated to a population expansion, with a high rate of male gene flow in the Turkish–Greek area.

Correspondence: Andrea Novelletto, Dept. of Cell Biology, University of Calabria, Via P. Bucci, 7030 Arcavacata di Rende, Italy. Tel: +39-0984-492933; Fax: +39-0984-492911.
E-mail: novelletto@bio.uniroma2.it

introduction

The non-recombining portion of the human Y chromosome (NRPY) represents, together with mtDNA, a uniparentally inherited polymorphic system. This property is also associated with the ability of the NRPY to detect high levels of structuring within and between populations, thus providing a first framework to be compared with weaker patterns detected by markers with alternative modes of inheritance. The final goal is the description of the history of populations in terms of the different realizations of the dispersal of lineages at loci that represent different portions of the genome (Owens & King, 1999; Sykes, 1999; Bertranpetit, 2000; Lell & Wallace, 2000; Renfrew *et al.* 2000; Barbujani & Bertorelle, 2001).

Major advances in understanding the build-up of the overall DNA variation of NRPY include: (*a*) the reconstruction of unequivocal phylogenies based on different sets of indel and single nucleotide polymorphic (SNP) binary markers (Karafet *et al.* 1999; Santos *et al.* 1999, Hammer *et al.* 2000; Rosser *et al.* 2000; Underhill *et al.* 2000) and (*b*) the estimation of the antiquity of lineages based on either the infinite site mutation model for binary markers or the stepwise mutation model at microsatellite loci (Shen *et al.* 2000; Thomson *et al.* 2000; Forster *et al.* 2000; Stumpf & Goldstein, 2001).

A strong structuring of the extant populations of Europe, the Middle East and North Africa has recently been confirmed by analyses of independent sets of binary indel and SNP markers of the NRPY (Semino *et al.* 2000; Rosser *et al.* 2000). In addition, the combined use of binary and microsatellite markers (Malaspina *et al.* 1998, 2000) has enabled clarification of the phylogeography in this area, based on an approach which partitions the microsatellite variability into the binary-defined lineages (de Knijff, 2000) and takes into account multirepeat mutations disrupting the regular stepwise process. In fact, uncommon microsatellite multirepeat mutational events which could be regarded as a disturbance in the regular process of accumulation of variation are instead an important resource, since they may identify groups of chromosomes with a common ancestry (Forster *et al.* 1998; Kayser *et al.* 2001) that cannot otherwise be recognized by binary markers.

All the above studies have identified clear patterns in the geographic distributions of haplotypes or haplogroups as well as sharp genetic boundaries, and have proposed models for the underlying population movements. In this context, the integration of all the available markers into a single phylogeny could increase the number of detectable lineages to attain a better description of the molecular evolution of the NRPY, and to address specific questions concerning the peopling of the area being examined.

In this work we focus on a lineage (network 1.2) we have identified and reported in previous studies (Malaspina *et al.* 1998, 2000) whose geographic distribution and antiquity appear to be compatible with the Neolithic spread of farmers. Here, we set network 1.2 in the NRPY phylogenetic tree, date it with respect to other lineages associated with the same movements by other authors (Eu9, Eu10, Semino *et al.* 2000; Med, Hammer *et al.* 2000; HG 9, Rosser *et al.* 2000; Quintana-Murci *et al.* 2001), suggesting that the different sublineages contributing to HG9 probably had different evolutionary histories in the Mediterranean area.

materials and methods

Subjects

An overall number of 1802 males from Europe, West Asia and Egypt was collected from a total of 55 sampling locations. To obtain suitable sample sizes, the local samples were pooled according to nationality except for the Mediterranean Islands, which were kept separate from the mainland countries to which they belong in view of their possible isolation. A total of 24 sample pools was thus gathered. Many of the local samples are a subset of those described previously (Malaspina *et al.* 1998, 2000; Stefan *et al.* 2001), although minor discrepancies in the number of subjects can occasionally be found.

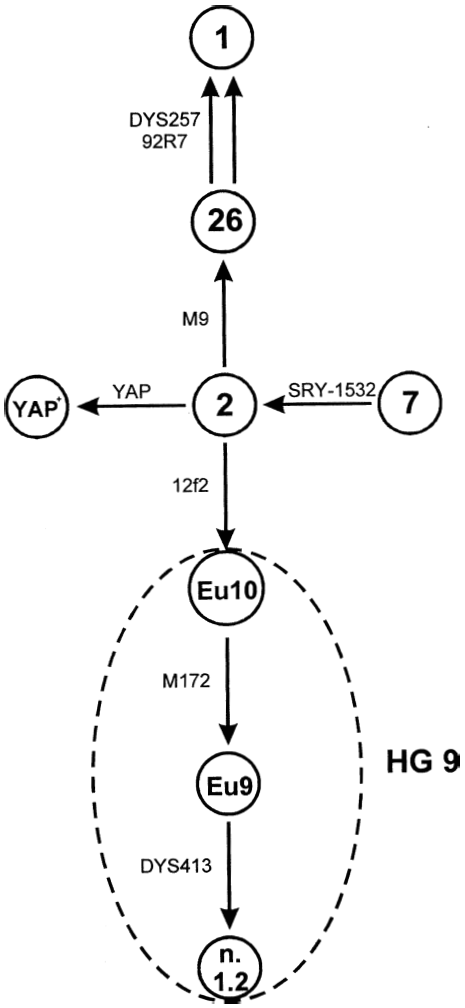

Fig. 1. Simplified version of the NRPY maximum parsimony network by Rosser *et al.* (2000) showing haplogroups 1, 2, 7, 26 and YAP+. Arrows represent the biallelic defining mutations at the indicated loci. Haplogroup 9 is represented by a dashed ellipse. Sublineages within HG 9 correspond to haplogroups Eu10 and Eu9 (Semino *et al.* 2000) and to network 1.2.

Additional samples described here for the first time include: two local samples from the Azerbaijan Republic, collected in Baku and Lenkoran, respectively; one sample collected around the city of Aleppo, Syria; an enlargement of the sample of Spanish Basques from the province of Guipuzcoa; four and seven samples from the different administrative districts of the Island of Crete and mainland Greece, respectively; one sample from the Aegean Island of Mitilene; one sample of Russians from the Perm region (Russian Federation); one sample from the region of Erzurum (Eastern Turkey); one sample from Bulgaria, collected in Varna; three samples from different provinces of Central and Southern Italy.

Markers

The presence of the rearrangement defining haplogroup 9 (HG 9) of Rosser *et al.* (2000) was detected by failure in producing an 88 bp fragment in the PCR assay described therein. This rearrangement is also detectable as an 8 kb *Taq*I band detectable by probe p12f2 (Casanova *et al.* 1985), making HG 9 coincide with haplotype Med by Hammer *et al.* (2000). With the exception of the previously typed Romanian local samples (Stefan *et al.* 2001), the assay was applied as a new typing to all carriers of the ancestral allele states at both the YAP insertion (Hammer & Horai, 1995) and DYS257 (Hammer *et al.* 1998) (Fig. 1). All subjects were typed for the DYS413 dinucleotide microsatellite system (Mathias *et al.* 1994; Malaspina *et al.* 1997).

A subset of HG 9 males was typed for microsatellites DYS388 and DYS390 as a duplex PCR under the conditions described by Thomas *et al.* (1999). They were also typed for trinucleotide DYS392 and tetranucleotide DYS393 under the conditions described by Perez-Lezaun *et al.* (1999). Allele size was determined by comparison with specimens previously typed (Caglià *et al.* 1997; Perez-Lezaun *et al.* 1999). Primers were 32-P end labelled and PCR products resolved on 6% polyacrylamide gels.

The T-G transversion at M172 (Underhill *et al.* 2000) was assayed by PCR followed by dot–blot hybridization with the 32-P labelled 17-mer ASO probes 5′-TTTTAAGTCAAGCATCA-3′(G allele) and 5′-TGATGCTTTACTTAAAA-3′(T allele). Hybridization was carried out in 5 × SSPE, 0.5% SDS, 5 × Denhardt's solution at 43 °C for 2 h. Filters were washed in 5 × SSPE, 0.1% SDS at 43 °C.

Statistical analyses

All calculations were performed with SSPS v. 6.1. Indexes to test the null hypothesis of population constant size were calculated

according to eqs. 1–3 by Reich & Goldstein (1998) and eqs. 5 and 7–8 by King *et al.* (2000).

An estimate of the microsatellite mutation rate for the loci here used was obtained by pooling the data available in the literature (Heyer *et al.* 1997; Bianchi *et al.* 1998; Foster *et al.* 1998; Kayser *et al.* 2000). Six mutational events in a total of 2958 meioses were observed, giving an average mutation rate (μ) of 2.03×10^{-3}. Confidence interval (CI) was calculated as $\mu \pm 1.96[\mu(1-\mu)/N]^{1/2}$, where N is the number of meioses, thus obtaining the interval $0.41–3.65 \times 10^{-3}$.

RESULTS

Phylogenetic relationships

Chromosomes grouped into network 1.2 are identified by short CA repeats ($\leqslant 18$) in both PCR fragments at DYS413. All chromosomes within this group can be linked to each other in a network by assuming insertion or deletion of a single CA unit in one of the fragments. By the same criterion, they could not be linked to any other chromosome in a sample of 1801 chromosomes (Malaspina *et al.* 2000) from Western Eurasia and North Africa.

Of the 250 network 1.2 chromosomes identified in the present survey (see below), 249 belong to HG 9. A single subject from Azerbaijan produced an 88 bp amplification product under the conditions by Rosser *et al.* (2000), and therefore was not considered as HG 9. The molecular basis of this result is currently under investigation.

Additional 154 chromosomes were identified as HG 9 but showed DYS413 fragment sizes outside the network 1.2 range.

All of network 1.2 chromosomes carried the derived G allele at M172, with the exception of the single non-HG 9 subject who carried M172-T. Both G and T alleles were found among the rest of HG 9 chromosomes.

Based on the observation that the derived 12f2 8 kb allele is shared by haplogroups Eu9 and Eu10 (Semino *et al.* 2000), we assigned to these haplogroups HG 9 chromosomes with M172 alleles G and T, respectively. Our results then define a tripartite phylogeny within HG 9, with the deepest branching represented by haplogroups Eu10 and Eu9, and a subsequent branching defined by network 1.2 (Fig. 1).

Population distributions

Table 1 reports the absolute and relative frequencies of HG 9 in the 24 national and island sample pools. A clear trend can be observed across the continental mainland, with the highest frequencies in Central and South West Asian countries, decreasing westward in the Balkan and Italian peninsula and dropping sharply in Central, Western and Northern Europe. Three island populations appear to be genetic outliers: Crete contrasts for its peak frequency of .40 with the rest of Greece, whereas Sardinia and Sicily contrast with Italy for their low frequencies.

Table 1 also reports the absolute and relative frequencies of network 1.2. Among the continental sample pools the highest frequencies are observed in Azerbaijan and Turkey. Overall, the dependence of the relative frequencies on geography parallels that of HG 9, with some notable exceptions. These can be better visualized when the frequency of network 1.2 is reported as a percentage of HG 9 (Table 1, last col.; Fig. 2). Four countries of Central and Western Europe are characterized by null contributions to a low (< 0.15) HG 9 frequency (the value of the Basques is not reliable due to a single network 1.2-HG 9 observation). Oman, Syria and UAE are characterized by high (> 0.35) HG 9 frequencies and by an intermediate (15–40%) contribution of network 1.2. All other sample pools with HG 9 frequencies $\geqslant 0.20$ are characterized by large ($> 50\%$) contributions of network 1.2, which reach approx. 80% in the large samples from Crete and Continental Italy. Overall, the proportions of network 1.2 within HG 9 are highly heterogeneous ($\chi^2 = 168$, 23 d.f., $p < 10^{-5}$) and both the sampling method and the geographical structuring of their variation exclude the possibility that they are due to local drift/founder effects.

Table 1. *Haplogroup 9 and network 1.2 absolute and relative frequencies in 24 sample pools*

Geographic region sample	No. of local sub-samples	No. of subjects	HG9 (freq.)	Network 1.2 (freq.)	Network 1.2 as % of HG9
Central Western Asia					
Azerbaijan Rep.	2	58	22 (0.379)	14 (0.241)	63.6
Syria	1	54	20 (0.370)	8 (0.148)	40.0
Cont. Turkey	5	186	65 (0.349)	37 (0.199)	56.9
Cyprus (Turkish)	1	46	13 (0.283)	11 (0.239)	84.6
South Western Asia					
Iran	1	6	1 (0.167)	1 (0.167)	100
Oman	1	13	5 (0.385)	2 (0.154)	40.0
United Arab Emirates	1	34	13 (0.382)	2 (0.059)	15.4
North Africa					
Egypt	2	46	11 (0.239)	4 (0.087)	36.4
South Eastern Europe					
Romania	7	165	22 (0.133)	15 (0.091)	68.2
Bulgaria	1	39	9 (0.231)	6 (0.154)	66.7
Albania	1	45	11 (0.244)	3 (0.067)	27.3
Cont. Greece	8	182	37 (0.203)	20 (0.110)	55.6
Crete	5	205	82 (0.400)	66 (0.322)	80.5
Mitilene	1	28	8 (0.286)	5 (0.179)	62.5
South Central Europe					
Cont. Italy	8	202	54 (0.267)	43 (0.213)	79.6
Sicily	1	23	2 (0.087)	1 (0.043)	50.0
Sardinia	2	56	9 (0.161)	4 (0.071)	44.4
South Western Europe					
Spain (Basques)	1	85	1 (0.012)	1 (0.012)	100
Portugal	2	44	2 (0.045)	0	0
Central Europe					
U.K.	1	20	1 (0.050)	0	0
Czech Rep.	1	46	3 (0.065)	1 (0.022)	33.3
Slovak Rep.	1	23	3 (0.130)	0	0
Northern Europe					
Rep. of Moldova	1	72	2 (0.028)	0	0
Russian Fed.	3	124	7 (0.056)	5 (0.040)	71.4

We may then conclude that the HG 9 distribution in the surveyed area is not due to the uniform spread of people from a single parental population, since it appears as a collection of lineages with different evolutionary histories.

Diversity analysis

With the aim of determining the most likely region of origin for network 1.2, and its antiquity with respect to the other lineages within HG 9, we haplotyped a geographically representative subset of 207 chromosomes for the DYS388, DYS390, DYS392 and DYS393 microsatellites. Allele size distributions are markedly unimodal at all four loci, suggesting the signature of an expanding population (Fig. 3).

Mean and variance of allele sizes are reported in Table 2 for the sample pools with meaningful size ($n \geqslant 15$), as well as for larger geographic aggregations. The largest variances, after averaging across the four loci, are found in Continental Greece, Crete and Romania (> 0.40), followed by Continental Turkey (0.36) and Italy (0.32). A super-pool consisting of all typed network 1.2 chromosomes from West Asia, except Turkey, produced the low value of 0.31. Considering that the area from which a population spread is generally characterized by a comparatively higher genetic variance than the areas colonized later (Wooding & Ward, 1997; Barbujani, 2000), these data identify the Balkans, Aegean and Anatolia as the possible homeland harbouring the largest variation within network 1.2, with decreasing values both east/south-east and west of it.

Within network 1.2 the commonest haplotype by far is 15-23-11-12 at loci DYS388-DYS390-DYS392-DYS393, which accounts for 38% (78/207). The second and third most common

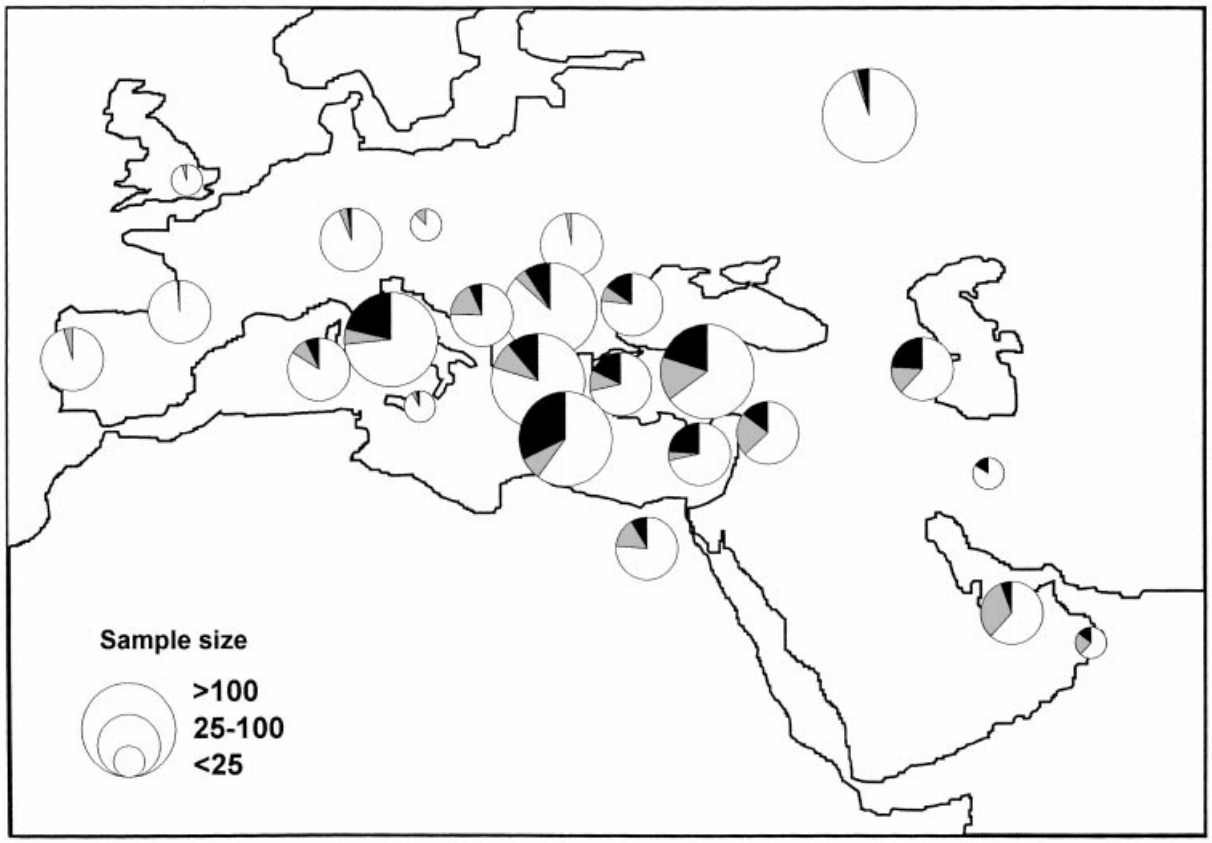

Fig. 2. Pie charts of HG 9 and network 1.2 frequencies in 24 sample pools in Europe, West Asia and North Africa. Black sectors indicate network 1.2 frequencies. Grey sectors indicate frequencies of the rest of HG 9 chromosomes. Three discrete pie sizes are used for samples consisting of < 25, 25–100 and > 100 subjects.

haplotypes are characterized by the gain of a single unit at DYS388 (24 chromosomes) and the loss of a unit at DYS390 (15), as compared to the major haplotype, respectively. Except for Crete, the analysis of the sampling locations of each haplotype does not reveal any particular clustering of peculiar molecular types in specific areas.

A founder effect in Crete

Within network 1.2, two relevant haplotypes related by the gain/loss of a single repeat unit are haplotypes 15-25-11-12 and 15-26-11-12. The first was observed in 6 subjects, 5 from Crete and one from Northern Turkey, whereas the second was observed in 8 subjects, all from Crete. An additional feature of this latter group of chromosomes is the presence of the DYS413 pattern (CA)18-(CA)16. In the entire study this pattern was observed in 21 subjects only, 18 of whom are from Crete. Within the island, 9/18 (as compared to 31/205 sampled chromosomes) were from the Western district of Chania. As a consequence of the excess presence of the 25 and 26 repeat alleles at DYS390, the high average variance value for the 60 network 1.2 chromosomes sampled in Crete (Table 2) is strongly affected by the large value of DYS390. We attributed the above findings to a recent founder or drift effect and therefore did not include the Cretan sample for analysis of population expansion and comparative dating.

Analysis of population expansion

Allele size distributions were used to test the null hypothesis of a constant population size during network 1.2 spread (Reich & Goldstein, 1998; King *et al.* 2000). Among the tests to evaluate departures from the above assumption,

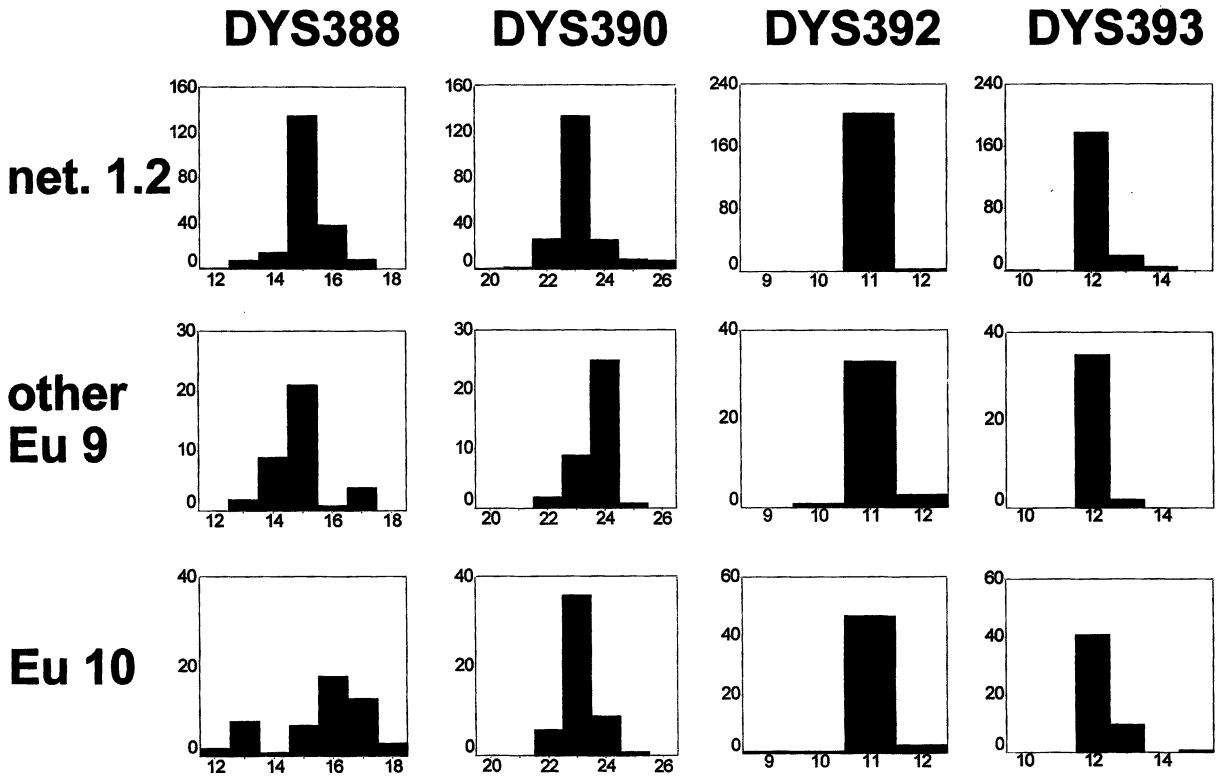

Fig. 3. Allele size distributions at four microsatellite loci in 207 network 1.2, 37 other Eu9 and 52 Eu10 chromosomes.

the imbalance test produced ln β_1 and ln β_2 values of -0.601 ($p \ll 0.05$) and -0.18 (n.s.), respectively; the within-locus k test produced negative values at all loci ($p = 0.055$) and the interlocus g test produced the value of 0.329 ($p < 0.05$). We may then conclude that the spread of network 1.2 was accompanied by a demographic population expansion.

Comparative dating

In order to compare the diversity attained by network 1.2 with that of its parental lineages (see above), we haplotyped 37 HG 9 chromosomes carrying G at M172 (Eu9) but not belonging to network 1.2 and 52 HG 9 chromosomes carrying the ancestral allele T at M172 (Eu10). The two groups display a unimodal allele size distribution at DYS390, DYS392 and DYS393 although a shift in the modal allele is observed at DYS390 (Fig. 3). In both groups DYS388 produces ragged distributions with modes at 15, 17 (other Eu9 chromosomes) and 13, 16 repeats (Eu10).

Among the Eu9 chromosomes not belonging to network 1.2, the commonest haplotype (15-24-11-12) at loci DYS388-DYS390-DYS392-DYS393, respectively, accounts for 40.5% (15 chromosomes), whereas among the Eu10 chromosomes the commonest haplotype (16-23-11-12) accounts for only 26.9% (14 chromosomes). Eu9 chromosomes not belonging to network 1.2 produced an average variance similar to network 1.2 whereas the Eu10 chromosomes produced a value nearly double that (Table 2). This difference can be considered significant, since two out of four F ratios for variance comparisons are significant at $p < 0.01$ (in both Eu10 vs. Eu9 and Eu10 vs. network 1.2).

By assuming a linear relationship between accumulation of variance and time (Thomas *et al.* 1998), the figure of 2.03×10^{-3} as the average mutation rate at microsatellites and its confidence interval of $0.41–3.65 \times 10^{-3}$, the age of network 1.2, and of the rest of Eu9 and Eu10 lineages, can be estimated at 195 (CI 108–963), 182 (CI 101–902) and 417 (CI 232–2067)

Table 2. *Mean and variance of repeat size at four microsatellite loci in network 1.2, other Eu9 and Eu10 chromosomes*

Chromosomes Sample	No. of typed chr.	Mean Variance				
		DYS388	DYS390	DYS392	DYS393	Avg.
Network 1.2						
Cont. Turkey	31	15.32	22.97	11.03	12.03	
		1.09	0.30	0.03	0.03	0.3625
Rest of West Asia	23	14.91	23.30	11.00	12.22	
		0.45	0.22	0.00	0.54	0.3100
All West Asia	54	15.15	23.11	11.02	12.11	
		0.85	0.29	0.02	0.25	0.3525
Romania	15	14.87	23.13	11.07	12.20	
		0.98	0.41	0.07	0.17	0.4075
Cont. Greece	19	15.21	22.89	11.11	12.16	
		1.06	0.54	0.10	0.25	0.4870
Crete	60	15.08	23.58	11.00	12.02	
		0.31	1.47	0.00	0.05	0.4570
Balkan penins. (Greece excl.)	24	15.00	23.21	11.04	12.12	
		0.96	0.52	0.04	0.37	0.4737
All Balkan peninsula	43	15.09	23.07	11.07	12.14	
		0.99	0.54	0.07	0.31	0.4775
Cont. Italy	19	15.00	22.58	11.00	12.42	
		0.33	0.37	0.00	0.59	0.3225
All Network 1.2	207	15.11	23.16	11.02	12.13	
		0.62	0.84	0.02	0.24	0.4300
All Network 1.2 (Crete excl.)	147	15.12	22.99	11.03	12.18	
		0.74	0.49	0.03	0.31	0.3950
Other Eu9						
All areas	37	14.89	23.68	11.05	12.05	
		0.93	0.39	0.11	0.05	0.3700
Eu10						
All areas	52	15.58	23.10	11.00	12.25	
		2.56	0.36	0.16	0.31	0.8475

generations, respectively. Stumpf & Goldstein (2001) recently discussed an estimation method which allows for a length-dependent mutation rate. We did not consider this refinement here, in view of the very slight differences (< 0.7 repeats) in the mean allele size at each locus across the three lineages (Table 2). Assuming an average generation time of around 25 years (Aris-Brosou & Excoffier 1996), the figures reported above roughly translate into 4900 (CI 2700–24000), 4550 (CI 2500–22500), and 10400 (CI 5800–51700) years, respectively.

Discussion

In previous studies we hypothesized that a group of chromosomes (network 1.2) defined solely on the basis of the pattern at dinucleotide microsatellites could represent a monophyletic lineage stemming from a rare multirepeat deletion event. This hypothesis was supported by phylogeographical observations (Malaspina *et al.* 1998). Here we show that such a hypothesis resists the more stringent definition of lineages based on rare or unique events binary polymorphisms, since network 1.2 falls entirely in HG 9 and shows a subset of chromosomes within Eu9. The present work provides a solid overview of the network 1.2 distribution, based on large population samples. Furthermore it provides new data for the description of the HG 9 frequency over a wide area, including the Mediterranean islands, complementing those of Rosser *et al.* (2000) and Quintana-Murci *et al.* (2001). Our data were obtained on entirely independent population samples and show full agreement with those for the shared geographic areas or countries.

As far as HG 9 is concerned, the population data by Semino *et al.* (2000) show large variations in the relative proportions of Eu9 and Eu10 across Europe and the Middle East, with a progressively larger contribution of Eu9 as one moves westward. Our data fit with this trend, since we found Eu9 : Eu10 ratios not greater than 1 : 1 in Syria, the Arab peninsula and Egypt, but up to 12.5 : 1 in the Balkans and the Italian peninsula. Moreover we show that this trend is accounted for by a subset of Eu9, network 1.2.

Overall, our data suggest that two independent evolutionary processes led to the current distribution of HG 9 chromosomes in Europe and in the Mediterranean area. Quintana-Murci *et al.* (2001) identified Northern Iran as the possible focal area from which HG 9 has dispersed. Chromosomes belonging to network 1.2 probably underwent further differentiation in relative isolation, and their current distribution reflects a more recent gene flow. This latter process seems to have originated in the Turkish-Greek area with further spread to the West and a partial back migration to the East/South East. The absence of obvious geographical confinement of network 1.2 haplotypes points towards a massive and rapid exchange of people across the Turkish-Greek area. Hammer *et al.* (2000) described the features of a common Jewish and non-Jewish Y-chromosomal pool in the Middle East, characterized by high HG 9 frequencies. Network 1.2 is a relevant component of this pool. By typing a limited number of Italian Cohanim (A. N. unpublished obs.) for the STRs used here, we determined that the Cohen Modal Haplotype ('an important component in the sharing of Ashkenazic and Sephardic Israelite Y chromosomes', Thomas *et al.* 2000) does indeed belong to network 1.2.

The two phases of HG 9 dispersal seem to be widely separated in time, too. For network 1.2, the use of tri- and tetranucleotide haplotyping produced a more recent dating estimate than the one based on dinucleotide STRs (Malaspina *et al.* 2000). The number of chromosomes haplotyped here does not allow us to distinguish the diversity of network 1.2 from that of its immediate parental lineage (Eu9-non network 1.2). On the other hand, the diversity of the Eu10 lineage is remarkably and significantly greater.

The dating values for HG 9 by Quintana-Murci *et al.* (2001) are likely to be affected by the presence of network 1.2 chromosomes in their data. Their major HG 9 haplotype in Iran shares alleles DYS388-15, DYS390-23, DYS392-11 and DYS393-12 with the major haplotype found in network 1.2, whereas major haplotypes in the rest of Eu9 and Eu10 chromosomes differ at DYS390 and DYS388, respectively. Moreover, we have already reported an overall frequency of 5.5% for network 1.2 among Pakistanis (Malaspina *et al.* 2000).

Translating diversity values into absolute dating figures could give clues to link the observed genetic patterns with archaeologically documented population processes. Unfortunately, absolute dating figures are heavily affected by great uncertainties in mutation rates and ignorance of generation time resulting from life-styles of past human societies (Stumpf & Goldstein, 2001). In this context comparative diversity analysis indicates an older HG 9 and a more recent network 1.2 dispersals. Semino *et al.* (1996) first hypothesized that the overall HG 9 frequency pattern in Europe marked the Neolithic spread of farmers, this view also being shared by Rosser *et al.* (2000). Hammer *et al.* (2000) proposed that more recent sea migrations of the Phoenicians could also have contributed to it. Our two step model includes high rates of male gene flow across the lands of modern day Turkey and the lower Balkan peninsula, with the spread of sublineages of HG 9 whose age is one half of that of the entire HG 9. The age difference of these lineages, as approximate as it may be, suggests that they have not been carried to where they are at present by a single large-scale dispersal of early Neolithic farmers from the Levant (Ammerman & Cavalli-Sforza, 1984).

Finally, further complexity can be hypothesized for the overall HG 9 spread : the bimodal allele size distributions at DYS388 can be interpreted as the result of secondary population expansions. Also, the lower network 1.2 diversity

at the western edge of the area of high HG 9 frequencies, represented by Continental Italy, is compatible with later events including, as previously proposed (Cavalli-Sforza *et al.* 1993; Sokal *et al.* 1997), the Greek colonization.

We are grateful to Dr M. Jobling for his advice on a first draft of this manuscript. We thank Dr M. Lo Ponte for revising the paper. Work supported by PRIN MURST 1999 (L.T.), and C.N.R. grants 97.00712.PF36 (L.T.). Sampling campaigns were partially funded by NATO grant LST.CLG975057 to A.I.K., M.G.K. and A.N.

REFERENCES

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

Aris-Brosou, S. & Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Mol. Biol. Evol.* **13**, 494–504.

Barbujani, G. & Bertorelle, G. (2001). Genetics and the population history of Europe. *Proc. Natl. Acad. Sci. USA* **98**, 22–25.

Barbujani, G. (2000). Geographic patterns: How to identify them and why. *Hum. Biol.* **72**, 133–153.

Bertranpetit, J. (2000). Genome, diversity, and origins: the Y chromosome as a storyteller. *Proc. Natl. Acad. Sci. USA* **97**, 6927–6929.

Bianchi, N. O., Catanesi, C. I., Bailliet, G., Martinez-Marignac, V. L., Bravi, C. M., Vidal-Rioja, L. B., Herrera, R. J. & Lopez-Camelo, J. S. (1998). Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. *Am. J. Hum. Genet.* **63**, 1862–1871.

Caglia', A., Novelletto, A., Dobosz, M., Malaspina, P., Ciminelli, B. M. & Pascali, V. L. (1997). Y-chromosome STR loci in Sardinia and Continental Italy reveal islander-specific haplotypes. *Eur. J. Hum. Genet.* **5**, 288–292.

Casanova, M., Leroy, P., Boucekkine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G. & Siniscalco, M. (1985). A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. *Science* **230**, 1403–1406.

Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1993). Demic expansions and human evolution. *Science* **259**, 639–646.

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.

Forster, P., Kayser, M., Meyer, E., Roewer, L., Pfeiffer, H., Benkmann, H. & Brinkmann, B. (1998). Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. *Mol. Biol. Evol.* **15**, 1108–1114.

Forster, P., Rohl, A., Lunnemann, P., Brinkmann, C., Zerjal, T., Tyler-Smith, C. & Brinkmann, B. (2000). A short tandem repeat-based phylogeny for the human Y chromosome. *Am. J. Hum. Genet.* **67**, 182–196.

Foster, E. A., Jobling, M. A., Taylor, P. G., Donnelly, P., De Knijff, P., Mieremet, R., Zerjal, T. & Tyler-Smith, C. (1998). Jefferson fathered slave's last child. *Nature* **396**, 27–28.

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., Redd, A. J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, A. S., Oppenheim, A., Jobling, M. A., Ostrer, H. & Bonné-Tamir, B. (2000). Jewish and Middle-Eastern non-Jewish populations share a common pool of Y chromosome biallelic haplotypes. *Proc. Natl. Acad. Sci. USA* **97**, 6769–6774.

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.

Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., De Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R. & Hammer, M. F. (1999). Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**, 817–831.

Kayser, M., Roewer, L., Hedman, M., Henke, L., Henke, J., Brauer, S., Krüger, C., Krawczak, M., Nagy, M., Dobosz, T., Szibor, R., De Knijff, P., Stoneking, M. & Sajantila, A. (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* **66**, 1580–1588.

Kayser, M., Brauer, S., Weiss, G., Schifenhövel, W., Underhill, P. A. & Stoneking, M. (2001). Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**, 173–190.

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

Lell, J. & Wallace, D. C. (2000). The peopling of Europe from the maternal and paternal perspectives. *Am. J. Hum. Genet.* **67**, 1376–1381.

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., 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., 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.

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

Owens, K. & King, M. C. (1999). Genomic views of human history. *Science* **286**, 451–453.

Perez-Lezaun, A., Calafell, F., Comas, D., Mateu, E., Bosch, E., Martinez-Ariaz, R., Clarimon, J., Fiori, G., Luiselli, D., Facchini, F., Pettener, D. & Bertranpetit, J. (1999). Sex-specific migration patterns in central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* **65**, 208–219.

Quintana-Murci, L., Krausz, C., Zerjal, T., Sayar, S. H., Hammer, M. F., Mehdi, S. Q., Ayub, Q., Qamar, R., Mohyuddin, A., Radhakrishna, U., Jobling, M. A., Tyler-Smith, C. & McElreavey, K. (2001). Y-chromosome lineages trace diffusion of peoples and languages in Southwest Asia. *Am. J. Hum. Genet.* **68**, 537–542.

Reich, D. E. & Goldstein, D. B. (1998). Genetic evidence for a Paleolithic human population expansion in Africa. *Proc. Natl. Acad. Sci. USA* **95**, 8119–8123.

Renfrew, C., Forster, P. & Hurles, M. (2000). The past within us. *Nature Genet.* **26**, 253–254.

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.

Santos, F. R., Pandya, A., Tyler-Smith, C., Pena, S. D. J., Schanfield, M., Leonard, W. R., Osipova, L., Crawford, M. H. & Mitchell, R. J. (1999). The central Siberian origin for Native American Y chromosomes. *Am. J. Hum. Genet.* **64**, 619–628.

Semino, O., Passarino, G., Brega, A., Fellous, M. & Santachiara-Benerecetti, A. S. (1996). A view of the neolithic demic diffusion in Europe through two Y chromosome-specific markers. *Am. J. Hum. Genet.* **59**, 964–968.

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.

Shen, P., Wang, F., Underhill, P. A., Franco, C., Yang, W. H., Roxas, A., Sung, R., Lin, A. A., Hyman, R. W., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (2000). Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci. USA* **97**, 7354–7359.

Sokal, R. R., Oden, N. L., Rosenberg, M. S. & Digiovanni, D. (1997). The patterns of historical population movements in Europe and some of their genetic consequences. *Am. J. Hum. Biol.* **9**, 391–404.

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.

Stumpf, M. P. H. & Goldstein, D. B. (2001). Genealogical and evolutionary inference with the human Y chromosome. *Science* **291**, 1738–1742.

Sykes, B. (1999). Using genes to map population structure and origins. In *The human inheritance* (ed. B. Sykes), pp. 93–117. Oxford University Press.

Thomas, M. G., Bradman, N. & Flinn, H. M. (1999). High throughput analysis of 10 microsatellite and 11 diallelic polymorphims on the human Y chromosome. *Hum. Genet.* **105**, 577–581.

Thomas, M. G., Parfitt, T., Weiss, D. A., Weiss, A., Skorecki, K., Wilson, J. F., Le Roux, M., Bradman, N. & Goldstein, D. B. (2000). Y chromosomes traveling South: the Cohen modal haplotype and the origins of the Lemba-the 'black Jews of Southern Africa'. *Am. J. Hum. Genet.* **66**, 674–686.

Thomas, M. G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N. & Goldstein, D. B (1998). Origins of Old Testament priests. *Nature* **394**, 138–140.

Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. (2000). Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl. Acad. Sci. USA* **97**, 7360–7365.

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.

Wooding, S. & Ward, R. (1997). Phylogeography and Pleistocene evolution in the North American black bear. *Mol. Biol. Evol.* **14**, 1096–1105.