ORIGINAL INVESTIGATION

F. Di Giacomo · F. Luca · L. O. Popa · N. Akar · N. Anagnou · J. Banyko ·
R. Brdicka · G. Barbujani · F. Papola · G. Ciavarella · F. Cucci · L. Di Stasi ·
L. Gavrila · M. G. Kerimova · D. Kovatchev · A. I. Kozlov · A. Loutradis ·
V. Mandarino · C. Mammi' · E. N. Michalodimitrakis · G. Paoli · K. I. Pappa ·
G. Pedicini · L. Terrenato · S. Tofanelli · P. Malaspina · A. Novelletto

Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe

Abstract In order to attain a finer reconstruction of the peopling of southern and central-eastern Europe from the Levant, we determined the frequencies of eight lineages internal to the Y chromosomal haplogroup J, defined by biallelic markers, in 22 population samples obtained with a fine-grained sampling scheme. Our results partially resolve a major multifurcation of lineages within the haplogroup. Analyses of molecular variance show that the area covered by haplogroup J dispersal is characterized by a significant degree of molecular radiation for unique event polymorphisms within the haplogroup, with a higher incidence of the most derived sub-haplogroups on the northern Mediterranean coast, from Turkey westward; here, J diversity is not simply a subset of that present in the area in which this haplogroup first originated. Dating estimates, based on simple tandem repeat loci (STR) diversity within each lineage, confirmed the presence of a major population structuring at the time of spread of haplogroup J in Europe and a punctuation in the peopling of this continent in the post-Neolithic, compatible with the

F. Di Giacomo · L. Terrenato · P. Malaspina
Department of Biology, University "Tor Vergata",
Rome, Italy

F. Luca · V. Mandarino · A. Novelletto (✉)
Department of Cell Biology, University of Calabria,
Via P. Bucci,
87030 Rende, Italy
e-mail: novelletto@bio.uniroma2.it
Tel.: +39-0984-492933
Fax: +39-0984-492911

L. O. Popa · L. Gavrila
Genetics Department, University of Bucharest,
Bucharest, Rumania

N. Akar
Pediatrics Department, Ankara University,
Ankara, Turkey

N. Anagnou · E. N. Michalodimitrakis · K. I. Pappa
School of Medicine, University of Athens ,
Athens, Greece

J. Banyko
University of P. J. Safarik,
Kosice, Slovak Republic

R. Brdicka
Institute for Haematology and Blood Transfusion,
Prague, Czech Republic

G. Barbujani
Department of Biology, University of Ferrara,
Ferrara, Italy

F. Papola
Centro Regionale di Immunoematologia e Tipizzazione
Tissutale,
L'Aquila, Italy

G. Ciavarella
IRCCS Casa Sollievo della Sofferenza,
S. Giovanni Rotondo, Italy

F. Cucci
Az. Osp. Perrino,
Brindisi, Italy

L. Di Stasi
A.S.L. 1,
Paola, Italy

M. G. Kerimova
Department of Hygiene, Azerbaijan Medical University,
Baku, Azerbaijan

D. Kovatchev
Department of Biology, Medical University,
Varna, Bulgaria

A. I. Kozlov
Arct-An C Innovative Laboratory,
Moscow, Russian Federation

A. Loutradis
Ministry of Health Center for Thalassemia,
Athens, Greece

expansion of the Greek world. We also present here, for the first time, a novel method for comparative dating of lineages, free of assumptions of STR mutation rates.

Introduction

The extant distribution of Y chromosomal diversity is being increasingly used as a tool for reconstructing the peopling of the world by modern humans, at least from a male perspective (for reviews, see Underhill et al. 2001; Jobling and Tyler-Smith 2003). Major advancements in this field derive from (1) the discovery of numerous single nucleotide polymophisms (SNPs) and other polymorphisms at biallelic loci; (2) the possibility of investigating a further level of diversity determined by multiallelic simple tandem repeat loci (STR). The first set of markers (polymorphisms) has been used to reconstruct a robust phylogeny of the molecular types found today, based on the assumption of a monophyletic origin of the derived allele at each locus, to generate the so-called unique event polymorphism (UEP). The phylogeny is under continuous revision and has been given a unified nomenclature system (Y Chromosome Consortium 2002) to identify each internal UEP-defined lineage or haplogroup. Markers of the second set (STRs) are characterized by mutation rates far higher than in the first set and by a mutational pattern commonly leading to alleles equal in state. By virtue of these properties, STR markers accumulate variation within each haplogroup (de Knijff 2000). When the bulk of STR variation is subdivided according to the different haplogroups, a large part of the homoplasy between allelic states is resolved (Bosch et al. 1999; Forster et al. 2000), allowing us to reconstruct possible evolutionary paths (networks; Bandelt et al. 1999) with a manageable number of reticulations.

STR variation accumulates over time, and a variety of measures and methods have been devised to exploit this process and to infer the antiquity of the corresponding haplogroup (Bandelt et al. 1999; Goldstein et al. 1995; Stumpf and Goldstein 2001; Wilson and Balding 1998). While these methods represent genuine attempts to arrive at absolute dating estimates based solely on genetic data, they rely on a necessary parameter, i.e., the mutation rate at the loci under study, which is in turn the main determinant of the speed of the accumulation process. By virtue of the rare occurrence of each event, the measurement of mutation rate intrinsically suffers from large sampling errors. Large series of father-son transmissions have been examined (Heyer et al. 1997; Foster et al. 1998; Kayser et al. 2000) for a number of STR loci, resulting in estimates that have later been entered into dating studies (e.g. Thomas et al. 1998; Weale et al. 2001; Zerjal et al. 2003). With a few exceptions, the final results in terms of absolute dates are associated with large confidence intervals, and the above considerations prompt us to take them cautiously. A recent attempt to circumvent this problem has been proposed by Zhivotovsky et al. (2004) who have obtained an average mutation rate from population data rather than family data and used known foundation events as starting points for the production of the level of diversity observed today.

Y chromosome diversity across the European continent has revealed high levels of population structuring (Malaspina et al. 2000; Rosser et al. 2000; Semino et al. 2000a). However, different authors attribute the observed patterns of geographic variation to alternative scenarios for the peopling of the continent. In the first, the largest part of diversity is accounted for by a Paleolithic colonization, possibly re-established by Mesolithic repopulation from refugia after the last glacial maximum (LGM; Karafet et al. 1999; Semino et al. 2000a; Underhill et al. 2001; Wilson et al. 2001). In the second, a Neolithic demic diffusion from the Levant was triggered by the development of agricultural practices and introduced into the continent a large proportion of extant types which, despite their appearance, may have predated the migratory movement (Chikhi et al. 2002). Additional observations indicate that further testing of these inferential conclusions is required, e.g., the finding that the distributions of some haplogroups do not fit either model (Jobling and Tyler-Smith 2003), that drift or founder effects may obscure continent-wide distribution patterns (Di Giacomo et al. 2003), and that more recent population expansions and resettlements within the continent could result in spatial patterns resembling those predicted by one or the other of the above-mentioned models.

Testing each of the two models may benefit from the analysis of the appropriate haplogroup(s) in the geographic region(s) that are more informative for its/their dispersal. In this context, time estimates for each lineage are a useful piece of information to reconstruct the tempo of the dispersal and to distinguish between a continuous vs. a punctuated spreading process.

We have addressed the geographic distribution and internal molecular diversity of Y chromosomal haplogroup

G. Pedicini
Az. Osp. Rummo,
Benevento, Italy

C. Mammi′
A.S.L. BMM,
Reggio Calabria, Italy

G. Paoli · S. Tofanelli
Department of Ethology, Ecology and Evolution, University of Pisa,
Pisa, Italy

N. Anagnou · K. I. Pappa
Institute of Molecular Biology and Biotechnology,
Heraklion, Greece

E. N. Michalodimitrakis
University of Crete School of Medicine, Laboratory of Toxicology and Forensic Medicine,
Heraklion, Crete

Table 1 Absolute and percent (in *italics*) frequencies of haplogroup J sub-haplogroups in 22 population samples (*NA* not applicable)

Population sample (no. of subsamples)	Code	Sample size	Frequency of J	J*(xJ1, J2) p12f2	J1 p12f2, M267	J2*(xDYS413≤18, J2e) p12f2, M172	J2e p12f2, M172, M12	J2-(DYS413≤18)(xJ2a, J2f) p12f2, M172, DYS413	J2a p12f2, M172, DYS413, M47	J2f*(xJ2f1) p12f2, M172, DYS413, M67	J2f1 p12f2, M172, DYS413, M67, M92	Diversity within J
Albania (1)	AL	9					1					NA
			11.1				*11.1*					
Azerbaijan (2)	AZ	46			7			9		2		0.62±0.07
			39.1		*15.2*			*19.6*		*4.3*		
Belarus (1)	BE	39										NA
			0									
Bulgaria (1)	BU	39			2		2	3		1	1	0.86±0.09
			23.1		*5.1*		*5.1*	*7.7*		*2.6*	*2.6*	
Czech Republic (2)	CZ	94		2						1		0.67±0.31
			3.2	*2.1*						*1.1*		
Egypt (2)	EG	47			6	1		2		1	1	0.71±0.14
			23.4		*12.8*	*2.1*		*4.3*		*2.1*	*2.1*	
Crete (4)	CR	137		2	4		1	26		14	4	0.66±0.05
			37.2	*1.5*	*2.9*		*0.7*	*19.0*		*10.2*	*2.9*	
Greece (10)	GR	249		1	4	6	9	17		10	3	0.80±0.03
			20.0	*0.4*	*1.6*	*2.4*	*3.6*	*6.8*		*4.0*	*1.2*	
Iran (1)	IR	7						1				NA
			14.3					*14.3*				
Northern Italy (4)	NI	126			1		3	7		1	1	0.69±0.12
			10.4		*0.8*		*2.4*	*5.6*		*0.8*	*0.8*	
Southern Italy (14)	SI	407			16	1	10	39		17	16	0.76±0.02
			24.3		*3.9*	*0.2*	*2.5*	*9.6*		*4.2*	*3.9*	
Sardinia (2)	SA	51			3	1	1	2				0.81±0.13
			13.8		*5.9*	*2.0*	*2.0*	*3.9*				
Morocco (2)	MO	37			2							0.00±0.00
			5.4		*5.4*							
Oman (1)	OM	13		1	2			2				0.80±0.16
			38.5	*7.7*	*15.4*			*15.4*				
Romania (5)	RO	130			2			11		1	1	0.47±0.15
			11.6		*1.5*			*8.5*		*0.8*	*0.8*	
Russia (6)	RU	223			1			3				0.50±0.27
			1.7		*0.4*			*1.3*				
Slovak Rep. (1)	SR	23			1		1					1.00±0.50
			8.6		*4.3*		*4.3*					
Syria (1)	SY	50			9			5	1	1		0.62±0.10
			32.0		*18.0*			*10.0*	*2.0*	*2.0*		
Turkey (5)	TU	168			16	3	3	23	1	9		0.72±0.04
			32.8		*9.5*	*1.8*	*1.8*	*13.7*	*0.6*	*5.4*		
UAE (1)	UA	34			11			2				0.28±0.14
			38.3		*32.4*			*5.9*				
Ukraine (1)	UI	6										NA
			0.0									
UK (1)	UK	20			1							NA
			5.0		*5.0*							
Total		1955		6	88	12	31	152	2	58	27	
Typed for STRs		242		4	61	5	21	99	1	37	19	
% of typed for STRs				*66.7*	*69.3*	*41.7*	*67.8*	*65.1*	*50.0*	*63.8*	*70.4*	

J in the Middle East, central-eastern Mediterranean, and central-eastern Europe. It is generally agreed that this haplogroup was dispersed by the westward movement of people from the Middle East (Semino et al. 1996, 2000a; Quintana-Murci et al. 2001). Our data show a higher diversity of this haplogroup in areas reached in later phases of this process. Thus, the present-day distribution of haplogroup J cannot be explained by the expansion of a repertoire of types previously present in the area in which this haplogroup supposedly originated.

Materials and methods

Subjects

We studied an overall number of 1,955 males from Europe, west Asia, and north Africa, collected at 68 sampling locations. Many of the local samples represent a subset of those previously described (Malaspina et al. 2001; Di Giacomo et al. 2003). Local samples for which the typing of J sub-haplogroups could not be completed for all subjects according to the protocol described below were excluded in their entirety. However, data of microsatellite typings obtained from J chromosomes from these local samples were retained for dating analyses.

The local samples were pooled according to nationality, except for the two Mediterranean islands of Sardinia and Crete, for which a suitable sample size and more than one sampling location were available. In addition, northern Italy was kept separate from southern Italy in view of the genetic discontinuity first detected by Barbujani et al. (1990) and which we confirmed on the basis of Y chromosomal haplogroup distribution (Di Giacomo et al. 2003).

Overall, we obtained the 22 samples reported in Table 1. The above pooling strategy was pursued, when possible, to attain a fair representation of the Y chromosomal diversity in each of the examined nations, smoothing the effects of local vagaries in haplogroup frequencies attributable to the reduced population effective size for the Y chromosome.

Haplogroup nomenclature

Throughout this paper, we will reserve the term "haplogroup" for the entire J lineage, whereas internal lineages defined by derived states at additional markers will be referred to as "sub-haplogroups". The term haplotype will be used to indicate groups of chromosomes recognizable by variation at STR loci (de Knijff 2000; Hammer and Zegura 2002). For sub-haplogroups, we adhere to the nomenclature system proposed by the Y Chromosome Consortium (2002). For each sub-haplogroup, the allele states at the markers examined in this work are detailed in Table 1 and Fig. 1.

Fig. 1 Simplified phylogenetic tree of J sub-haplogroups based on the results reported here and those by Scozzari et al. (2001). Sub-haplogroups J2b and J2d, whose origin with respect to the DYS413≤18 mutation could not be resolved, are not shown. Other haplogroups, internal to J2e and J2f and reported by the Y Chromosome Consortium (2002), are not shown. The positions of the nodes of the tree are according to age estimates obtained with BATWING (Table 3, line 5)

DNA typings

Each subject was initially screened for UEP markers, which enabled the detection of the majority of haplogroups in the population of origin. The biallelic marker 12f2 was typed (Rosser et al. 2000) in all subjects who escaped this stage of detection. The sequential subtyping of haplogroup J carriers proceeded as follows: M172 was typed as described (Malaspina et al. 2000) in all 12f2-derived subjects. M267 was typed in all M172-ancestral subjects. The multirepeat deletion at DYS413 was typed (Malaspina et al. 1997) in all the M172-derived and most of M172-ancestral subjects. We had previously shown (Malaspina et al. 2001) that the derived state at the DYS413 marker is represented by a multirepeat deletion with alleles of ≤18 repeats, which are found only on M172-derived chromosomes.

M67 and M92 were initially searched in all M172-derived subjects. As M67-T (derived allele) was found only in DYS413-derived subjects, we completed the screening of this group to obtain the corresponding sub-haplogroup frequency in each population sample. M92 was typed in M67-derived subjects only. M47 was tested in all M172-derived/M67-ancestral individuals. All subjects who were M172-derived and without the DYS413 multirepeat deletion (ancestral) were screened for M12. M68 and M158 were typed in 163 and 172 subjects, respectively, with or without the multirepeat deletion at DYS413.

M67, M12, M92, and M267 were typed by amplification of the corresponding locus under the conditions and with the primers described by Underhill et al. (2000) or Cinnioglu et al. (2004), spotted on nylon membranes, and hybridized with the 32-P labeled ASO probes M67-A (ancestral) 5′-AAAAACAAATATAGAGG-3′, M67-T (derived) 5′-CCTCTATATATGTTTTT-3′ (hybridization and

washing temperature 42°C); M12-G (ancestral) 5′-CCCATCTCTACAAATA-3′, M12-T (derived) 5′-CCCA-TATCTCTACAAATA-3′ (44°C); M92-T (ancestral) 5′-CAAAAATATGTAGGCTC-3′, M92-C (derived) 5′-GAGCCTACGTATTTTTG-3′ (47°C); M267-T (ancestral) 5′-TGTAAATATACGTTCAG-3′, M267-G (derived) 5′-TGTAAATAGACGTTCAG-3′ (47°C).

M47, M68, and M158 were all amplified in a multiplex reaction under the conditions and with primers described by Paracchini et al. (2002). The polymerase chain reaction (PCR) products were spotted in replicate on nylon filters and hybridized with the 32-P labeled ASO probes: M47-G (ancestral) 5′-TTAACATAGAAGTTTAC-3′, M47-A (de-rived) 5′-TTAACATAAAAGTTTAC-3′ (41°C); M68-A (ancestral) 5′-GTAGGAAGAGTGGAAAG-3′, M68-G (derived) 5′-GTAGGAAGGGTGGAAAG-3′ (51°C); M158-G (ancestral) 5′-AAGCCCCACCTTGCCT-3′, M158-A (derived) 5′-AAGCCCCATCTTGCT-3′ (51°C).

The STR markers DYS19, DYS388, DYS390, DYS392, and DYS393 were analyzed in 247 subjects (Table 1, bottom lines, and Appendix) by fluorescent PCR assays and detection by a ABI310 DNA sequencer. DYS19 was typed in a single PCR, whereas DYS388, DYS390, DYS392, and DYS393 were typed in a separate multiplex reaction as described by Thomas et al. (1999). A subset of the same subjects had been previously typed for the four latter markers with a radioactive PCR assay and polyacrylamide gel fractionation. Cross-checking of the two methods always gave consistent results.

Data analysis

Diversity indexes, AMOVA computations, and the matri-ces of pairwise Fst and Phist values were obtained with the Arlequin 2.000 package (Schneider et al. 2000). Spatial analysis of molecular variance was performed with the program SAMOVA 1.0 (Dupanloup et al. 2002). This program implements an approach to search for groups of populations that are geographically homogeneous and maximally differentiated from each other, also leading to the identification of genetic barriers between groups. The method extracts the populations contributing to each group without a-priori assumptions. We applied AMOVA and SAMOVA only to haplogroup J data, solely to evaluate its internal diversity across populations. Thus, the fixation indexes reported here are not comparable with those that are usually reported in the literature and that include all chromosomes within a population sample. The entries in the matrix of distances among sub-haplogroups consisted of the number of UEP events separating them (the DYS413 multirepeat deletion was considered as a UEP).

Correspondence analysis was performed with SPSS v.6.1 on the table of sub-haplogroup frequencies. Multi-dimensional scaling (MDS) was performed with SPSS, by importing the matrices of either pairwise Fst or Phist generated by Arlequin (Schneider et al. 2000).

Dating estimates of haplogroup antiquity based on microsatellite diversity were obtained by three different

methods. (1) The program YMRCA (Stumpf and Gold-stein 2001) was used on the STR haplotypes within each sub-haplogroup, with mutation rates of 0.0021, 0.0020, 0.0054, 0.0023, and 0.0011 for DYS19, DYS388, DYS390, DYS392, and DYS393, respectively. These values result from direct observations of mutational events (Weale et al. 2001 and references therein, except for DYS388) and also represent the means of the gamma distributions used with BATWING. (2) The program BATWING (Wilson and Balding 1998) was used under two different conditions: in the first, each J sub-haplogroup was treated independently, whereas in the second, the phylogenetic succession of UEP mutations was considered, by using the "infsites" option. In both cases, priors for mutation rates at DYS19, DYS390, DYS392, and DYS393 were the same used by Weale et al. (2001), whereas for DYS388 we used GAMMA(2,1000). In contrast to Weale et al. (2001), we used more stringent priors for the initial population size, population growth rate, and start of growth (GAMMA[1, 0.001], UNIFORM [0.03, 0.05], and UNIFORM[0.10, 0.20], respectively), to exclude unrealistically large values especially for the growth rate (Ammerman and Cavalli-Sforza 1984). Each BATWING run consisted of 1,000 Markovian steps, after a warmup of 200. Increasing the run length to 5,000 and 10,000 did not change the results sensibly. (3) A third method was implemented to obtain estimates of the timing of the appearance of sub-haplogroups, relative to the age of the entire J haplogroup. In this method, no assumption is made regarding mutation rates at the microsatellites, but each locus is assumed to have the same mutation rate across sub-haplogroups and to follow a strict stepwise mutation model (SMM). Under this model, the measure ASD or average squared distance (ASD between STR alleles of a haplotype and the alleles found on the ancestral haplotype) accumulates linearly with time according to the equation $ASD = \mu \sigma^2 \tau$ (see Stumpf and Goldstein 2001 for a description of the theory and criteria to identify ancestral haplotypes), where τ is the genealogical branch length separating alleles within a sub-haplogroup from the most recent common ancestor of the sample, μ the mutation rate, and σ^2 the variance in mutational step (=1 in the SMM). With five STR loci typed in seven sub-haplogroups (J2a is ignored here), five regressions can be computed for ASD as the dependent variable, which must share the same set of values for the independent variable (τ). Within a well-defined UEP phylogeny, these latter estimate the antiquity of each node of the tree, which must be the same for all STR loci. The slope of each regression in turn estimates the mutation rate at each locus (in ASD units/arbitrary time unit), thus allowing a comparison across loci.

The procedure thus aims at finding the slopes of the regression lines (s_i, with $i = 1 \dots n$ of loci) that best fit the linearity of accumulation of ASD with time, conditional on values that are fixed for all STR loci within a sub-haplogroup. This is performed by the least-squares method.

Briefly, ASD is calculated for each STR locus and each sub-haplogroup as described (Stumpf and Goldstein 2001). The age of the most ancestral J sub-haplogroup is given the arbitrary value of one (zero being the present). Locus-specific slopes (s_i), total sum of squares ($SSQtot_i$) and sum of squares about regression ($SSQab_i$) are calculated for 250,796 combinations of x_j values (with $j=1...n$ of sub-haplogroups), representing the positions of the nodes of the phylogenetic tree of the sub-haplogroups (with topology univocally defined by UEPs) at discrete steps of 0.083 (1/12 of the total haplogroup age) as:

$$s_i = \frac{\sum_{j=1}^{n} A_{ij} x_j}{\sum_{j=1}^{n} x_j^2}$$

where A_{ij} is the ASD for the ith locus in the jth sub-haplogroup and n the number of sub-haplogroups;

$$SSQtot_i = \sum_{j=1}^{n} (A_{ij})^2 - \frac{\left(\sum_{j=1}^{n} A_{ij}\right)^2}{n}$$

$$SSQab_i = \sum_{j=1}^{n} \left[A_{ij} - E(A_{ij})\right]^2$$

where

$$E(A_{ij}) = s_i x_j$$

The set of x_j values producing the minimum cumulative SSQ about regressions is chosen as the best-fitting. The statistical significance of the five regressions is evaluated by an F ratio between the variance explained by the regressions (obtained by subtracting the sum of $SSQab_i$ from the sum of $SSQtot_i$) and the variance about regressions with 5/30 degree of freedom (df). Reliable intervals for the estimate are given by the range, mean, and standard deviation of values obtained in the significant ($P<0.01$) solutions.

Two variants of this method can be used, in which SSQs are calculated with or without giving each sub-haplogroup a weight equal to the number of observations, respectively. The entire procedure has been implemented in SPSS syntax language and is available on request.

Results

Phylogenetic relationships among sub-haplogroups

Our results expand those obtained previously for M172, DYS413, and M12 (Malaspina et al. 2001; Scozzari et al. 2001). Indeed, all chromosomes carrying the derived state at M67 (defining J2f) were found also to be derived at DYS413. In addition, the two chromosomes carrying the derived state at M47 (defining J2a) were derived at DYS413. Thus, both J2a and J2f are internal to J2-

Fig. 2 Plot of the frequencies of haplogroup J sub-haplogroups in the 22 populations screened (codes as in Table 1). Each *bar* is based on the geographic area it represents and is proportional to the absolute number of haplogroup J observations. This number and the overall sample size are reported for each *bar*
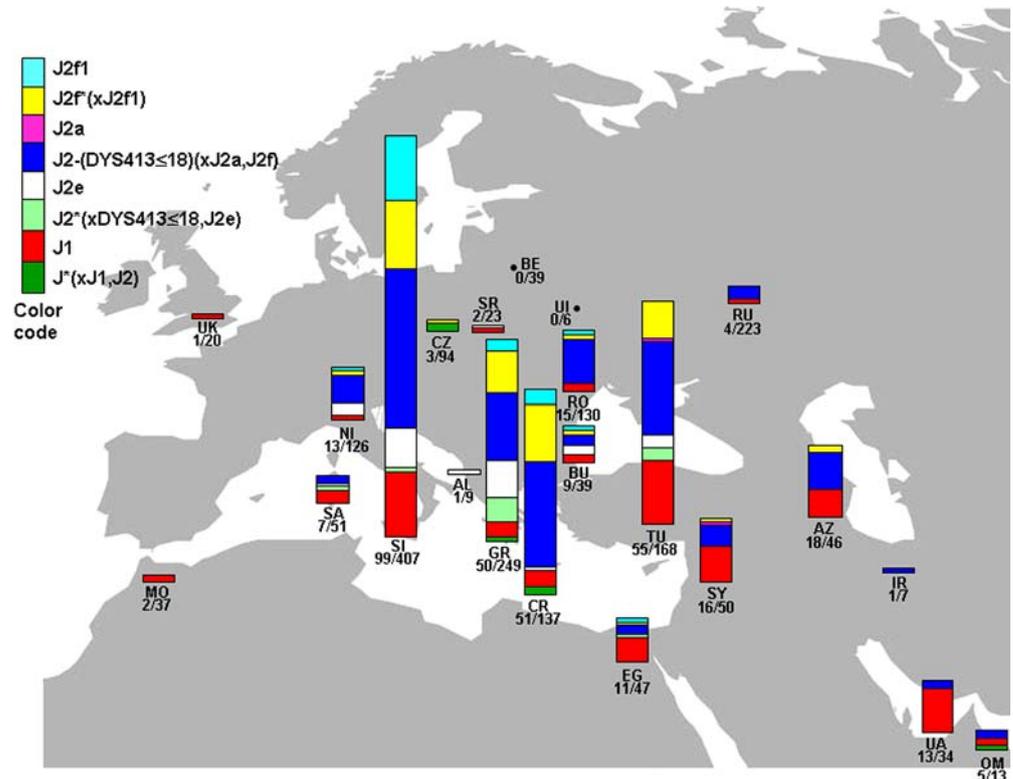
(DYS413≤18). The phylogenetic relationships of the eight sub-haplogroups that could be resolved here are shown in Fig. 1. Uncertainty remains regarding the position of sub-haplogroups J2b and J2d with respect to DYS413, as we did not find any subjects carrying the derived state at the corresponding markers M68 and M158. Generally, these results refine those by the Y Chromosome Consortium (2002), Jobling and Tyler-Smith (2003), and Cinnioglu et al. (2004), as they resolve the order of J2e with respect to the pair J2a-J2f, thereby reducing the multifurcation downstream to M172.

Affinity of populations for haplogroup J

The frequencies of the entire haplogroup J in the 22 samples are highly heterogeneous (χ^2=175, df=21, P<0.001; Fig. 2, Table 1). Frequencies greater than 0.30 are observed in Crete and Turkey and in the other middle-eastern samples (Azerbaijan, Oman, Syria, and UAE), with the exception of the few Iranians. Frequencies in the 0.20–0.30 range are present in the south-west (Egypt) and west (Greece, Bulgaria, and southern Italy) with respect to this group of samples. Albania, Sardinia, and northern Italy have frequencies in the 0.10–0.20 range. The frequencies fall sharply in northern and central Europe, north of the Carpathians.

Heterogeneity among the Russian samples confirms this trend. Indeed, three local samples at the western (Yaroslav, n=39), northern (Republic of Udmurtia, n=40), and eastern (Republic of Bashkiria, n=39) edges of our sampling area show null frequencies, the observations of haplogroup J being concentrated in two samples from the southern Urals. Moreover, in northern Africa, a drastic change of haplogroup J frequency is seen from the eastern to the western edge (Egypt and Morocco, respectively).

The diversity index for sub-haplogroups is highest in the few Omani and in four samples of southern Europe, i.e. Bulgaria, continental Greece, southern Italy, and Sardinia (0.86–0.76). It is intermediate (0.62–0.72) in Turkey and the Middle East, with a minimum in the UAE (0.28).

We assayed the affinity among population samples for haplogroup J by considering the frequency of eight sub-haplogroups. In order to use reliable sub-haplogroup frequencies, only samples with ten or more observations of J were included. These represent 1,394 subjects (71% of the total), 341 of whom carry a J sub-haplogroup (91% of all carriers of J in the study). Heterogeneity across these samples is highly significant (χ^2=149, df=63, P<0.001). AMOVA performed considering sub-haplogroups without their molecular distances also confirms this heterogeneity, with an overall Fst value of 0.069 (P<0.0001), and pairwise values as high as 0.5. Correspondence analysis and MDS produce highly consistent results (not shown). UAE, Egypt, and Syria are widely separated from the rest of samples on dimension 1, Azerbaijan and Turkey occupy an intermediate position, and the remaining samples all cluster together. This cluster is resolved on dimension 2, with Romania at one extreme, Crete, southern Italy and northern Italy in the middle, and continental Greece at the opposite extreme.

When the mutational steps separating sub-haplogroups are considered, additional heterogeneity is highlighted. The fixation index in AMOVA increases (Table 2, first line), and a clearer separation of samples upon MDS is apparent (Fig. 3). The four samples with positive values on dimension 1 are those with the highest relative frequencies of J1. In the three samples clustering on top left, J2-(DYS413≤18)(xJ2f, J2a) is the most common sub-haplogroup within J. The three samples at bottom left have high relative frequencies of both J2f(xJ2f1) and J2f1. The above similarities are visually apparent in Fig. 2; they are geographically consistent and can be somewhat extended to samples with less than ten haplogroup J observations. For example, the results in Omani show results are in agreement with the high prevalence of J1 and J2-(DYS413≤18)(×J2f, J2a) in the Middle East, and Moroccans exhibit the high prevalence of J1 in northern Africa.

Table 2 SAMOVA obtained on the ten population samples with ten or more observations of haplogroup J

No. of clusters	Fixation indexes	Populations in cluster						
		1	2	3	4	5	6	7
1[a]	Fst=0.09**	All	–	–	–	–	–	–
2	Fsc=0.06** Fct=0.31	UAE	Others	–	–	–	–	–
3	Fsc=0.05** Fct=0.22*	UAE	Egypt	Others	–	–	–	–
4	Fsc=0.03* Fct=0.16*	UAE	Egypt, Syria	N. Italy	Others	–	–	–
5	Fsc=0.03* Fct=0.16*	UAE	Egypt	Syria	Azerbaijan	Others	–	–
6	Fsc=0.04** Fct=0.12*	UAE	Egypt	Syria	N. Italy	Romania	Others	–
7	Fsc=0.00 Fct=0.13**	UAE	Egypt, Syria	Azerbaijan, Turkey	Cont. Greece, N. Italy	Crete	Romania	S. Italy

*P<0.05
**P<0.001
[a]This line reports the Fst obtained by ordinary AMOVA with molecular distances among sub-haplogroups
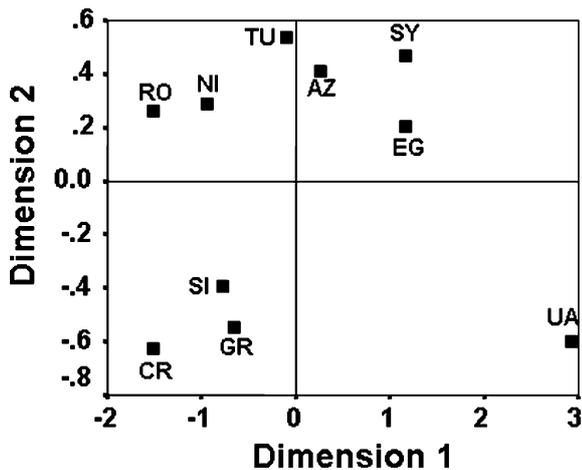
Fig. 3 Plot of the ten population samples with more than ten haplogroup J observations obtained by multidimensional scaling on the matrix of pairwise Phist values (codes as in Table 1)

The search for genetic discontinuities was performed by exploring the partitioning of the ten samples with ten or more J observations in two to seven clusters. SAMOVA recognized the middle-eastern samples as those contributing high Fct values. With two to six clusters, the Turkish, Greek, and southern Italian samples are all grouped together, indicating a relative homogeneity between them, despite significant Fsc values. Only, when seven clusters are used, Fsc drops to an insignificant value, and the above grouping is disrupted.

All of the above analyses show that the area here investigated is characterized by a detectable degree of molecular radiation for UEPs within haplogroup J, with a higher incidence of the most derived sub-haplogroups on the northern Mediterranean coast, from Turkey westward. The overlay of molecular radiation onto geographic dispersal in determining the J diversity is particularly evident in the three central Mediterranean locations of continental Greece, Crete, and southern Italy. These appear to share a highly similar J pool, which is characterized by the maximum internal diversity and which distinguishes them from the rest of the sampled locations.

Dating

We used three different methods to date the nodes of the phylogenetic tree showed in Fig. 1. Methods that took into account each sub-haplogroup separately (Table 3, lines 1–4) produced fluctuating results. YMRCA produced estimates 1.5-fold to 3-fold lower than BATWING, despite our using the same mutation rates at the five STR loci.

Table 3 Estimates of age of haplogroup J, J sub-haplogroups, and mutation rate at five STR loci, obtained with three independent methods (*st. dev.* standard deviation, *NA* not applicable)

	Haplogroup or sub-haplogroup age							
	J*(xJ1, J2)	J1	J2*(xDYS413≤18, J2e)	J2e	J2-(DYS413≤18) (xJ2a, J2f)	J2a	J2f*(xJ2f1)	J2f1
Derived state at locus	p12f2	p12f2, M267	p12f2, M172	p12f2, M172, M12	p12f2, M172, DYS413	p12f2, M172, DYS413, M47	p12f2, M172, DYS413, M67	p12f2, M172, DYS413, M67, M92
Estimation method								
1 YMRCA[a]	19,300	8,400	3,600	3,000	6,900	NA	4,700	5,000
2 st. dev.[b]	10,260	12,600	3,330	2,460	4,620		4,410	7,260
3 BATWING[a] (subhaplogroups treated separately)	32,200	18,000	8,600	11,500	19,900	NA	10,900	7,300
4 5th–95th centile[b]	11,010–69,090	7,590–33,990	6,720–21,420	4,080–24,300	7,290–44,880		9,060–25,740	2,580–14,550
5 BATWING[a] (subhaplogroups and UEPs)	22,200	8,300–11,300	9,500–18,000	3,500–5,800	7,300–9,100	0–840	3,800–4,700	2,300–2,900
5th–95th centile[b]	11,850–40,170	4,260–23,070	5,760–32,490	2,010–10,290	5,130–14,640	0–2,160	2,430–7,680	1,650–4,200
7 ASD-time linearity[c] (optimal solution)	1	0.50	0.33	0.17	0.33	NA	0.25	0.08
8 Reliable interval[d] (range)		0.17–1.0	0.17–1.0	0.0 –0.83	0.17–1.0		0.0 –0.92	0.0–0.67
9 Reliable interval[d] (mean ± SD)		0.78±0.18	0.77±0.18	0.31±0.21	0.57±0.16		0.40±0.16	0.19±0.15

	STR locus mutation rate				
	DYS19	DYS388	DYS390	DYS392	DYS393
10 BATWING[e] (subhaplogroups and UEPs)	3.8 E−03	4.9 E−03	4.4 E−03	0.73 E−03	1.6 E−03
11 5th–95th centile	2.6–5.2 E−03	3.4–6.7 E−03	3.1–6.0 E−03	0.36–1.2 E−03	0.87–2.5 E−03
12 ASD-time linearity[f] (optimal solution)	1.32	2.91	1.83	0.17	0.67
13 Reliable interval[d] (range)	0.59–2.49	1.12–4.55	0.69–4.15	0.05–0.30	0.26–1.31
14 Reliable interval[d] (mean ± SD)	0.86±0.22	1.86±0.46	1.12±0.37	0.10±0.03	0.43±0.12

[a]Age in years rounded to the nearest hundred, assuming 30 years/generation (Zerjal et al. 2002, 2003)
[b]Age as above, rounded to the nearest tenth
[c]Age as fraction of the antiquity of the entire haplogroup
[d]Obtained from 56,496 significant (P<0.01) solutions
[e]Mutation rate in events/gamete/generation
[f]Mutation rate as slope of the ASD-time linear regression

Moreover, with both methods, the values for sub-haplogroup J2*(xDYS413≤18, J2e) is lower than that for sub-haplogroup J2-(DYS413≤18)(xJ2a, J2f), which carries an additional UEP mutation and cannot be older. The estimates for the former are clearly severely affected by the low number of observations (5).

When UEP phylogeny is taken into account (Table 3, lines 5–6), BATWING returns figures with narrower confidence intervals. With the exception of J2* (xDYS413≤18, J2e), the ages of all sub-haplogroups are shortened by about a factor two. J2e and J2f*(xJ2f1) are associated with similar estimates.

The method based on the linear accumulation of ASD with time, here used for the first time, can only be partly compared with the previous methods. Indeed, it returns the age of each node in terms of the fraction of the age of the entire haplogroup. In order to render the contribution of rare sub-haplogroups appropriately, the estimation of ASD-time regression coefficients and of explained and residual variances was performed by giving each sub-haplogroup a weight equal to the number of observations (Table 1, last two rows). Indeed, we observed that the haplogroup J2*(xDYS413≤18, J2e) and the paragroup J* (xJ1, J2) did not display a modal STR haplotype (see Appendix), thus making the identification of the ancestral haplotype and the calculation of ASD uncertain.

The best-fitting solution for the position of tree nodes (Table 3, line 7) produced an F-ratio between the overall explained and residual variances of 6.18 ($P=4.7\times10^{-4}$). The ranges of values represented in the 56,496 significant solutions (Table 3, line 8) are inflated by a few outlying figures. The means and standard deviations of the same values exclude an age for the two deepest branches, J1 and J2, below 42%, and an age for the most derived J2f1 branch above 49% of the entire haplogroup.

The proportions obtained with the optimization of the ASD-time linearity largely overlap those obtained with BATWING conditional on the UEP phylogeny (Table 3, lines 5 vs. 7). Both methods agree in showing an age of J1 at approximately one half of the age of J, an age of the J2 branch between one third and two thirds of that of J, an age for J2-(DYS413≤18)(xJ2a, J2f) of less than one half of that of J, similar ages for J2e and J2f*(xJ2f1) and, finally, an age for J2f1 of roughly 10% of that of the entire haplogroup.

The consistency of the two methods can be reinforced by considering the relative magnitude of the mutation rates obtained at the end of the two entirely independent estimation procedures. For BATWING, these represent the result of modification of the priors during the Markov chain process, whereas in the ASD-Time method, they represent the optimal solutions for the slopes of ASD-time regressions.

The results reported in Table 3 (lines 10–11) show that the Markovian process tends to reduce the mutation rate at DYS390 (from 5.4 to 4.4×10^{-3}) and DYS392 (from 2.3 to 0.73×10^{-3}) and to increase that at DYS388 (from 2.0 to 4.9×10^{-3}). The ASD-time method, without any prior assumption regarding mutation rates, actually catches these differences, assigning the highest rate to DYS388, followed by DYS390 and DYS19. DYS393 and DYS392 are assigned a rate reduced to 1/4 and less than 1/10 of DYS388, respectively (Table 3, lines 12–14).

Taken together, these results show that BATWING and the ASD-time method both respond in the same way to the different rate of mutation at the five STR loci, thus supporting each other. In this context, the consistencies in the estimates of antiquity of the J sub-haplogroups become more solid.

Discussion

Haplogroup J has been considered to represent a signature of the Neolithic demic diffusion associated with the spread of agriculture (Semino et al. 1996). We provide new data on the molecular radiation within this haplogroup and population data giving insights into the modes of spread of this haplogroup.

Phylogenesis

Haplogroup J can be subdivided into two major clades, viz., J1 and J2, characterized by the markers M267 and M172, respectively, plus the paragroup J*(xJ1, J2). Weale et al. (2003) have recommended caution in drawing phylogeographic inferences based on paragroups; nevertheless, the analysis of STR loci on these chromosomes is promising for understanding the accumulation of variation in the entire haplogroup (see below).

Although several markers have been described within J1, the overall UEP diversity within this clade and its informativeness for phylogeography have been reported as being low (Cinnioglu et al. 2004; Semino et al. 2004) and have not been extensively sought here.

Within J2, the analysis of a multirepeat deletion in the dinucleotide STR locus DYS413 has allowed us to resolve a major multifurcation of six independent lineages (Y Chromosome Consortium 2002), recently increased to eight (Cinnioglu et al. 2004). This additional mutational step within J2 enhances the possibility of performing phylogeographic studies of the entire J2 sub-haplogroup. In particular, we have found 12 subjects carrying the derived allele M172(G), but not the derived state at either M12 or DYS413 (Fig. 1). Of the eight markers immediately downstream of M172, two (M47, M67) are also downstream to DYS413≤18, whereas five could not be assigned because they were either monomorphic (M68 and M158) or not tested (M137, M330, and M339). A final sub-haplogroup, defined by M12 (J2e), had been previously shown (Scozzari et al. 2001) to carry the ancestral state at DYS413.

Population data

Our data on the overall occurrence of the entire haplogroup display an area of high frequencies (>20%) stretching from the Middle East to the central Mediterranean.

A review of the frequency data concerning Europe, the Caucasus, Iran, Iraq, and northern Africa reveals that, in the Mediterranean, this haplogroup is mainly confined to coastal areas (Al-Zahery et al. 2003; Barac et al. 2003; Behar et al. 2003; Bosch et al. 2001; Brion et al. 2003, 2004; Capelli et al. 2003; Francalacci et al. 2003; Manni et al. 2002; Nasizde et al. 2003; Quintana-Murci et al. 2001; Rosser et al. 2000; Semino et al. 2000a, 2000b, 2004; Wells et al. 2001).

The high frequencies in Turkey, in local Jewish and non-Jewish populations (Hammer et al. 2000; Nebel et al. 2001; Behar et al. 2003), and in the Caucasus identify the Middle East and the east Mediterranean as the focal area for the westward dispersal of the haplogroup. However, both the data reported here and in the literature agree in showing that this haplogroup did not leave a strong signature in the peoples of the northern Balkans and central Europe, this being the most likely route for the entry of agriculturalists in the European continent north of the Alps, under the demic diffusion model. Instead, the raw frequency data from within the Iberian, Italian, and Balkan peninsulas are more in line with alternative routes of westward spread, possibly maritime.

Internal J diversity and dating

We observe the highest UEP diversity in Turkey, Egypt, and three locations in southern Europe. The two most derived sub-haplogroups typed in this survey (J2f1 and J2e) have only been found in Turkey and locations west of it, boosting the UEP internal diversity. The sub-haplogroup distribution that we have observed in Turkey is similar to that reported by Cinnioglu et al. (2004; contingency χ^2 not significant). While we did not detect J2f1 in Turkey, Cinnioglu et al. (2004) report a frequency of 7.8%, with most of the observations (11/14) in central Anatolia, Istanbul, and European Turkey, three areas not represented in our sampling.

The diversity within J2 is lower in the Middle East (0.43 ±0.11) compared with both Turkey (0.60±0.07) and the European locations (0.67±0.02). In conclusion, the UEP diversity of J in Turkey and southern Europe does not seem to be a simple subset of that present in the area in which this haplogroup first originated. This finding, also confirmed by the data by Semino et al. (2004), points to Turkey and the Aegean as a relevant source for the J diversity observed throughout Europe.

STR loci haplotyping (see Appendix) provides additional evidence for J dispersal. The antiquity of J*(xJ1, J2) is denoted by the lack of a central haplotype in the network, by the shortest DYS388 alleles (12–14), and by the spread of DYS390 alleles over four repeat units (21–25).

Within sub-haplogroup J1, the putative ancestral haplotype is 14-16-23-11-12 (for loci DYS19-DYS388-DYS390-DYS392-DYS393), found in 17 out of 61 chromosomes. This haplotype is compatible with the major haplotype found in the middle-eastern J pool, which in turn contains the so-called Cohen modal haplotype (Thomas et al. 1998). Within J1, the marker DYS388 identifies two interesting subgroups. First, chromosomes with DYS388(17) have been linked with the expansion of Arabian tribes in southern Levant and northern Africa (Nebel et al. 2002). Our data agree with this scenario, as out of 13 instances of DYS388(17), ten are from Oman, UAE, and Egypt. Second, chromosomes with DYS388 (13) have been reported by Cinnioglu et al. (2004) with no DYS388(14) intermediate. We have found eight chromosomes of the former type, three of which are from northern Turkey and Azerbaijan, but we have also observed the DYS388(14) intermediate in southern Italy. In addition, this allele is also present on some M172(G) chromosomes. Thus, the geographic clustering observed in Turkey probably represents more a drift or a founder effect than a deviation from the stepwise mutational model.

Within J2-(DYS413≤18)(xJ2a, J2f), two major STR haplotypes are found. The first (14-15-23-11-12) is the most common (16/99) and is compatible with the haplotype mostly represented in the M172(G) chromosomes reported by Nebel et al. (2001). The second (15-15-23-11-12) is found in 12/99 chromosomes but is considered ancestral because of the higher number of one- and two-step neighbors within this sub-haplogroup STR network. This bimodality should prompt the search for additional UEP variation within this sub-haplogroup to improve its phylogeographic informativeness.

Chromosomes J2f*(xJ2f1) are characterized by haplotype 14-15-23-11-12 (15/37), whereas chromosomes J2f1 invariably carry the DYS390(22) allele, mostly as 14-15-22-11-12. This discrepancy is similar to that observed by Cinnioglu et al. (2004). Finally, the J2e chromosomes are characterized by the allele DYS390(24), found in 13 out of 21 cases.

When variation at each locus is analyzed across sub-haplogroups, DYS388 turns out to be the most variable, with ASD figures >0.35 in six out of seven cases; on the other hand, DYS392 is by far the less variable, with ASD values <0.15 in six out of seven cases. These results show a clear structuring of the total STR variation.

Our absolute dating estimates are based on three different methods, only one of which makes full exploitation of the data by considering UEP and STR information at the same time (Table 3, lines 5–6). The age returned for the entire J clade obtained with BATWING and its confidence interval fall within the range reported in previous works involving either different methods or the same method but with different priors (39.6–10.5 thousands of years ago; Hammer et al. 2000; Nebel et al. 2001; Quintana-Murci et al. 2001; Cinnioglu et al. 2004). It is to be noted that all these figures have large confidence

intervals and were obtained from local samples, whereas our repertoire of J chromosome types is derived from an area more representative of the entire haplogroup home-range.

As far as sub-haplogroups are concerned, the ages of J1 and J2 in previous publications are almost indistinguishable from the entire J (Nebel et al. 2001; Cinnioglu et al. 2004), whereas our results suggest ages corresponding to 63%–33%. We have also resolved the dating for J2f* (xJ2f1) and J2f1. The difference in their antiquity is not apparent in the data of Cinnioglu et al. (2004), as they pooled all J2f-M67 chromosomes, although these are well differentiated, at least for DYS390 (see above).

Our dating estimates for five J sub-haplogroups can be compared with those of Semino et al. (2004). Although the confidence intervals for absolute dates largely overlap, the results of Semino et al. (2004) tend to be older. This discrepancy can be partially attributed to the average mutation rate used, which is 3.7-fold lower than the average rate used in our BATWING runs. Their figure was obtained (Zhivotovsky et al. 2004) from lineages and populations carrying shorter alleles at DYS388 and DYS390, compared with haplogroup J. These two loci may have contributed substantially to the accumulation of variation on J, in view of the relationship between repeat count and variance (Kayser et al. 2004). When the absolute estimates by Semino et al. (2004) are expressed as percentages of the age of the entire haplogroup, their results closely match those reported here.

Conclusions

Our estimates are in agreement with the appearance of J1 and J2 in the Levant at the time of the Neolithic agriculture revolution. Implicitly, this figure makes them of little help in identifying population splitting that may have accompanied the westward dispersal of the entire haplogroup.

Our data and those by Semino et al. (2004) show that J2f1 is predominantly found in the northern Mediterranean, from Turkey westward. In particular, our estimates for this latter sub-haplogroup are barely compatible with its presence among the early Levantine agriculturalists. The coalescence of J2f1 well after the beginning of the Neolithic revolution thus identifies a major population structuring already present at the time of the spreading of haplogroup J in southern Europe and central Mediterranean, thus differentiating the Aegean area from the Middle East. We favor the emergence of J2f1 in the Aegean area, possibly during the population expansion phase also detected by Malaspina et al. (2001) and coincident with the expansion of the Greek world to the European coast of the Black sea. This scenario would agree with the clustering of J2f1 chromosomes in north–west Turkey (Cinnioglu et al. 2004).

We tried to parallel the BATWING analysis with an independent method, free of assumptions concerning STR mutation rates. Indeed, inter-locus heterogeneity for mutation rates is responsible for a large part of variation in dating estimates (Stumpf and Goldstein 2001). This method provides a formal approach to the concept of "comparative dating" (Malaspina et al. 2001) by exploiting the UEP–STR data set to its fullest extent. Compared with the dating of a single haplogroup or sub-haplogroup, as implemented in YMRCA, it seems more powerful in detecting the "signal" of increasing diversity with time over the "noise" of the poor correlation between ASD and time observed for some loci. This is achieved by including, in the analysis, information concerning UEP phylogeny, which provides a unique solution for the order of nodes in the tree to be analysed. As the number of UEP-defined lineages in the Y phylogeny is steadily increasing, this method will become of wider applicability.

Our method has been able to rank the five loci used here in the same order as BATWING according to their mutation rates, indicating that the two programs can detect the same basic trend. In particular, the remarkable evolutionary stability of the marker DYS392 has emerged, in agreement with its constancy in many sets of Y chromosomes typed (Bosch et al. 1999; Forster et al. 2000; Thomas et al. 2000; Nebel et al. 2001; Weale et al. 2001; Cinnioglu et al. 2004). This enables us to review, on a microevolutionary basis, the estimates based on direct observation of father–son transmissions (Heyer et al. 1997; Bianchi et al. 1998; Kayser et al. 2000), a result that may be relevant in forensic applications.

The ASD vs. time method returns relative timings for the branching within the J tree in close agreement with those of BATWING. In particular, it shows a burst of UEP radiation in the second half-life of the entire J haplogroup, followed by the origin of J2f1 in recent times. Unless assuming the origin of J at >50–100 thousand of years ago, this is again incompatible with an J2f1 origin in the Levant at the time of the rise of agriculture.

In summary, our data are in agreement with a major discontinuity for the peopling of southern Europe. Here, haplogroup J constitutes not only the signature of a single wave-of-advance from the Levant but, to a greater extent, also of the expansion of the Greek world, with an accompanying novel quota of genetic variation produced during its demographic growth. In the analysis by Cavalli-Sforza et al. (1994), the two peopling contributions can be distinguished, as they are caught in the first and the fourth principal component, respectively, but the relevance of the latter may have been underestimated. The two processes, widely spaced in time, are associated with dramatically different travel technologies. This implies that, in the central and west Mediterranean, the entry of J chromosomes may have occurred mainly by sea, i.e., in the south–east of both Spain and Italy.

Appendix

Haplotype information for the 247 subjects typed for the five STR markers

Sub-haplogroup	DYS19	DYS388	DYS390	DYS392	DYS393	Frequency
J*(xJ1, J2)						
	13	13	25	11	13	1
	14	14	23	11	13	1
	15	12	21	11	13	1
	15	12	22	10	15	1
J1						
	13	15	24	11	13	1
	13	15	24	11	12	1
	13	16	23	11	12	1
	14	13	23	11	13	6
	14	13	23	11	12	2
	14	14	24	11	12	1
	14	15	22	11	12	2
	14	15	23	11	12	1
	14	15	24	11	12	2
	14	16	23	11	12	17
	14	16	23	12	12	2
	14	16	24	11	13	1
	14	16	24	11	12	1
	14	17	22	12	12	1
	14	17	23	9	12	1
	14	17	23	11	12	10
	14	17	24	11	12	1
	14	18	23	11	12	2
	15	12	22	11	12	1
	15	16	23	11	12	3
	15	16	23	11	14	1
	15	17	23	11	12	1
	15	18	23	11	12	1
	16	16	24	11	12	1
J2*(xDYS413≤18, J2e)						
	14	14	23	11	12	1
	14	14	24	11	13	1
	14	14	25	11	12	1
	14	15	24	11	12	1
	14	15	23	11	12	1
J2e						
	14	15	24	11	12	3
	14	17	24	11	12	1
	15	15	23	11	12	2
	15	15	23	12	12	2
	15	15	24	11	12	5
	15	15	25	11	12	2
	15	17	23	11	12	1
	16	15	23	11	12	1
	16	15	24	11	12	3
	15	15	24	12	12	1
J2-(DYS413≤18)(xJ2a, J2f)						
	12	15	24	11	12	1
	13	15	23	11	12	1

Sub-haplogroup	DYS19	DYS388	DYS390	DYS392	DYS393	Frequency
	14	14	23	11	12	2
	14	14	24	11	13	2
	14	14	24	11	12	3
	14	15	22	11	12	3
	14	15	22	11	14	2
	14	15	23	11	12	16
	14	15	23	11	13	2
	14	15	24	11	13	1
	14	15	24	11	12	2
	14	16	23	11	13	1
	14	16	23	11	12	1
	14	17	23	11	12	4
	15	13	24	11	12	1
	15	14	25	11	12	1
	15	15	23	11	14	2
	15	15	23	11	12	12
	15	15	23	11	13	3
	15	15	24	11	13	1
	15	15	24	11	14	1
	15	15	24	11	12	2
	15	15	25	11	12	4
	15	15	26	11	12	3
	15	16	23	11	12	10
	15	16	24	11	12	2
	15	16	26	11	12	1
	15	17	22	11	12	1
	15	17	24	11	12	1
	16	14	23	11	12	1
	16	15	23	11	12	5
	16	16	23	11	12	1
	16	17	23	11	12	1
	17	13	23	11	13	1
	17	13	25	11	12	1
	17	15	23	11	12	3
J2a						
	14	15	23	11	12	1
J2f*(xJ2f1)						
	14	13	21	11	12	2
	14	13	23	11	12	2
	14	14	23	11	12	1
	14	15	22	11	12	1
	14	15	23	11	12	15
	14	16	22	11	12	1
	14	16	23	11	12	3
	14	16	24	11	12	2
	14	16	24	12	12	2
	14	17	23	11	12	1
	15	15	23	11	10	1
	15	15	23	11	12	3
	16	15	23	11	12	3
J2f1						
	13	15	22	11	12	1
	14	15	22	11	13	1
	14	15	22	11	12	8
	14	15	22	12	12	1
	14	16	22	11	12	1

Sub-haplogroup	DYS19	DYS388	DYS390	DYS392	DYS393	Frequency
	15	15	22	11	12	2
	15	15	22	11	14	2
	15	15	22	11	13	3

References

Al-Zahery N, Semino O, Benuzzi G, Magri C, Passarino G, Torroni A, Santachiara-Benerecetti AS (2003) Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. Mol Phyl Evol 28:458–472

Ammerman AJ, Cavalli-Sforza LL (1984) The Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Barac L, Pericic M, Klaric IM, Rootsi S, Janicijevic B, Kivisild T, Parik J, Rudan I, Villems R, Rudan P (2003) Y chromosomal heritage of Croatian population and its island isolates. Eur J Hum Genet 11:535–542

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

Behar DM, Thomas MG, Skorecki K, Hammer MF, Bulygina E, Rosengarten D, Jones AL, Held K, Moses V, Goldstein D, Bradman N, Weale ME (2003) Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. Am J Hum Genet 73:768–779

Bianchi NO, Catanesi CI, Bailliet G, Martinez-Marignac VL, Bravi CM, Vidal-Rioja LB, Herrera RJ, Lopez-Camelo JS (1998) Characterization of ancestral and derived Y-chromosome haplotypes of New World native populations. Am J Hum Genet 63:1862–1871

Bosch E, Calafell F, Santos FR, Perez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623–1638

Bosch E, Calafell F, Comas D, Oefner PJ, Underhill PA, Bertranpetit J (2001) High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am J Hum Genet 68:1019–1029

Brion M, Salas A, Gonzalez-Neira A, Lareu MV, Carracedo A (2003) Insights into Iberian population origins through the construction of highly informative Y-chromosome haplotypes using biallelic markers, STRs, and the MSY1 minisatellite. Am J Phys Anthropol 122:147–161

Brion M, Quintans B, Zarrabeitia M, Gonzalez-Neira A, Salas A, Lareu V, Tyler-Smith C, Carracedo A (2004) Micro-geographical differentiation in northern Iberia revealed by Y-chromosomal DNA analysis. Gene 329:17—25

Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MP, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N, Goldstein DB (2003) A Y chromosome census of the British Isles. Curr Biol 13:979–984

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

Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the neolithic demic diffusion model. Proc Natl Acad Sci USA 99:11008–11013

Cinnioglu C, King R, Kivisild T, Kalfoglu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, Underhill PA (2004) Excavating Y-chromosome haplotype strata in Anatolia. Hum Genet 114:127–148

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

Di Giacomo F, Luca F, Anagnou N, Ciavarella G, Cresta M, Cucci F, Di Stasi L, Agostiano V, Giparaki M, Loutradis A, Mammi' C, Michalodimitrakis EN, Papola F, Pedicini G, Plata E, Terrenato L, Tofanelli S, Malaspina P, Novelletto A (2003) Clinal patterns of human Y chromosomal diversity in Continental Italy and Greece are dominated by drift and founder effects. Mol Phyl Evol 28:387–395

Dupanloup I, Schneider S, Excoffier L. (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571–2581

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 EA, Jobling MA, Taylor PG, Donnelly P, De Knijff P, Mieremet R, Zerjal T, Tyler-Smith C (1998) Jefferson fathered slave's last child. Nature 396:27–28

Francalacci P, Morelli L, Underhill PA, Lillie AS, Passarino G, Useli A, Madeddu R, Paoli G, Tofanelli S, Calo CM, Ghiani ME, Varesi L, Memmi M, Vona G, Lin AA, Oefner P, Cavalli-Sforza LL (2003) Peopling of three Mediterranean islands (Corsica, Sardinia, and Sicily) inferred by Y-chromosome biallelic variability. Am J Phys Anthropol 121:270–279

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92:6723–6727

Hammer MF, Zegura SL (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Annu Rev Anthropol 31:303–321

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti AS, Oppenheim A, Jobling MA, 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

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612

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 (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, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183–1197

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 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 (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 AI, Barbujani G, Vernesi C, Papola F, Ciavarella G, Kovatchev D, Kerimova MG, Anagnou N, Gavrila L, Veneziano L, Akar N, Loutradis A, Michalodimitrakis En, 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

Manni F, Leonardi P, Barakat A, Rouba H, Heyer E, Klintschar M, McElreavey K, Quintana-Murci L (2002) Y-chromosome analysis in Egypt suggests a genetic regional continuity in north-eastern Africa. Hum Biol 74:645–658

Nasidze I, Sarkisian T, Kerimov A, Stoneking M (2003) Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. Hum Genet 112:255–261

Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A (2001) The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. Am J Hum Genet 69:1095–1112

Nebel A, Landau-Tasseron E, Filon D, Oppenheim A, Faerman M (2002) Genetic evidence for the expansion of Arabian tribes into the southern Levant and north Africa. Am J Hum Genet 70:1594–1596

Paracchini S, Arredi B, Chalk R, Tyler-Smith C (2002) Hierarchical high throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. Nucleic Acids Res 30: e27

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 (2001) Y-chromosome lineages trace diffusion of peoples and languages in southwest Asia. Am J Hum Genet 68:537–542

Rosser ZH, 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 DG, Brede G, Cooper G, Côrte-Real HBSM, De Knijff P, Decorte R, Dubrova YE, Grafo O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Lavinha J, Livshits LA, Malaspina P, Syrrou M, McElreavey K, Meitinger TA, Melegh B, Mitchell RJ, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, João Prata M, Previderé C, Rajczy K, Roewer L, Rootsi S, Rubinsztein DCM, Saillard J, Santos FR, Stefanescu G, Sykes BC, Olun A, Villems R, Tyler-Smith C, Jobling MA (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, Roessli D, Excoffier L (2000) Arlequin v 2000: a software for population genetics 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 MM, Romano V, De Leo G, Gennarelli M, Jaruzelska J, Villems R, Parik J, Macaulay V, Torroni A (2001) Human Y-chromosome variation in the western Mediterranean area: implications for the peopling of the region. Hum Immunol 62:871–884

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

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

Semino O, Passarino G, Quintana-Murci L, Liu A, Beres J, Czeizel A, Santachiara-Benerecetti AS (2000b) MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. Eur J Hum Genet 8:339–346

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, Santachiara-Benerecetti AS (2004) Origin, diffusion, and differentiation of Y-chromosome haplogroups E and J: inferences on the neolithization of Europe and later migratory events in the Mediterranean area. Am J Hum Genet 74:1023—1034

Stumpf MPH, Goldstein DB (2001) Genealogical and evolutionary inference with the human Y chromosome. Science 291:1738–1742

Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of old testament priests. Nature 394:138–140

Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphims on the human Y chromosome. Hum Genet 105:577–581

Thomas MG, Parfitt T, Weiss DA, Weiss A, Skorecki K, Wilson JF, Le Roux M, Bradman N, Goldstein DB (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

Underhill PA, Shen P, Lin A, Jin AA, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad M, Wells S, Piazza A, Davis RW, Feldman MF, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Underhill PA, Passarino G, Lin AA, Shen P, Mirazon Lahr M, Foley RA, Oefner PJ, Cavalli-Sforza LL (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62

Weale ME, Yepiskoposian L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum Genet 109:659–674

Weale ME, Shah T, Jones AL, Greenhalgh J, Wilson JF, Nymadawa P, Zeitlin D, Connell BA, Bradman N, Thomas MG (2003) Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. Genetics 165:229–234

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249

Wilson IJ, Balding DG (1998) Genealogical inference from microsatellite data. Genetics 150:499–510

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions 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

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into central Asia. Am J Hum Genet 71:466–482

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. Am J Hum Genet 72:717–721

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am J Hum Genet 74:50–61