Hum Genet (2004) 115: 357–371
DOI 10.1007/s00439-004-1168-9

ORIGINAL INVESTIGATION

. D . . . . . A . . A . . B .
. B . . B . . . . C . . C . . D .
. . . . D. . A. . . A. .
. . C. . . . . .
. . . . . A.

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. 1999f(c73ot(Tvar026)(varlpop263(Y]ranva40ch-2.osom8n)c09-.23iverseact)40p4((ncrossr(;).[68sio

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

		Sample size	Frequency of J	Sub-haplogroup								
				J*(xJ1, J2)	J1	J2*(xDYS413≤18, J2e)	J2e	J2-(DYS413≤18) (xJ2a, J2f)	J2a	J2f*(xJ2f1)	J2f1	Diversity within J
				Derived state at locus								
Population sample (no. of subsamples)	Code			p12f2	p12f2, M267	p12f2, M172	p12f2, M172, M12	p12f2, M172, DYS413	p12f2, M172, DYS413, M47	p12f2, M172, DYS413, M67	p12f2, M172, DYS413, M67, M92	
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		2Egypt (2)0.1407				

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.

. 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≤

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-TATCTACAAATA-3′

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;

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

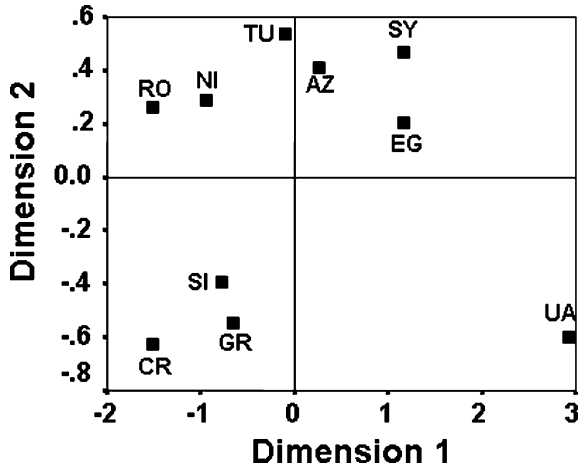
$$SSQab_i = \sum_{j=1}^{n} \left[A_{ij} - E(A_{ij})\right]^2$$

where

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

The set of x

(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



. 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.

. . 3

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,

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.

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

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
	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

–

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

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