doi: 10.1111/j.1529-8817.2005.00224.x

Population Structure in the Mediterranean Basin: A Y Chromosome Perspective

C. Capelli[1,2,*], N. Redhead[1], V. Romano[3], F. Calì[4], G. Lefranc[5], V. Delague[6,†], A. Megarbane[6], A. E. Felice[7], V. L. Pascali[2], P. I. Neophytou[8], Z. Poulli[9], A. Novelletto[10], P. Malaspina[11], L. Terrenato[11], A. Berebbi[12], M. Fellous[13], M. G. Thomas[1] and D. B. Goldstein[1]

[1]*Department of Biology, University College of London, London, UK*
[2]*Istituto di Medicina Legale,Università Cattolica del Sacro Cuore, Rome, Italy*
[3]*Department of Biopathology and Biomedical Methodologies, University of Palermo, Palermo, Italy*
[4]*Oasi Institute for Research on mental Retardation and Brain Aging (IRCCS), Troina, Italy*
[5]*Université Montpellier II et Institut de Génétique Humaine, CNRS UPR 1142, Montpellier, France*
[6]*Unité de Génétique Médicale, Faculté de Médecine, Université Saint-Joseph, Beyrouth, Lebanon*
[7]*Laboratory of Molecular Genetics, Department of Pathology, University of Malta, Malta*
[8]*Mendel Center for Biomedical Sciences, Eglomi, Nicosia, Cyprus*
[9]*Cyprus Research Center (KYKEM), Nicosia, Cyprus*
[10]*Department of Cell Biology, University of Calabria, Rende, Italy*
[11]*Department of Biology, University "Tor Vergata", Rome, Italy*
[12]*Hematology Institute, Kaplan Medical Center, Rehovot, Israel*
[13]*Department of Human Genetics, Université Paris, Paris, France*

Summary

The Mediterranean region has been characterised by a number of pre-historical and historical demographic events whose legacy on the current genetic landscape is still a matter of debate. In order to investigate the degree of population structure across the Mediterranean, we have investigated Y chromosome variation in a large dataset of Mediterranean populations, 11 of which are first described here. Our analyses identify four main clusters in the Mediterranean that can be labelled as North Africa, Arab, Central-East and West Mediterranean. In particular, Near Eastern samples tend to separate according to the presence of Arab Y chromosome lineages, suggesting that the Arab expansion played a major role in shaping the current genetic structuring within the Fertile Crescent.

Keywords : Mediterranean, Y chromosome, population genetic structure, STRs, UEPs.

Introduction

The Mediterranean basin has witnessed a number of dramatic demographic events throughout time. The area has been peopled since the very beginning of the human diaspora from the Africa continent. The archeological sites of Skhul and Qafzeh in the Middle-East (Stringer & Gamble, 1993) contain the oldest *Homo sapiens* remains outside Africa, dating to 90–100000 years before present (yBP), and Southern Greece, Sicily and South Coastal Spain were peopled well before 10–12000 yBP. By 9000 yBP agricultural development occurred in the Near East and this new technology spread throughout Europe in the next millennia (Cunliffe, 2001). The demographic effects the Neolithic revolution had on the European peninsula are still a matter of debate among both geneticists (Barbujani *et al*. 1998; Chikhi *et al*. 1998; Richards *et al*. 2000; Simoni *et al*. 2000; Semino *et al*. 2000; Rosser *et al*. 2000; Torroni *et al*. 2001; Chikhi *et al*.

* Corresponding author: Cristian Capelli, Genetic Identity Europe, Promega Corporation, Madison WI, U.S.A. Promega Italia, Via Decembrio 28, 20137 Milano. Tel: +39 0773 458535; Fax: +39 0773 458535. E-mail: cristian.capelli@promega.com
† Current address: INSERM U491, Laboratoire Génétique Médicale et Développement Faculté de Médecine de la Timone, Marseille, France. The opinions expressed in this article by C. Capelli do not necessarily reflect those of Promega Corporation

© University College London 2005

2002; Richards *et al*. 2002) and archaeologists (Ammerman & Cavalli-Sforza, 1984; Zvelebil & Zvelebil, 1988), even though in the frame of hypothesis testing, actual available data seems to support a Neolithic Demic Diffusion (NDD) scenario (Barbujani & Goldstein, 2004). The complexity of Mediterranean genetic history is testified also by the number of historical events occurring in this area: the foundation of Greek colonies in the west Mediterranean (South Italy, Spain and coastal France; Burn, 1990), Phoenician and Carthaginian settlements (Cunliffe, 2001), and the creation of the Roman Empire, which at its peak dominated the entire Mediterranean coast. Later in time, the Arab conquest had a major historical impact, starting from the Arabian Peninsula, moving through the Near East and finally reaching North Africa, the Iberian Peninsula and Sicily (Dussaud, 1955; Finley, 1968; Ricci, 1984; Abun-Nasr, 1987; Hitti, 1990). The numbers of population contacts that occurred along the Mediterranean coast following commercial routes (Braudel, 1998) are an additional element, suggesting that the gene flow across groups was very large and that it possibly erased genetic differences across the basin. A recent study of Y chromosome variation in Mediterranean populations showed the distinctiveness of North Africans compared to the rest of the Mediterranean, but failed to find any heterogeneity among other Mediterranean populations (Quintana-Murci *et al*. 2003). Using a more formal inferential framework, Chikhi *et al*. (2002) estimated the current degree of Neolithic component in European populations, and revealed similar Near Eastern contributions in Mediterranean and continental European samples. Flores *et al*. (2004) investigated the structure within Iberia and included in their analysis a group of samples representing Mediterranean populations. The North African populations tended to separate clearly from other samples, with some indication of structure between Near East and Mediterranean samples, but substantial homogeneity within Northern Mediterranean groups was observed. Nebel *et al*. (2002) reported the presence of Arab markers in some Near Eastern populations, possibly as a result of the Arab expansion, but did not specifically question if this had any effect on population structure within the Near East, and Hammer *et al*. (2000) pointed to a common male origin for Jews and Middle Easterners.

So far little attention has been devoted to differences within the Near East populations and across Mediterranean groups. The current set of data now makes it possible to specifically address this issue.

Here we explicitly assess Mediterranean genetic structure by analysing Y chromosome distribution in 11 different populations with a battery of both fast and slow evolving markers (Fig. 1). A total of 656 Y-chromosomes were typed for 6 Y linked STRs and 16 Unique Event Polymorphisms (UEPs). Additionally, we included published Y chromosome data for SNPs (58 populations) and STRs (34 populations) (Bosch *et al*. 1998, 2001; Nebel *et al*. 2001; Semino *et al*. 2000; Malaspina *et al*. 2001; Wilson *et al*. 2001; Luis *et al*. 2004; Flores *et al*. 2004; Cinnioglu *et al*. 2004; Thomas *et al*. 2000; Arredi *et al*. 2004).Our analyses revealed significant structure in the Mediterranean Y chromosome gene pool, in particular pointing to structure within the Middle East due to the presence of an Arab genetic component, a result also supported by a reanalysis of published mtDNA and autosomal markers (Richards *et al*. 2002; Thomas *et al*. 2002; Plaza *et al*. 2003; Rosenberg *et al*. 2002).

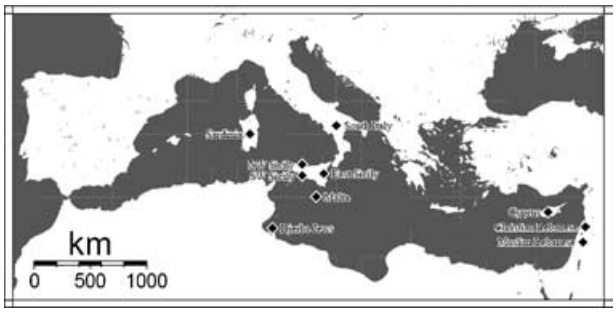

Figure 1 Geographic origin of the investigated samples.

Materials and Methods

Samples

Geographic origin of populations genotyped in this study is indicated in Figure 1. The Sicilians, a subset of Sardinians and Tunisian sample origins have already been described (Romano *et al*. 2003; Ciminelli *et al*. 1995; Comas *et al*. 2000; Plaza *et al*. 2003). Sample sizes are as in Table 1. Haplogroup and haplotype published data included in the analyses are described in Table 1.

© University College London 2005

Y Chromosome Genotypings

Y-STRs were genotyped as described in Thomas *et al*. (1999) and included DYS388, DYS393, DYS392, DYS19, DYS390 and DYS391 loci. UEPs were genotyped using monoplex and multiplex reactions (Table 2). YAP, SRY 10831, SY81 and SRY 4064 were genotyped following published protocols (Thomas *et al*. 1999). RPS4Y polymorphism was genotyped as described in Capelli *et al*. (2001). Tat and 12f2 markers were scored as indicated in Rosser *et al*. (2000). All other markers were genotyped as described below. The genealogical relationship between the selected markers is shown in Figure 2, following the Y Chromosome Consortium nomenclature (2002). PCR was carried out in a final volume of 10 μl. The reaction contained 1X buffer (HT Biotech, Cambridge), 200 μM dNTPs, and 0.13 units of *Taq* polymerase enzyme (HT Biotech, Cambridge), as well as 9.3 nM TaqStart Antybody (Clontech). The cycling conditions were 94°C for 5 minutes, and then 35 cycles at 95°C for 40 sec, 59°C/58°C/55°C for 40 sec and 40 sec at 72°C, with a final incubation at 72°C for 10 minutes. For the M173/M17/M172/M170/M9/92R7 multiplex, the final concentration of $MgCl_2$ was 2.2 mM. Primers and Endonucleases are shown in Table 2. Restriction reactions were performed following the supplier's information (New England Biolabs, Beverly, Mass.)

The PCR mix preparation was performed following suggestions described in Thomas *et al*. (1999).

Appropriate allelic control DNAs where included to confirm successful restriction reaction conditions. The digested products were run on a 377 ABI instrument (Applied BioSystems, Foster City, CA).

The M173/M17/M172/M170/M9/92R7 multiplex was firstly genotyped in all the samples. M9 derived samples, ancestral at 92R7, were also tested for tat. M9, M170, M172 underived samples were tested for M89, 12f2 and M35. Samples ancestral at these markers were additionally tested for SRY 10831, YAP and RPS4Y. YAP derived samples ancestral at M35 were tested for SRY4064 and SY81. M26 was tested only on M170 derived chromosomes.

Two chromosomes found in Cyprus, ancestral at M89 and YAP markers, could not be tested for RPS4Y and SRY10831. Haplotype comparison with unpublished data suggested considering them as hg A. Similarly, due to DNA not being available, three Djerba chromosomes, underived at M9, 12f2, M170, were assigned to haplogroups E3b (1 chromosome) and FxIJK (2 chromosomes) by haplotype comparison with fully genotyped available samples. For hgs analysis, chromosomes were clustered in 8 groups following genealogical information: DE; Y*(xA,D,E,J,K); P*(xR1a1); J*(xJ2); J2; A; K*(xP), R1a1.

Statistical Analysis

Principal Components (PC) analysis was performed on haplogroup frequencies by using POPSTR software (Henry Harpending, pers comm.). Geographic structuring of variation was assessed using AMOVA as implemented in the Arlequin software (Schneider *et al*. 2000). AMOVA analysis was performed taking in consideration STR data only and using Rst genetic distance (Slatkin, 1995). Samples included in the analyses are indicated in Table 1. MtDNA PC analysis was perfomed on a set of published data (Richards *et al*. 2002; Thomas *et al*. 2002; Plaza *et al*. 2003) Admixture estimations were calculated by LEA and ADMIX software (Chikhi *et al*. 2001; Bertorelle & Excoffier, 1998; Dupanloup & Bertorelle, 2001) using hg information (i.e. binary markers only) and considering the E3b North African modal type as separate from the rest of the E3b group. Selected autosomal data from Rosenberg *et al*. (2002) was investigated by the use of the STRUCTURE program (Pritchard *et al*. 2000). Each STRUCTURE run was performed twice with a burn-in and run length of 200,000 each.

Results and Discussion

Mediterranean Population Structure

We analysed 16 biallelic markers (otherwise known as Unique Event Polymorphisms - UEPs) identifying hgs (Fig. 2, Table 3), and 6 Y linked microsatellites (Thomas *et al*. 1999) defining hpts (Appendix). To include additional data from other reference populations (Table 1) Y chromosome PC analysis (Fig. 3a) was performed with a lower level of resolution, collapsing the 14 observed haplogroups into 8 haplogroups, as described in

© University College London 2005

Table 1 Geographic origin of all the samples included in the Y chromosome and STRs analyses. Codes, sample sizes and references are indicated. na: not available

Haplogroup data	Haplotype data	Code	Sample size	Reference
Central-East Mediterranen				
East Sicily	East Sicily	ESC	87	this study
South-West (SW) Sicily	South-West (SW) Sicily	SWS	55	this study
North-West (NW) Sicily	North-West (NW) Sicily	NWS	70	this study
South Italy	South Italy	SIT	68	this study
Sardinia	Sardinia	SAR	81	this study
Malta	Malta	MAL	90	this study
Cyprus	Cyprus	CYP	65	this study
Turkey	na	TRS	20	Semino *et al*. 2000
na	Turkey		68	Wilson *et al*. 2001
na	Turkey		523	Cinnioglu *et al*. 2004
Turkey-1	na	TR1	52	Cinnioglu *et al*. 2004
Turkey-2	na	TR2	29	Cinnioglu *et al*. 2004
Turkey-3	na	TR3	83	Cinnioglu *et al*. 2004
Turkey-4	na	TR4	82	Cinnioglu *et al*. 2004
Turkey-5	na	TR5	43	Cinnioglu *et al*. 2004
Turkey-6	na	TR6	33	Cinnioglu *et al*. 2004
Turkey-7	na	TR7	90	Cinnioglu *et al*. 2004
Turkey-8	na	TR8	30	Cinnioglu *et al*. 2004
Turkey-9	na	TR9	81	Cinnioglu *et al*. 2004
West Mediterranean (Iberians)				Cinnioglu *et al*. 2004
Seville	na	SEV	155	Flores *et al*. 2004
Huelva	na	HUE	22	Flores *et al*. 2004
Cadiz	na	CAD	28	Flores *et al*. 2004
Cordoba	na	COR	27	Flores *et al*. 2004
Malaga	na	MAG	26	Flores *et al*. 2004
North Portugal	na	NPF	109	Flores *et al*. 2004
Leon	na	LEO	60	Flores *et al*. 2004
Galicia	na	GAL	19	Flores *et al*. 2004
Cantabria	na	CAN	70	Flores *et al*. 2004
Valencia	na	VAL	31	Flores *et al*. 2004
Castile	na	CAS	21	Flores *et al*. 2004
Madeira	na	MAD	129	Goncalves *et al*. 2005
Acorres	na	ACO	121	Goncalves *et al*. 2005
North Portugal	na	NPC	101	Goncalves *et al*. 2005
Central Portugal	na	CPC	102	Goncalves *et al*. 2005
South Portugal	na	SPC	100	Goncalves *et al*. 2005
Basques	Basques	BAS	42	Bosch *et al*. 1998; 2001
Catalans	Catalans	CAT	14	Bosch *et al*. 1998; 2001
Andalusians	Andalusians	AND	37	Bosch *et al*. 1998; 2001
Middle East and Arabian Peninsula				
Oman Arabs	Oman Arabs	OMA	121/122*	Luis *et al*. 2004
na	Syria		72	Wilson *et al*. 2001
Syria	na	SYR	20	Semino *et al*. 2000
na	Yemen		76	Wilson *et al*. 2001
United Arab Emirates	na	UAE	33	Malaspina *et al*. 2001
Iraq	na	IRQ	139	Al-Zahery *et al*, 2003
Palestinians	Palestinians	PAL	143	Nebel *et al*. 2001
Beduins	Beduins	BED	32	Nebel *et al*. 2001
Muslim Kurds	Muslim Kurds	MKU	95	Nebel *et al*. 2001
Kurdish Jews	Kurdish Jews	KJE	99	Nebel *et al*. 2001
Ashkenazi Jews	Ashkenazi Jews	AJE	79	Nebel *et al*. 2001
Sephardic Jews	Sephardic Jews	SJE	78	Nebel *et al*. 2001

 © University College London 2005

Table 1 (continued)

	Haplogroup data	Haplotype data	Code	Sample size	Reference
	Muslim Lebanese	Muslim Lebanese	MLE	39	this study
	Christian Lebanese	Christian Lebanese	CLE	43	this study
North Africa					
	Algeria Arabs	Algeria Arabs	AAR	35	Arredi *et al*. 2004
	Algeria Berbers	Algeria Berbers	ABE	19	Arredi *et al*. 2004
	North Egypt	North Egypt	NEG	44	Arredi *et al*. 2004
	South Egypt	South Egypt	SEG	29	Arredi *et al*. 2004
	Egypt Arabs	Egypt Arabs	EAR	147/148*	Luis *et al*. 2004
	Tunisia	Tunisia	TUA	148	Arredi *et al*. 2004
	Tunisia	Tunisia	TUN	39	this study
	Djerba Jews	Djerba Jews	DJE	19	this study
	Moroccan Arabs	Moroccan Arabs	MAR	44	Bosch *et al*. 1998; 2001
	Moroccan Berbers	Moroccan Berbers	MBA	60	Bosch *et al*. 1998; 2001
	Saharawis	Saharawis	SAH	29	Bosch *et al*. 1998; 2001
Total populations	59	34			
Total individuals	3807	2692			

* Haplogroup/haplotype available individuals.

the Methods section. PC results were shaped by hgs P*(xR1a1), J*(xJ2) and DE along axis one (loading factors: 0.484, −0.292, and −0.299, respectively) and by hgs J2 and DE on axis 2 (loading factors: 0.253 and −0.324, respectively). The PC plot suggested the presence of four main groups (Fig. 3a): 1) North Africa, 2) Near East/Arabs (including Muslim Lebanese and Ashkenazi Jews), 3) Central-East Mediterranean grouping, including Christian Lebanese and 4) West Mediterranean. A population differentiation test was performed using an analogue of the Fisher Exact Test as implemented in the Arlequin software (Schneider *et al*. 2000)

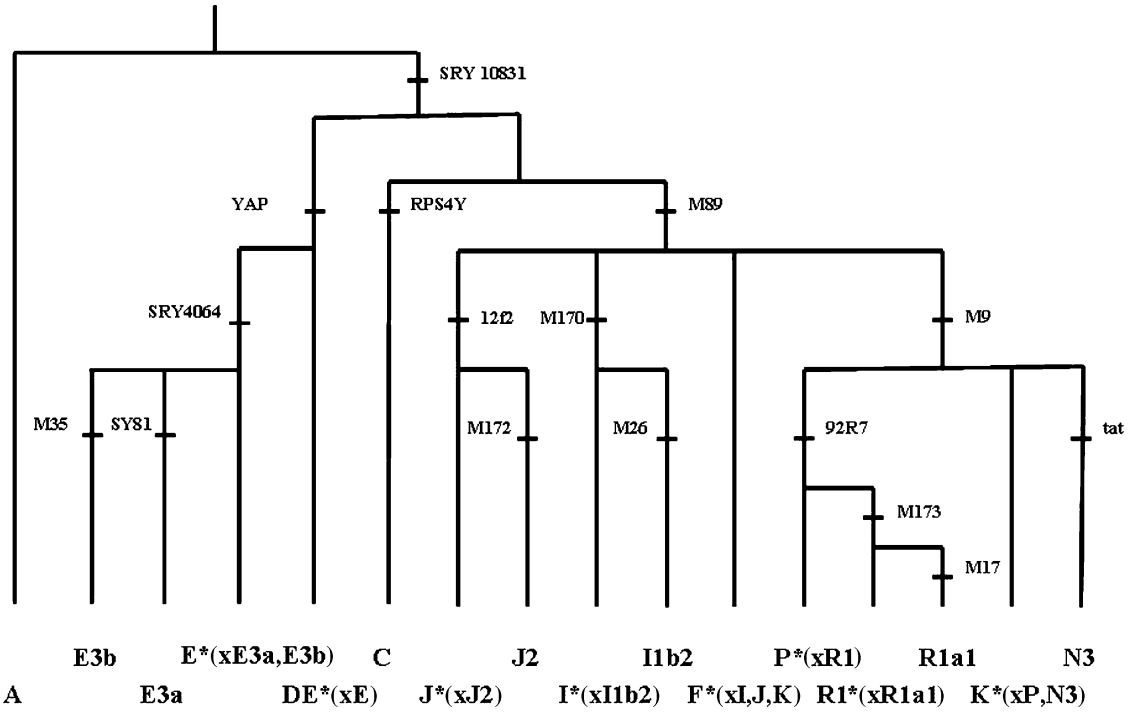

Figure 2 Genealogical relationships of the selected UEPs. Nomenclature as suggested by YCC (2002).

© University College London 2005

Table 2 Information on UEP genotyping protocols.[a)] Underlined are mismatch nucleotides used to create/eliminate restriction sites;[b)] Sizes were estimated using a 4.25% acrylamide gel run on a 377 ABI automated sequencer with TAMRA 350 (Applied Biosystem) as internal size standard;[c)] M17 primers were described in Thomas *et al.* (1999);[d)] 12f2 primers were described in Rosser *et al.* (2000); e) tat primers were described in Rosser *et al.* (2000)

Reaction	Locus	Primers	Label	Primers sequence (5′ -3′)[a)]	Primer concentration (μM)	Amplicon size[b)]	Discriminating enzyme	Fragment size ancestral nucleotide	Fragment size derived nucleotide
1									
	92R7	92R7u	HEX	TCAGAAAGATAGTAAGAGGAACACTTC	0.15	95	*Hind* III	95+66 (C)	95(T)
		92R7r		GCATTGTTAAATATGACCAGCA					
	M9	M9L F		CATTGAACGTTTGAACATGTC	0.15	114	*Hinf* I	67 (C)	97(G) (cut by *Afl* III)
		M9L R	TET	TGCAGCATATAAAACTTTCAGG					
	M17[c)]	M17 F		GTGGTTGCTGGTTGTTACGT	0.15	125	*Afl* III	125 (+G)	105 (-G)
		M17 R	TET	AGCTGACCACAAACTGATGTAGA					
	M170	M170 F		TTACTATTTTATTTACTTAAAAATCATTGATC	0.5	111	*Bcl* I	83 (A)	111 (C)
		M170 R	HEX	CCAATTACTTTCAACATTTAAGACC					
	M172	M172 F	TET	TTAGCCAGATGACCAGGATGC	0.2	173	*Hinf* I	173 (A)	143 (G)
		M172 R		GAAAATAATAATTGAAGACCTTTTGAGT					
	M173	II M173 F		ACAATTCAAGGGCATTTTGTAC	0.4	117	*Bsr G* I	100 (A)	117 (C)
		M173 R	FAM	CTTACTCAGTATGGGTAAAAGAAATGC					
2									
	M89	M89 F	HEX	GAAAGTGGGGCCCACAG	0.35	98	*Nla* III	80 (C)	98 (T)
		M89 R		AACTCAGGCAAAGTGAGACAT					
	12f2 [d)]	12f2 D	HEX	CTGACTGATCAAAATGCTTACAGATC	0.35	88	-	88	locus deleted
		12f2 G		GGATCCCTTCCTTACACCTTATAC					
	tat[e)]	Tat F		GACTCTGAGTGTAGACTTGTGA	0.35	112	-	PCR control	PCR control
		Tat R	FAM	GAAGGTGCCGTAAAAGTGTGAA					
3	M26	M26F		CAATTTCTTTCTGAATTAGAATGATC	0.5	169	*Bcl* I	169 (G)	147 (A)
		M26R	HEX	CCATACACAAGGATGCAGCAC					
4	M35	IIM35F	FAM	GAAACTGAGAGGGCAAGGTC	0.5	161	*Bsr I*	131 (G)	161 (C)
		IIM35R		GGAGCTTCTGCCTGTTGC					

© University College London 2005

on hgs frequencies, using the same 8 hgs tested for PC analysis. The pattern of population similarities paralleled that displayed by PC analysis, with populations in the same cluster being mostly not significantly different from each other ($P > 0.05$, data not shown). Population relationships were also investigated by the use of microsatellite variation on a total of 34 populations (Table 1). Analysis of the distribution of molecular variation on STRs haplotypes (performed using Arlequin software; Schneider *et al*. (2000)) was also consistent with the four group clustering suggested by the PC analysis, with the lowest within groups variation (1.57%) and largest between groups variation (7.70%, Φst 0.077, $P \ll 0.01$). Clustering of the West and Central-East Mediterranean populations lowered the intergroup variation to 6.95 % and increased the intragroup value to 2.35% ($P \ll 0.01$). Following a geographical clustering scheme as indicated in Table 1, genetic variation showed a between groups value of 5.18% and a within groups percentage of 3.06 ($P \ll 0.01$). Finally, clustering the North African Arab samples with the Arab cluster of Figure 3a displayed a between groups value of 6.18% and a within groups percentage of 2.71 ($P \ll 0.01$).

We also evaluated the degree of differentiation shown by Mediterranean populations by performing PC analysis on previously published mtDNA haplogroup data (Richards *et al*. 2002; Thomas *et al*. 2002; Plaza *et al*. 2003). In general agreement with the Y chromosome results, a Mediterranean, an Arab and a North African cluster were apparent (Fig. 3b). Contrary to the Y chromosome PC plot (Figure 3a) West and Central Mediterranean samples did not cluster independently, while a certain tendency to separate was shown by Middle Eastern samples. Mediterraneans and North Africans were closer to each other than to Arabic populations (Bedouins and Yemenite), a result previously interpreted as a signature of female mediated gene flow across the Mediterranean basin (Plaza *et al*. 2003). Near Eastern Arab populations (Palestinians, Jordanians and Syrians) did not cluster with the Arabians and were instead closer to Mediterranean populations. This result is in contrast with the Y chromosomal data, where Arabian and non-Arabian Arabs grouped together (Fig. 3a). Sex mediated differential contacts between the Arabian peninsula and the Near East could be responsible for the different observed pattern. In the Middle East, only Bedouins seem

Table 3 Population Y hg frequencies (in brackets absolute number of chromosomes) investigated in this study. No hg D and N3 chromosomes were found

Sample	A		E3b		E3a		E*(xE3a, E3b)		C		J*(xJ2)		J2		I*(xI1b2)		I1b2		F*(xI, J,K)		K*(xP, N3)		P*(xR1)		R1*(xR1a1)		R1a1		n
ESC			0.287	(25)							0.069	(6)	0.287	(25)	0.034	(3)	0.011	(1)	0.046	(4)	0.046	(4)			0.195	(17)	0.023	(2)	87
SWS			0.164	(9)									0.273	(15)	0.055	(3)			0.145	(8)	0.055	(3)			0.291	(16)	0.018	(1)	55
NWS			0.214	(15)					0.014	(1)	0.071	(5)	0.114	(8)	0.157	(11)			0.114	(8)	0.014	(1)	0.014	(1)	0.257	(18)	0.029	(2)	70
SIT			0.265	(18)							0.044	(3)	0.162	(11)	0.044	(3)	0.015	(1)	0.147	(10)	0.029	(2)	0.015	(1)	0.25	(17)	0.029	(2)	68
SAR	0.012	(1)	0.099	(8)							0.049	(4)	0.099	(8)	0.049	(4)	0.259	(21)	0.185	(15)	0.012	(1)	0.025	(2)	0.21	(17)			81
MAL			0.089	(8)	0.011	(1)	0.011	(1)			0.078	(7)	0.211	(19)	0.122	(11)			0.067	(6)	0.044	(4)	0.011	(1)	0.322	(29)	0.033	(3)	90
TUN			0.513	(20)							0.231	(9)	0.103	(4)	0.026	(1)			0.051	(2)					0.051	(2)	0.026	(1)	39
DJE			0.053	(1)							0.316	(6)	0.526	(10)					0.105	(2)									19
MLE			0.179	(7)							0.308	(12)	0.256	(10)					0.128	(5)	0.051	(2)			0.051	(2)	0.026	(1)	39
CLE			0.163	(7)							0.093	(4)	0.349	(15)	0.047	(2)			0.186	(8)	0.047	(2)	0.07	(3)	0.047	(2)			43
CYP		0.031	(2)	0.2	(13)						0.062	(4)	0.369	(24)	0.077	(5)			0.092	(6)	0.046	(3)			0.092	(6)	0.031	(2)	65
																													656

© University College London 2005

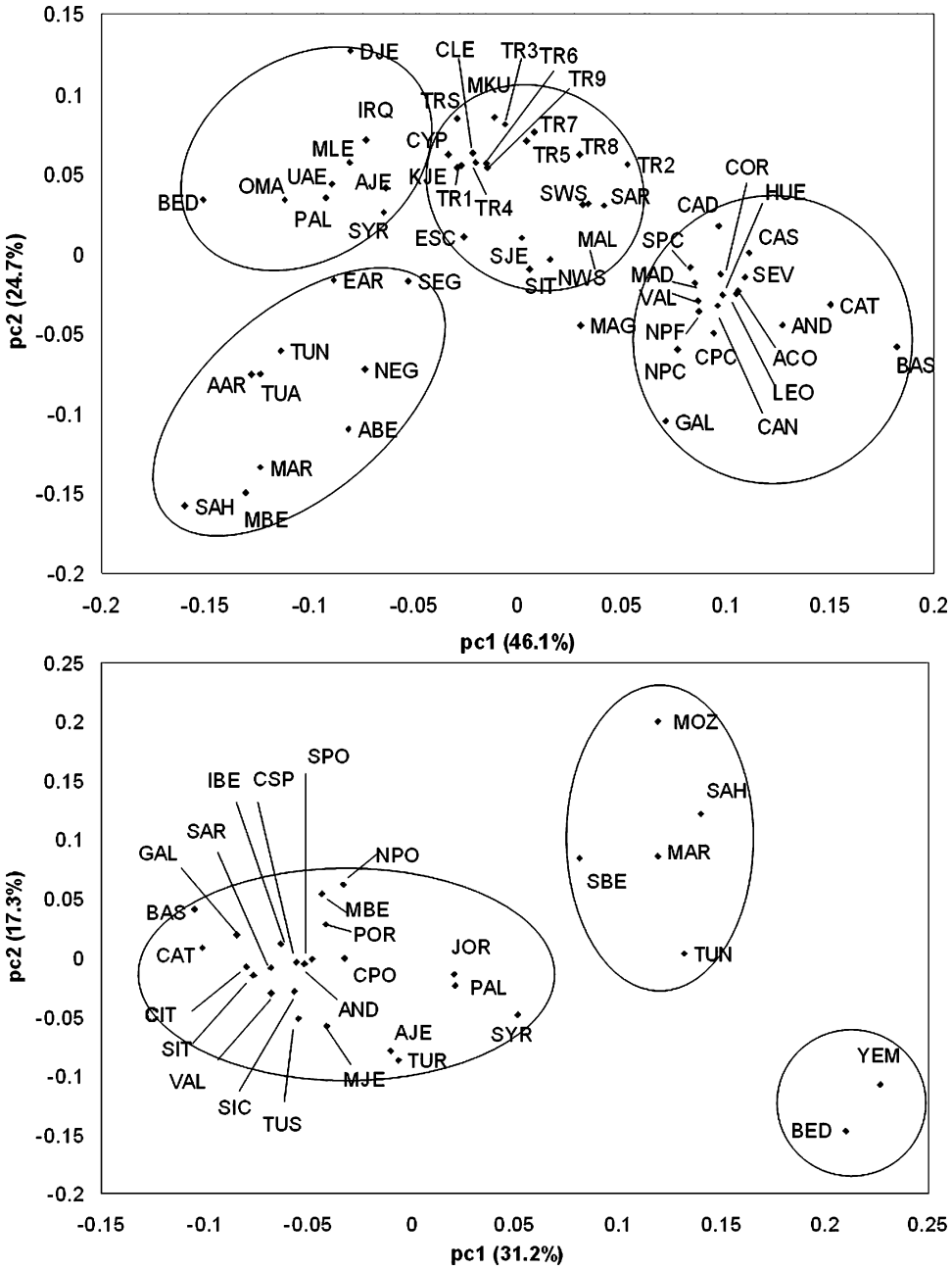

Figure 3 a)Y chromosome hgs Principal Components plot. Codes as in Table 1; b) MtDNA hgs Principal Components plot. Codes for Figure 3b as follows: MAR, Moroccan Arabs; SBE, South Berbers; MOZ, Mozabites; SAH, Saharawis; TUN, Tunisians; MBE, Moroccan Berbers; AND, Andalusians; BAS, Basques; CAT, Catalans; CSP, Central Spanish; GAL, Galicians; VAL, Valencians; POR, Portuguese; NPO, North Portuguese; CPO, Central Portuguese; SPO, South Portuguese; IBE, Iberians; CIT, Central Italians; SIT, South Italians; SIC, Sicilians; TUS, Tuscans; SAR, Sardinians; AJE, Ashkenazi Jews; MJE, Moroccan Jews; TUR, Turks; JOR, Jordanians; PAL, Palestinians; SYR, Syrians, YEM, Yemenites; BED, Bedouins. Indicated are the percentages of genetic variation expressed by each principal component.

 © University College London 2005

to have had extensive contacts with Arabians, as in both mtDNA and Y chromosome analyses they cluster together. Recently Rosenberg *et al*. (2002) published an investigation of world-wide populations genotyped at almost 400 microsatellite loci. They assessed population differentiation using the clustering scheme implemented in the STRUCTURE software (Pritchard *et al*. 2000). Using the same approach we focused our attention on samples present in their dataset that are comparable to ours, namely Sardinians and Italians, Mozabites from North Africa, Bedouins and Palestinians. Previous analysis revealed independent clustering on Basques from other European samples (Rosenberg *et al*. 2002); therefore, in order to make the computation time feasible, we did not include this group in our analysis. Several runs were performed and the highest likelihood values were obtained when testing a number of clusters K >= 4. These K values were associated with overlapping likelihood values across runs, so that it was not possible to identify the most likely number of clusters (data not shown). However, in all these cases at least three different groups were present: group 1 clustered the majority (over 70%) of Italians and Sardinians, and a small part of Bedouins and Palestinians; group 2 was represented by almost 90% of North Africans; and group 3 contained the majority of Palestinians and a subset of Bedouins. Additional groups were composed by subsequent subdivision of the Bedouin sample. Clusters 1, 2 and 3 broadly mirrored Mediterranean, North African and Arab Y chromosome and mtDNA groupings. Taking into consideration the independent clustering of Basques from other European samples (Rosenberg *et al*. 2002), the four clustering scheme suggested by Y chromosome analysis seems to be supported also at an autosomal level. Additional groups, almost exclusively represented by Bedouins, were also observed. These additional clusters most likely reflect specific gene flow or drift occurring in Bedouins. Richards *et al*. (2003) found genetic evidence of female gene flow from Sub-Saharan Africa into Near East Arab populations, resulting from the slave trade. When we included in the STRUCTURE analysis the available Sub-Saharan samples from Rosenberg *et al*. (2002), the three African populations (Bantu-speaking Kenians, Yoruba and Mandenka) strongly clustered together, with a minor portion (not above 10%) of North African and Bedouins. Assuming correct identification of source populations, this suggests minor Sub-Saharan contribution to current Near East Arab populations.

Near East, Jewish communities and Arabians: evidence of Arab male genetic contribution to Levantine populations

The independent clustering of Near Eastern Arab and non-Arab populations is driven by J*(xJ2) frequencies, above 30% in all populations in the Arab group (Table 1 and Semino *et al*. 2004). Nebel *et al*. (2002) suggested this group (H71 in their nomenclature) as a possible marker of the Arab expansion in the Southern Levant and North Africa. The authors found a J*(xJ2) haplotype mainly restricted to Arab populations or Arabised ones (Galilee Modal Haplotype, GMH, Nebel *et al*. 2000), and virtually absent from Jewish groups and the North Mediterranean coast (this study and Nebel *et al*. 2001). The same haplotype has also been identified in various North African groups (this study and Bosch *et al*. 2000, 2001) and it is the modal hpt in Syria and Yemen (Thomas *et al*. 2000). J*(xJ2) hg frequency has an important influence on the distribution of populations on the PC plot axis 2, the axis along which Arab and Arabian populations tend to cluster. These populations also show high frequencies for the GMH and its one microsatellite mutational step neighbours. J*(xJ2) and GMH appear then to be non-randomly distributed across populations, and together may represent a signature for Arabian influence. Interestingly, considering the two Lebanese samples, the Muslims contained GMH and its one microsatellite mutation step cluster at frequencies of 4.6% and 14%, respectively, while on the contrary the Christians displayed only the GMH (2.5%). Hg J*(xJ2) frequencies in Muslim and Christian Lebanese were 31% and 9%, respectively.

Nebel *et al*. (2002) suggested a "working model" with two main waves of contacts between the Near East and the Arabian peninsula. An earlier Near East expansion toward the Arabian Peninsula during Neolithic time would be identified by types shared also with Jewish communities, while Arabian and Near East group uniquely shared types could be instead related to a more recent population movement, following the dispersion of nomadic tribes out of the Arabian peninsula.

© University College London 2005

Arabian male admixture with Near East populations could explain the independent grouping of these from Mediterranean populations, as shown by Y chromosome data. The different pattern shown by mtDNA (Fig. 3b), with non-Arabian Arabs clustering with Mediterranean groups, suggests that Arab introgression had different effects in different groups possibly due to sex-biased gene flow.

This scenario has important implications also on the investigation of the origin of Jewish communities. Genetic variation in Jewish populations has been investigated in several recent papers using different genetic systems (e.g. Hammer *et al*. 2000; Nebel *et al*. 2001; Thomas *et al*. 2002; Behar *et al*. 2003). Basing their analysis on Y chromosome variation, Hammer *et al*. (2000) recently found similarity between Near Eastern and Jewish communities, and suggested a common Near Eastern origin for all these groups. We have focused on Mediterranean and Near Eastern populations and highlight a significant separation between "Arabian" Near Easterners and Jewish populations. Both PC and AMOVA analyses suggest a more complex scenario, with Sephardic and Kurdish Jews closer to Mediterraneans, and Djerba and Ashkenazi instead showing a higher affinity to Arab populations. A larger Djerban Jewish sample confirms this observation (Leonardi *et al*. personal communication). These results contrast with those of Hammer *et al*. (2000). However, the main differences between the two investigations are the specific focus on Mediterranean populations and the additional level of resolution within hg J, offered by the inclusion of the M172 marker and microsatellite data. Hg J frequencies are in fact similar across Jewish communities and Near Easterners, while J2 and J* (xJ2) hgs frequencies highlight important differences. Hg J* (xJ2) is always more frequent (sometime even as much as twice as frequent) than J2 in the "Arabian" Near Eastern populations, while J2 is more frequent than J* (xJ2) in the Jewish populations (Table 2; Nebel *et al*. 2001; Cinnioglu *et al*. 2004; Luis *et al*. 2004; Semino *et al*. 2004) The inclusion of genetically different worldwide populations such as Sub Saharan African, European and North African in the MDS analysis by Hammer *et al*. (2000) probably favoured the clustering of broadly similar populations such as Near Easterners and Jews, due to their high hg J frequencies (between 28% and 84%). All the Europeans included in their analysis had low hg J frequencies except the only two Mediterranean samples, Italy and Greece, which were in fact the closest to Near Eastern/Jews in their MDS analysis. The heterogeneity we observed between the Jewish groups included here and the Near East populations might be related to the independent genetic histories and European admixture (Thomas *et al*. 2002; Nebel *et al*. 2002; Behar *et al*. 2003), and to male specific Arabic gene flow into Arab Near Eastern populations.

North Africa, West and Central Mediterranean clusters

Genetic distinctness of North African populations has been shown by a number of investigations (Arredi *et al*. 2004; Bosch *et al*. 2001; Flores *et al*. 2004; Rosser *et al*. 2000). This uniqueness seems to be only partially related to SubSaharan gene flow, as little introgression from this area has been suggested (Salas *et al*. 2004; Bosch *et al*. 2001). North African populations show little Y chromosomal variation, a result interpreted as a consequence of the Neolithic expansion (Arredi *et al*. 2004). North African populations have E3b frequencies above 50% (Table 1 and Bosch *et al*. 2001; Arredi *et al*. 2004), while this hg is not above 29% in other Mediterranean and European samples (Table 1, Semino *et al*. 2000, 2004; Rosser *et al*. 2000). The E3b North African types are mostly derived from the M81 marker, while European E3b types are defined by the M78 marker (both not included in our UEP panel) (Semino *et al*. 2004).

Very limited North African gene flow to Europe seems to have occurred (Bosch *et al*. 2001). In order to quantitatively estimate the North African male genetic introgression on the northern shores of the Mediterranean basin, we used likelihood and least-square approaches, as implemented in the LEA and ADMIX software (Chikhi *et al*. 2001; Bertorelle & Excoffier, 1998; Dupanloup & Bertorelle, 2001). We selected Moroccan Berbers/Basques (Bosch *et al*. 2000, 2001) and Tunisian/Southern Italians, respectively, as possible representatives of the source populations for the Iberian and Sicilian samples. Hpt 44, modal in North Africa (see appendix and Bosch *et al*. 2000; 2001; Arredi *et al*. 2004) has been considered as separate from the rest of the E3b chromosomes. The median likelihood

 © University College London 2005

estimations ranged between 0.15 and 0.271, while the ADMIX values were between 0.004-0.21. The two different approaches support a scenario with limited genetic contribution of North African populations, in agreement with their independent clustering in the STRUCTURE analysis, and as suggested also by mtDNA and Y chromosome analysis (Figure 3a, 3b; Plaza *et al*. 2003) and in agreement with previous data (Bosch *et al*. 2001; Cruciani *et al*. 2004).

Of interest is the genetic separation that West Mediterranean samples from Iberia display vs. Central and East Mediterranean samples, as shown in this study by Y chromosome SNPs and STRs analyses and by autosomal data (Rosenberg *et al*. 2002). Investigation of the mtDNA distribution of genetic variation instead seems to support a more homogeneous situation for European Mediterranean populations. Higher female than male gene flow across populations and/or difference in population sizes of breeding individuals between the two genders have been suggested as possible explanations for this observation (Seielstad *et al*. 1998; Dupanloup *et al*. 2003). The observed autosomal independent clustering (Rosenberg *et al*. 2002) could possibly be due to unique characteristics of the Basque population, but not of all of Iberia, and reflects drift occurring in this group. A more exhaustive sampling and genotyping of multiple genetic systems for the same individuals from populations from this area and from other European populations would help to clarify this issue.

Conclusions

The significant genetic structuring of populations facing the Mediterranean basin into three groupings, Near Eastern Arab, Mediterranean and North African, is related to the demographic processes that have occurred since first populating the area. The distribution of Neolithic technologies was probably paralleled by demographic expansion in the Mediterranean basin, and subsequent westward migration by Phoenicians and Greeks contributed to the distribution of Y chromosome types of most likely Near East origin. The Arab conquest in particular appears to have had a dramatic influence on the East and South Mediterranean coasts, with differential sex-related gene flow playing a major role in the distribution of genetic variation. The presence of Arab Y chromosome lineages in the Middle East suggests that most have experienced substantial gene flow from the Arabian peninsula. This result raises the issue of the correctness of identifying all Near Eastern populations as reliable representations of the original Neolithic groups that expanded from the Middle East towards the European peninsula.

The identified genetic structure raises important issues not only for historical but also for medical genetics. Differential distribution of malaria resistance variants among Mediterranean populations is well known, and thought in part to be the result of local selection pressure (Tishkoff *et al*. 2001;Verrelli *et al*. 2002). The presence of structure as identified by multiple genetic systems suggests that other polymorphisms might be not randomly distributed as a result of the indicated demographic phenomena. In particular, variants involved in the response to drug metabolism are important candidates for such investigation, and indeed some of them, for example the *CYP2D6* and *PKU* genes, show significant differences across European and Mediterranean populations (Bradford, 2002; Cali' *et al*. 1997). More detailed sampling and investigation focused on this area will be critical for extensive evaluation of the degree of population structuring present, and the demographic processes that have shaped them.

Acknowledgements

We would like to thank the following people: Julia Abernethy and Fiona Gratrix helped in genotypings; Almut Nebel and Elena Bosch kindly provided data in an electronic format; Prof. SantaCruz provided a subset of the Sardinian samples; Jim Wilson and Martin Richards commented on an earlier version of the manuscript; Henry Harpending gave permission to use POPSTR software. G.L., V.D. and A.M. would like to additionally thank Nabiha Salem and Eliane Choueiry. V.L.P. was supported by a grant PRIN-MIUR 2002 (N.2002063871); A.N. by grant PRIN-MIUR 2003. D.B.G. is a Royal Society/Wolfson Research Merit Award holder.

References

Abun-Nasr, J. M. (1987) *A history of the Maghrib in the Islamic Period*. Cambridge: Cambridge University Press.

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

© University College London 2005

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

Arredi, B., Poloni, E. S., Paracchini, S. *et al.* (2004) A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. *Am J Hum Genet* **75**, 338–345.

Barbujani, G., Bertorelle, G. & Chikhi, L. (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. *Am J Hum Genet* **62**, 488–492.

Barbujani, G. & Goldstein, D. B. (2004) Africans and Asians abroad: genetic diversity in Europe. *Annu Rev Genomics Hum Genet* **5**, 119–150.

Bertorelle, G. & Excoffier, L. (1998) Inferring admixture proportions from molecular data. *Mol Biol Evol* **15**, 1298–1311.

Behar, D. M., Thomas, M. G., Skorecki, K. *et al.* (2003) Multiple origins of Ashkenazi Levites: Y chromosome evidence for both Near Eastern and European ancestries. *Am J Hum Genet* **73**, 768–779.

Bosch, E., Calafell, F., Perez-Lezaun, A. *et al.* (2000) Y chromosome STR haplotypes in four populations from northwest Africa. *Int J Legal Med* **114**, 36–40.

Bosch, E., Calafell, F., Comas, D., Oefner, P. J., Underhill, P. A. & 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.

Bradford, L. D. (2002) CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. *Pharmacogenomics* **3**, 229–243.

Braudel, F. (2002) *The Mediterranean in the Ancient*. London: World Penguin Books

Burn, A. R. (1990) *The Penguin History of Greece*. pp. 83–103. London: Penguin Books.

Cali, F., Dianzani, I., Desviat, L. R. *et al.* (1997) The STR252-IVS10nt546-VNTR7 phenylalanine hydroxylase minihaplotype in five Mediterranean samples. *Hum Genet* **100**, 350–355.

Capelli, C., Wilson, J. F., Richards, M. *et al.* (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am J Hum Genet* **68**, 432–443.

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

Chikhi, L., Bruford, M. W. & Beaumont, M. A. (2001) Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. *Genetics* **158**, 1347–1362.

Chikhi, L., Nichols, R. A., Barbujani, G. & Beaumont, M. A. (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci U S A* **99**, 11008–11013.

Ciminelli, B. M., Pompei, F., Malaspina, P. *et al.* (1995) Recurrent simple tandem repeat mutations during human Y-chromosome radiation in Caucasian subpopulations. *J Mol Evol* **41**, 966–973.

Cinnioglu, C., King, R., Kivisild, T., Kalfoglu, E., Atasoy, S., Cavalleri, G. L., Lillie, A. S., Roseman, C. C., Lin, A. A., Prince, K., Oefner, P. J., Shen, P., Semino, O. & Cavalli-Sforza, L. L., Underhill, P. A. (2004) Excavating Y-chromosome haplotype strata in Anatolia. *Hum Genet* **114**, 127–148.

Comas, D., Calafell, F., Benchemsi, N. *et al.* (2000) Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Hum Genet* **107**, 312–319.

Cruciani, F., La Fratta, R., Santolamazza, P. *et al.* (2004) Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. *Am J Hum Genet* **74**, 1014–1022.

Cunliffe, B. (2001) *The Oxford illustrated history of Prehistoric Europe*. Oxford: Oxford University Press.

Di Giacomo, F., Luca, F., Popa, L. O. *et al.* (2004) Y Chromosomal Haplogroup J as a signature of the post-Neolithic colonization of Europe. *Hum Genet* **115**, 357–371.

Dupanloup, I. & Bertorelle, G. (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Mol Biol Evol* **18**, 672–675.

Dupanloup, I., Pereira, L., Bertorelle, G., Calafell, F., Prata, M. J., Amorim, A. & Barbujani, G. (2003) A recent shift from polygyny to monogamy in humans is suggested by the analysis of worldwide Y-chromosome diversity. *J Mol Evol* **57**, 85–97.

Dussaud, R. (1955) *La penetration des Arabes en Syrie avant l'-Islam*. Paris: P Geuthner.

Finley, M. L. (1968) *A History of Sicily: Ancient Sicily to the Arab Conquest*. London: Viking Press.

Flores, C., Maca-Meyer, N., Gonzalez, A. M., Oefner, P. J., Shen, P., Perez, J. A., Rojas, A., Larruga, J. M. & Underhill, P. A. (2004) Reduced genetic structure of the Iberian peninsula revealed by Y-chromosome analysis: implications for population demography. *Eur J Hum Genet* **10**, 855–863.

Goncalves, R., Freitas, A., Branco, M., Rosa, A., Fernandes, A. T., Zhivotovsky, L. A., Underhill, P. A., Kivisild, T. & Brehm, A. (2005) Y-chromosome Lineages from Portugal, Madeira and Acores Record Elements of Sephardim and Berber Ancestry. *Ann Hum Genet* **69**, 443–454.

Hammer, M. F., Redd, A. J., Wood, E. T. *et al.* (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.

Hitti (1990) *The Arabs: a short history*. Gateway Editions. Washington, DC.

Luis, J. R., Rowold, D. J., Regueiro, M., Caeiro, B., Cinnioglu, C., Roseman, C., Underhill, P. A., Cavalli-Sforza,

 © University College London 2005

L. L. & Herrera, R. J. (2004) The Levant versus the Horn of Africa: evidence for bidirectional corridors of human migrations. *Am J Hum Genet*. **74**, 532–544.

Malaspina, P., Tsopanomichalou, M., Duman, T. *et al.* (2001) A multistep process for the dispersal of a Y chromosomal lineage in the Mediterranean area. *Ann Hum Genet* **65**, 339–349.

Nebel, A., Filon, D., Weiss, D. A. *et al.* (2000) High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. *Hum Genet* **107**, 630–641.

Nebel, A., Filon, D., Brinkmann, B., Majumder, P. P., 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.

Plaza, S., Calafell, F., Helal, A. *et al.* (2003) Joining the pillars of Hercules: mtDNA sequences show multidirectional gene flow in the western Mediterranean. *Ann Hum Genet* **67**, 312–328.

Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.

Quintana-Murci, L., Veitia, R., Fellous, M., Semino, O. & Poloni, E. S. (2003) Genetic structure of Mediterranean populations revealed by Y-chromosome haplotype analysis. *Am J Phys Anthropol* **121**, 157–171.

Ricci, L. (1984) L'expansion de l'arabie meridionale. In (Ed. J. Chelhod) *L'Arabie du Sud. Maisonneuve et Larose*, Paris, vol 1, pp 239–248.

Richards, M., Macaulay, V., Hickey, E. *et al.* (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* **67**, 1251–1276.

Richards, M., Macaulay, V., Torroni, A. & Bandelt, H. J. (2002) In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet* **71**, 1168–1174.

Richards, M., Rengo, C., Cruciani, F. *et al.* (2003) Extensive Female-Mediated Gene Flow from Sub-Saharan Africa into Near Eastern Arab Populations. *Am J Hum Genet* **72**, 1058–1064.

Romano, V., Cali, F., Ragalmuto, A. *et al.* (2003) A. Autosomal microsatellite and mtDNA genetic analysis in Sicily (Italy). *Ann Hum Genet* **67**, 42–53.

Rosser, Z. H., Zerjal, T., Hurles, M. E. *et al.* (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* **67**, 1526–1543.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A. & Feldman, M. W. (2002) Genetic structure of human populations. *Science* **298**, 2381–2385.

Salas, A., Richards, M., Lareu, M. V., Scozzari, R., Coppa, A., Torroni, A., Macaulay, V. & Carracedo, A. (2004) The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* **74**, 454–465.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN ver. 2.00, A software for Population Genetic Data Analysis*. Geneva: Genetics and Biometery Laboratory, University of Geneva.

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

Semino, O., Passarino, G., Oefner, P. J. *et al.* (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159.

Semino, O., Magri, C., Benuzzi, G. *et al.* (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.

Simoni, L., Calafell, F., Pettener, D., Bertranpetit, J. & Barbujani, G. (2000) Geographic patterns of mtDNA diversity in Europe. *Am J Hum Genet* **66**, 262–278.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462.

Stringer, C. & Gamble, C. (1993) *In Search of the Neanderthals: Solving the Puzzle of Human Origins*. pp. 96–122. Thames and Hudson.

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

Thomas, M. G., Parfitt, T., Weiss, D. A. *et al.* (2000) Y chromosomes travelling 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., Weale, M. E., Jones, A. L. *et al.* (2002) Founding Mothers of Jewish Communities: Geographically Separated Jewish Groups Were Independently Founded by Very Few Female Ancestors. *Am J Hum Genet* **70**, 1411–1420.

Tishkoff, S. A., Varkonyi, R., Cahinhinan, N. *et al.* (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* **293**, 455–462.

Torroni, A., Bandelt, H. J., Macaulay, V. *et al.* (2001) A signal, from human mtDNA, of postglacial recolonization in Europe. *Am J Hum Genet* **69**, 844–852.

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

© University College London 2005

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

Verrelli, B. C., McDonald, J. H., Argyropoulos, G. *et al.* (2002) Evidence for balancing selection from nucleotide sequence analyses of human G6PD. *Am J Hum Genet* **71**, 1112–1128.

Zvelebil, M. & Zvelebil, K. V. (1988) Agricultural transition and Indo-European dispersals. *Antiquity* **62**, 574–583.

Appendix

List of haplotypes (STRs plus hgs) and their occurrence in the populations investigated in this study. Codes as in Table A1. *Hg was identified as described in the text.

Received: 9 March 2005
Accepted: 18 July 2005

Table A1

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
1	9	13	11	13	24	10	E3b						1						1
2	9	14	12	15	23	11	F*(xI,J,K)					1							1
3	10	12	11	15	23	10	J2							1					1
4	10	13	11	13	24	10	E3b						2						2
5	10	13	13	15	24	12	R1*(xR1a1)							1					1
6	10	14	11	15	22	9	F*(xI,J,K)							1					1
7	11	13	11	15	22	11	A					1							1
8	11	13	13	14	23	10	K*(xP,N3)			1									1
9*	11	14	10	16	22	10	A	1											1
10*	11	14	11	13	24	10	A	1											1
11	12	11	14	14	22	10	K*(xP,N3)			1									1
12	12	11	14	14	25	10	K*(xP,N3)				1								1
13	12	12	11	13	23	11	E3b	1											1
14*	12	12	11	13	24	9	E3b		1										1
15	12	12	11	13	24	10	E3b							1					1
16	12	12	11	14	24	10	J2						1						1
17	12	12	11	14	24	11	R1*(xR1a1)							1					1
18	12	12	11	15	23	10	F*(xI,J,K)	2											2
19	12	12	13	13	24	10	R1*(xR1a1)						1						1
20	12	12	13	13	26	11	R1*(xR1a1)								1				1
21	12	12	13	14	23	10	K*(xP,N3)										1		1
22	12	12	13	14	23	10	R1*(xR1a1)	2							1				3
23	12	12	13	14	23	11	R1*(xR1a1)							1					1
24	12	12	13	14	24	10	R1*(xR1a1)				1			1	1		1		4
25	12	12	13	14	24	11	R1*(xR1a1)						1			1	1		3
26	12	12	13	14	25	10	R1*(xR1a1)						1					1	2
27	12	12	13	14	25	11	R1*(xR1a1)										1		1
28	12	12	13	15	25	10	R1*(xR1a1)									1			1
29	12	12	13	15	25	11	R1*(xR1a1)						1			1			2
30	12	12	14	14	23	10	P*(xR1)				1								1
31	12	12	14	14	23	10	K*(xP,N3)				1								1
32	12	12	14	14	24	10	R1*(xR1a1)							2		1			3
33	12	12	14	15	24	11	R1*(xR1a1)							1					1
34	12	12	15	14	24	10	R1*(xR1a1)				1								1
35	12	13	10	13	24	10	E3b					1							1
36	12	13	10	14	23	10	P*(xR1)					1							1
37	12	13	10	15	22	11	F*(xI,J,K)										1		1
38	12	13	11	9	24	11	E3b								1				1
39	12	13	11	13	21	8	E3b			1									1
40	12	13	11	13	22	10	E3b						1	1					2

 © University College London 2005

Table A1 (continued)

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
41	12	13	11	13	23	9	E3b							1			2		3
42	12	13	11	13	23	10	E3b			2	1		3	3					9
43	12	13	11	13	23	11	E3b				1								1
44	12	13	11	13	24	9	E3b					1		1	4	1		15	22
45	12	13	11	13	24	10	E3b	3		2			5	6	4	3	1		24
46	12	13	11	13	24	11	E3b	1				1			1		1	1	5
47	12	13	11	13	24	12	E3b			1									1
48	12	13	11	13	25	9	E3b					1						1	2
49	12	13	11	13	25	10	E3b				1	2		3			1		7
50	12	13	11	13	25	11	E3b							1	1				2
51	12	13	11	14	22	10	F*(xI,J,K)				1	1							2
52	12	13	11	14	23	10	I*(xI1b2)										1		1
53	12	13	11	14	23	11	E3b							2					2
54	12	13	11	14	24	9	E3b											1	1
55	12	13	11	14	24	10	E3b	3					2	2	2	1	1		11
56	12	13	11	14	24	11	E3b				1								1
57	12	13	11	15	21	10	F*(xI,J,K)										1		1
58	12	13	11	15	22	9	F*(xI,J,K)					1							1
59	12	13	11	15	22	10	F*(xI,J,K)			2			1	1					4
60	12	13	11	15	22	11	F*(xI,J,K)				1								1
61	12	13	11	15	24	9	E3b											1	1
62	12	13	11	15	24	10	E3b						1	1					2
63	12	13	11	15	25	10	R1a1	1											1
64	12	13	11	15	25	10	E3b									1			1
65	12	13	11	15	25	11	R1a1							1	1			1	3
66	12	13	11	15	26	11	R1a1								1				1
67	12	13	11	16	22	10	F*(xI,J,K)								1				1
68	12	13	11	16	23	10	F*(xI,J,K)				3		1						4
69	12	13	11	16	24	10	R1a1	1											1
70	12	13	11	16	24	10	E3b							1					1
71	12	13	11	16	25	10	R1a1						1				3		4
72	12	13	11	16	25	11	R1a1			1			1						2
73	12	13	11	17	21	11	E3a										1		1
74	12	13	12	13	23	10	F*(xI,J,K)	1											1
75	12	13	12	13	24	10	E3b	1											1
76	12	13	12	13	25	10	E3b										1		1
77	12	13	12	14	23	10	F*(xI,J,K)			1			1	1					3
78	12	13	12	15	21	10	F*(xI,J,K)					1							1
79	12	13	12	15	24	11	R1*(xR1a1)									1			1
80	12	13	12	16	21	9	F*(xI,J,K)								1				1
81	12	13	12	17	22	9	E*(xE3a,E3b)										1		1
82	12	13	13	13	22	10	K*(xP,N3)										1		1
83	12	13	13	13	23	10	K*(xP,N3)	1					1	1					3
84	12	13	13	13	23	11	F*(xI,J,K)									1			1
85	12	13	13	13	23	12	R1*(xR1a1)					1							1
86	12	13	13	13	24	9	K*(xP,N3)							1					1
87	12	13	13	13	24	10	K*(xP,N3)						1						1
88	12	13	13	13	24	11	R1*(xR1a1)	1						1					2
89	12	13	13	13	25	11	R1*(xR1a1)								1				1
90	12	13	13	14	23	9	K*(xP,N3)	1											1
91	12	13	13	14	23	10	R1*(xR1a1)					1			1	3	2		7
92	12	13	13	14	23	11	R1*(xR1a1)					2	1	1	3		3		10
93	12	13	13	14	23	12	R1*(xR1a1)					1	1	1					3
94	12	13	13	14	24	9	K*(xP,N3)									1			1

© University College London 2005

Table A1 (continued)

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
95	12	13	13	14	24	10	F*(xI,J,K)				1	4			1	4	1		11
96	12	13	13	14	24	11	R1*(xR1a1)					2	4	2	2	2	3		15
97	12	13	13	14	24	12	R1*(xR1a1)				1								1
98	12	13	13	14	25	10	R1*(xR1a1)								1				1
99	12	13	13	14	25	11	R1*(xR1a1)	1				1	1				2		5
100	12	13	13	15	23	10	P*(xR1)				1								1
101	12	13	13	15	23	10	K*(xP,N3)				1								1
102	12	13	13	15	23	10	R1*(xR1a1)						1	1					2
103	12	13	13	15	23	11	K*(xP,N3)									1			1
104	12	13	13	15	23	11	R1*(xR1a1)					1				1	1		3
105	12	13	13	15	23	12	K*(xP,N3)							1					1
106	12	13	13	15	24	10	R1*(xR1a1)								1				1
107	12	13	13	15	24	11	K*(xP,N3)							1					1
108	12	13	13	15	24	11	R1*(xR1a1)			1		1	1	1	2	1	1		8
109	12	13	13	15	24	12	R1*(xR1a1)	1								1			2
110	12	13	13	16	25	10	R1a1									1			1
111	12	13	14	14	23	10	R1*(xR1a1)											1	1
112	12	13	14	14	23	11	R1*(xR1a1)					1		1			1		3
113	12	13	14	14	24	10	R1*(xR1a1)	1						1					2
114	12	13	14	14	24	11	R1*(xR1a1)					1							1
115	12	13	14	15	24	11	R1*(xR1a1)						1	1					2
116	12	13	15	13	22	10	P*(xR1)						1						1
117	12	13	15	14	23	11	R1*(xR1a1)						1						1
118	12	14	10	14	23	10	P*(xR1)				1	1							2
119	12	14	10	15	22	10	F*(xI,J,K)						2						2
120	12	14	10	15	23	10	P*(xR1)								1				1
121	12	14	11	13	23	10	E3b					1					1		2
122	12	14	11	13	24	9	E3b	1				1				1		1	4
123	12	14	11	13	24	10	E3b	2					1						3
124	12	14	11	13	25	9	E3b						1						1
125	12	14	11	14	22	10	F*(xI,J,K)					3							3
126	12	14	11	14	23	10	F*(xI,J,K)	1											1
127	12	14	11	15	21	10	F*(xI,J,K)						2				3		5
128*	12	14	11	15	22	9	F*(xI,J,K)		1										1
129	12	14	11	15	22	10	J*(xJ2)		1										1
130	12	14	11	15	22	10	F*(xI,J,K)			2			1		3	2	1		9
131	12	14	11	15	22	11	F*(xI,J,K)								2				2
132	12	14	11	15	23	9	F*(xI,J,K)									1			1
133	12	14	11	15	23	10	F*(xI,J,K)			1			1			2			4
134	12	14	11	15	23	11	F*(xI,J,K)					1							1
135	12	14	11	15	24	10	E3b							1					1
136	12	14	11	16	23	10	F*(xI,J,K)				1								1
137	12	14	11	17	22	11	F*(xI,J,K)	1											1
138	12	14	12	15	22	10	F*(xI,J,K)					1							1
139	12	14	12	15	23	9	F*(xI,J,K)											1	1
140	12	14	12	16	24	10	F*(xI,J,K)						1						1
141	12	14	13	13	24	11	R1*(xR1a1)											1	1
142	12	14	13	14	23	10	K*(xP,N3)								1				1
143	12	14	13	14	23	10	P*(xR1)										1		1
144	12	14	13	14	24	10	R1*(xR1a1)						1				1		2
145	12	14	13	14	24	11	R1*(xR1a1)								1				1
146	12	14	13	15	22	10	K*(xP,N3)					1							1
147	12	14	14	14	25	10	K*(xP,N3)	1											1

 © University College London 2005

Table A1 (continued)

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
148	12	15	11	13	23	10	E3b							1	1				2
149	12	15	11	14	21	10	F*(xI,J,K)					1							1
150	12	15	11	15	23	10	F*(xI,J,K)					2							2
151	12	15	11	16	22	10	F*(xI,J,K)	1											1
152*	12	15	11	16	23	10	F*(xI,J,K)		1										1
153	12	15	11	16	24	10	R1a1							1					1
154	12	15	13	14	23	10	I*(xI1b2)									1			1
155	12	15	13	14	23	10	F*(xI,J,K)									1			1
156	13	12	11	14	24	9	C								1				1
157	13	12	11	16	22	10	R1*(xR1a1)										1		1
158	13	12	11	17	22	10	I*(xI1b2)						1						1
159	13	12	13	14	25	11	R1*(xR1a1)										6		6
160	13	12	13	14	26	10	R1*(xR1a1)										1		1
161	13	12	13	14	26	11	R1*(xR1a1)										1		1
162	13	12	13	15	23	10	J2							1					1
163	13	13	11	13	24	11	E3b	1											1
164	13	13	11	14	23	10	E3b									2			2
165	13	13	11	14	24	10	E3b				2								2
166	13	13	11	15	22	10	F*(xI,J,K)									1			1
167	13	13	11	15	23	10	I*(xI1b2)					1							1
168	13	13	11	15	23	10	I1b2					1							1
169	13	13	11	15	23	11	I1b2					1							1
170	13	13	11	15	24	10	E3b			1									1
171	13	13	11	15	24	11	I*(xI1b2)				1			1					2
172	13	13	11	15	25	10	I1b2					1							1
173	13	13	11	15	25	11	I*(xI1b2)						1						1
174	13	13	11	16	23	10	I1b2					5							5
175	13	13	11	16	23	11	I*(xI1b2)	1				1							2
176	13	13	11	16	24	9	I*(xI1b2)										1		1
177	13	13	11	16	24	10	I1b2					1							1
178	13	13	11	16	24	10	I*(xI1b2)	1									2		3
179	13	13	11	16	24	11	I*(xI1b2)	1										1	2
180	13	13	11	16	25	10	I*(xI1b2)										1		1
181	13	13	11	17	23	10	I1b2					5							5
182	13	13	11	17	24	10	I1b2					1		1					2
183	13	13	11	17	24	11	F*(xI,J,K)								1				1
184	13	13	11	17	25	10	I*(xI1b2)								1				1
185	13	13	12	15	23	10	I*(xI1b2)					1					1		2
186	13	13	12	16	24	11	I*(xI1b2)				1								1
187	13	13	13	14	23	11	R1*(xR1a1)					1							1
188	13	13	13	14	24	10	R1*(xR1a1)						1						1
189	13	14	11	15	22	10	F*(xI,J,K)					1						1	2
190	13	14	11	15	22	11	F*(xI,J,K)							1					1
191	13	14	11	15	24	10	I*(xI1b2)									1			1
192	13	14	11	15	24	11	I*(xI1b2)	1											1
193	13	14	11	16	23	10	I1b2					4							4
194	13	14	12	14	22	10	I*(xI1b2)							1					1
195	13	14	12	15	23	10	I*(xI1b2)					1		1	3	1			6
196	13	14	12	16	23	10	I*(xI1b2)								1				1
197	13	14	13	13	24	11	R1*(xR1a1)								1				1
198	13	14	13	15	22	10	I*(xI1b2)								1		4		5
199	13	14	13	15	22	11	I*(xI1b2)										1		1
200	13	14	13	16	22	10	I*(xI1b2)								2				2

© University College London 2005

Table A1 (continued)

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
201	14	12	11	14	23	9	J2									1			1
202	14	12	11	14	23	10	J2	1											1
203	14	12	11	14	23	11	J2		2										2
204	14	12	11	14	24	10	J2	1		2				1			1		5
205	14	12	11	14	25	10	J2				1								1
206	14	12	11	15	23	10	J2	1				1						2	4
207	14	12	13	14	24	10	R1*(xR1a1)										1		1
208	14	13	11	13	24	10	E3b						1						1
209	14	13	11	14	22	10	I*(xI1b2)								1				1
210	14	13	11	14	25	10	E3b							1					1
211	14	13	11	15	23	10	I1b2						1						1
212	14	13	11	16	23	10	I*(xI1b2)								1				1
213	14	13	11	16	24	11	I*(xI1b2)								1				1
214	14	13	11	17	23	10	I*(xI1b2)					2							2
215	14	13	12	15	22	11	I*(xI1b2)						1						1
216	14	13	13	14	23	10	R1*(xR1a1)										3		3
217	14	14	12	15	23	11	I*(xI1b2)	1											1
218	15	12	11	13	22	9	J2										1		1
219	15	12	11	13	24	10	J*(xJ2)							2					2
220	15	12	11	14	20	10	J2										2		2
221	15	12	11	14	22	9	J2				1								1
222	15	12	11	14	22	10	J2	1			2	1		1	1	1			7
223	15	12	11	14	22	11	J2								1				1
224	15	12	11	14	23	9	J2	1						2					3
225	15	12	11	14	23	10	J2	7		2	4		2	6		3			24
226	15	12	11	14	23	11	J*(xJ2)	2						1		4	1		8
227	15	12	11	14	24	10	J2	1	2			1	1		3		3		11
228	15	12	11	14	24	11	J2					1			1				2
229	15	12	11	14	25	11	J2	1											1
230	15	12	11	15	22	10	J2				1		1						2
231	15	12	11	15	23	10	J2	1						3	1		4		9
232	15	12	11	15	24	10	J2	1		1		1	1						4
233	15	12	11	15	24	11	J2						1						1
234	15	12	11	15	25	10	J*(xJ2)								1				1
235	15	12	11	16	23	10	J2	2									1		3
236	15	12	11	16	24	11	J2				3								3
237	15	12	11	16	24	12	J2				1								1
238	15	12	12	15	24	10	J2					1							1
239	15	12	14	14	24	10	J*(xJ2)		1										1
240	15	12	15	14	23	10	J*(xJ2)					1							1
241	15	13	11	14	22	10	J2			3			1						4
242	15	13	11	14	23	10	J2			1						2			3
243	15	13	11	14	24	11	J2							1					1
244	15	13	11	15	22	10	J2							1					1
245	15	13	11	15	22	11	J2							1			3		4
246	15	13	11	15	23	10	F*(xI,J,K)					1							1
247	15	13	11	15	23	10	J2							1		1	1		3
248	15	14	11	15	23	10	F*(xI,J,K)					1							1
249	16	12	11	13	24	10	J2							1					1
250	16	12	11	14	22	10	J2				1								1
251	16	12	11	14	22	10	J*(xJ2)			1									1
252	16	12	11	14	22	10	J2									1			1
253	16	12	11	14	23	9	J*(xJ2)			1									1

 © University College London 2005

Table A1 (continued)

Hpt	DYS 388	DYS 393	DYS 392	DYS 19	DYS 390	DYS 391	Hg	CYP	DJE	MLE	CLE	SAR	SIT	ESC	NWS	SWS	MAL	TUN	n
254	16	12	11	14	23	10	J*(xJ2)		3										3
255	16	12	11	14	23	10	J2						1	1	1			1	4
256	16	12	11	14	23	10	J*(xJ2)	1	1	2			3		2		5		14
257	16	12	11	14	23	11	J2											1	1
258	16	12	11	14	23	11	J*(xJ2)	1				1					1		3
259	16	12	11	14	23	12	J*(xJ2)										1		1
260	16	12	11	14	24	10	J*(xJ2)				1								1
261	16	12	11	15	23	9	J2		2	1				3		1	1		8
262	16	12	11	15	23	10	J*(xJ2)				1	1							2
263	16	12	11	15	23	10	J2		1						2				3
264	16	12	11	15	24	9	J2		1										1
265	16	12	11	15	24	9	J*(xJ2)		2										2
266	16	12	11	15	24	10	J2						1			1			2
267	16	12	11	16	23	9	J2						1						1
268	16	12	12	14	23	10	J2	1				1							2
269	16	12	12	14	23	10	J*(xJ2)			1								1	2
270	16	13	11	15	22	10	J*(xJ2)					1							1
271	16	14	11	14	23	10	J*(xJ2)			1									1
272	17	12	11	13	23	11	J*(xJ2)	1											1
273	17	12	11	14	22	11	J*(xJ2)			1									1
274	17	12	11	14	23	10	J*(xJ2)			1	1							2	4
275	17	12	11	14	23	11	J2				1								1
276	17	12	11	14	23	11	J*(xJ2)			2				1				3	6
277	17	12	11	14	24	10	J2	1											1
278	17	12	11	14	24	10	J*(xJ2)				1								1
279	17	12	11	15	23	10	J2	1											1
280	17	12	11	15	23	10	J*(xJ2)							1					1
281	17	12	11	15	23	11	J*(xJ2)			2									2
282	17	12	11	16	25	10	J2										1		1
283	17	12	11	17	23	9	J2					1							1
284	17	12	11	17	23	10	J*(xJ2)							1					1
285	17	13	11	14	23	10	J2	1											1
286	18	12	11	14	22	9	J*(xJ2)											1	1
287	18	12	11	14	22	11	J*(xJ2)											1	1
288	18	12	11	14	23	10	J2	1											1
289	18	12	11	14	23	11	J*(xJ2)											1	1
290	18	12	11	15	23	10	J*(xJ2)												1
								65	19	39	43	81	68	87	70	55	90	39	656

© University College London 2005