ARTICLE IN PRESS

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution xxx (2003) xxx–xxx

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations

N. Al-Zahery,[a,b] O. Semino,[a] G. Benuzzi,[a] C. Magri,[a] G. Passarino,[c] A. Torroni,[a] and A.S. Santachiara-Benerecetti[a,*]

[a] *Dipartimento di Genetica e Microbiologia "A. Buzzati Traverso", Università di Pavia, Via Ferrata, 1, 27100 Pavia, Italy*
[b] *Department of Biotechnology, College of Science, University of Baghdad, Baghdad, Iraq*
[c] *Dipartimento di Biologia Cellulare, Università della Calabria, Cosenza, Italy*

Received 29 May 2002; revised 11 December 2002

Abstract

Analyses of mtDNA and Y-chromosome variation were performed in a sample of Iraqis, a scarcely investigated population of the "Fertile Crescent." A total of 216 mtDNAs were screened for the diagnostic RFLP markers of the main Eurasian and African haplogroups. A subset of these samples, whose HVS-I sequences were previously obtained, was also examined by high-resolution restriction analysis. The Y-chromosome variation was investigated in 139 subjects by using 17 biallelic markers and the 49a,f/*Taq* I system. For both uniparental systems, the large majority of the haplogroups observed in the Iraqi population are those (H, J, T, and U for the mtDNA, and J(xM172) and J-M172 for the Y chromosome) considered to have originated in the Middle East and to have later spread all over Western Eurasia. However, about 9% of the mtDNAs and 30% of the Y-chromosomes most likely represent arrivals from distant geographic regions. The different proportion of long-range genetic input observed for the mtDNA and the Y chromosome appears to indicate that events of gene flow to this area might have involved mainly males rather than females.

1. Introduction

Iraq is an Arabian country bordered by the Arabian gulf, Kuwait, Saudi Arabia (South), Jordan and Syria (West), Turkey (North), and Iran (East) (Fig. 1). Iraq is composed of a mountainous region in the Northeast and a vast desert in the Southwest; in between is the heart of the country, a fertile low land region irrigated by the Tigris and Euphrates rivers, corresponding to the ancient Mesopotamia. The Iraqi population, ca. 21.5 million individuals (July 1995), consists of 75–80% Arabs, 15–20% Kurds, and 5% Assyrians, Turkmen, and others. The official language is Arabic, a Semitic language belonging to the Afro-Asiatic family, but other languages are also spoken: Kurdish (Indo-European), Assyrian (Semitic), and Armenian (Indo-European).

Modern humans are likely to have been present in this area from the time of the exit from Africa around 60,000 years ago, but little information is available about the Palaeolithic period. As for Neolithic times, Iraq represents one of the most important areas for the development and diffusion of agriculture dating from 10,000 years ago onward as testified by many archaeological remains at the Shanidar and Karim Shahir sites (Rashed, 1997). In addition, this region was the source of several ancient civilizations (Sumerians, who first introduced the use of writing, Babylonians, Assyrians, and Caldeans) that profoundly affected other populations of the Middle East, Europe, and elsewhere.

In spite of the importance of this region, genetic studies on the Iraqi population are scarse, aged, and generally restricted to analyses of classical markers (Cavalli-Sforza et al., 1994 and references therein), thus not allowing a dissection of its genetic structure. Analyses of mtDNA and Y-chromosome variation have

* Corresponding author. Fax: +39-0382-528496.
E-mail address: santa@ipvgen.unipv.it (A.S. Santachiara-Benerecetti).

1055-7903/03/$ - see front matter © 2003 Published by Elsevier Science (USA).
doi:10.1016/S1055-7903(03)00039-3

Fig. 1. Map of Iraq.

indicated the central role of the Middle Eastern area in the dispersal of human populations by revealing that it was the probable homeland of numerous mtDNA and Y-chromosome haplogroups (Macaulay et al., 1999; Quintana-Murci et al., 2001; Richards et al., 1998; Semino et al., 2000a; Torroni et al., 1998). However, while extensive studies were carried out in several populations of this region (Comas et al., 1998; Hammer et al., 2000; Nebel et al., 2001; Quintana-Murci et al., 2001; Ritte et al., 1993; Salem et al., 1996; Santachiara-Benerecetti et al., 1993), only few components of mtDNA and Y-chromosome Iraqi gene pool have been evaluated (Quintana-Murci et al., 1999b; Richards et al., 2000). Thus, to increase the knowledge of the genetic structure of that population and to better understand the relationships between the Middle East, Asia Minor, and Europe, we have investigated the extent and nature of mtDNA and Y-chromosome variation in a sample of Iraqi subjects.

2. Materials and methods

2.1. Subjects

Blood samples were collected in Baghdad from 216 unrelated adult males (178 Arabs, 25 Assyrians, and 13 Kurds) whose origin was ascertained, at least up to four generations, by oral interview, and who were from different villages and towns along the Tigris and Euphrates rivers. Genomic DNA was extracted from buffy-coats by using standard procedures.

2.2. MtDNA analysis

A total of 216 mtDNAs were analyzed for the diagnostic RFLP markers of the most common European, Asian, and African haplogroups (Macaulay et al., 1999; Saillard et al., 2000; Torroni et al., 1996). In addition, a subset ($N = 52$) was examined by high-resolution RFLP analysis using the PCR primers, restriction enzymes, and conditions described by Torroni et al. (1999). These RFLP data were combined with the corresponding sequence data from the first hypervariable segment (HVS-I) of the control-region previously included in Richards et al. (2000).

2.3. Y-chromosome DNA analysis

A sample of 139 individuals (115 Arabs, 19 Assyrians, and 5 Kurds) was examined for 17 biallelic markers, namely: 12f2 [DYS11] (Casanova et al., 1985), YAP [DYS287] (Hammer, 1994), DYS257 (Hammer et al., 1998), M9 (Underhill et al., 1997) and, in a hierarchical way, for M2 corresponding to DYS271 (Seielstad et al., 1994), M17 (Underhill et al., 1997), M35 (Underhill et al., 2000), M70, M74, M89, M170, M172, M173, M175, M201 (Shen et al., 2000), P2 (former PN2: Hammer et al., 1997), and M269 (Cruciani et al., 2002). These markers define the main groups of the Y chromosome phylogeny (Underhill et al., 2001 and, for a review, see The Y Chromosome Consortium—The YCC, 2002). The 49a,f/*Taq*I [DYS1] RFLPs (Ngo et al., 1986), which can provide additional information at the microevolutionary level, were also examined. The 12f2 and 49a,f/*Taq*I polymorphisms were analyzed according to Passarino et al. (1998). The YAP and DYS257 mutations were analyzed according to Hammer and Horai (1995) and Hammer et al. (1998), respectively. The mutations M2, M17, P2, M70, M74, M172, M173 M175, and M201 were detected by DHPLC according to Underhill et al. (2001). The mutations M9, M89, and M269 were typed through PCR/RFLP assay: M9, by using the primers described by Underhill et al. (2001) and the restriction enzyme *Hin*fI, whereas M89 and M269 were examined as reported by Akey et al. (2001) and Cruciani et al. (2002), respectively.

2.4. Data analysis

Phylogeny construction of the 52 mtDNAs examined for both high resolution restriction analysis and sequencing was performed by hand and confirmed by using the Network 2.0e (Bandelt et al., 1995; for the reduced median network).

Principal component (PC) analysis was performed on haplogroup frequencies by using Excel implemented by the XL-stat program. In addition to Iraqi data, for mtDNA analysis we used data from Italian (Torroni

et al., 1997), Armenian, Georgian (Tambets et al., 2000), Palestinian, Syrian (Richards et al., 2000), Norwegian (Passarino et al., 2002), Iranian, Arabian, and Anatolian (Kivisild et al., 2003, in press) populations; for the Y-chromosome analysis, we used populations data from Europe, the Middle East, and Caucasus (Semino et al., 2000a), from Ethiopia (Semino et al., 2002) and from North Africa (Cruciani et al., 2002). The proportion p_{ij} of the ith haplogroups in the jth population was used. These matrices were submitted to the PC analysis after the standardization of the data.

3. Results

3.1. MtDNA variation

Table 1 shows the haplogroup frequencies obtained by screening all Iraqi mtDNAs for the diagnostic RFLP markers. Their distribution is similar to that of Iran and, in general, to the Middle Eastern populations whereas it substantially differs from that observed in Arabia. Haplogroups HV, H, V, J, T, K, U, I, X, and W encompass about 78% of the Iraqi subjects. This overall frequency is in the range of the other Middle Eastern population values (75.1–80.4%) which are closer to Europe (>90%) than to the Arabian peninsula (60.4%). By considering the single haplogroups, it should be noted that the incidence of haplogroup H (20%) is much lower than in Europeans (33–50%) but higher than in Central Asians (14%). On the contrary, the incidence of haplogroup U (about 20%) is rather close to those observed in Trans-Caucasian and European populations. As for the other Western Eurasian haplogroups, HV, J, and T are all very well represented (about 10% each), whereas K, I, X, and W are rather uncommon and generally do not exceed a frequency of 3%.

Gene flow to Iraq from Europe, Africa, and Central Asia is revealed by the presence of some particular haplogroups. A low frequency of haplogroup V (Torroni et al., 2001) is observed in Iraq (0.5%). This finding confirms gene flow from Europe as proposed by Richards et al. (2000). The presence (4.2%) of haplogroups L1 and L2 identifies the mtDNAs of African origin, whereas mtDNAs M (1.4%) and B (0.9%) are most likely due to gene flow from Central Asia.

A feature that appears to distinguish Middle Eastern populations from other West Eurasian populations is the presence of haplogroup pre-HV mtDNAs. Its frequency (4.2%) in Iraq is similar to that observed in other Arabic-speaking populations which, however, have a frequency lower than that (15.2%) found in Arabia.

The nature and extent of the Iraqi mtDNA variation was further evaluated by performing high-resolution RFLP analysis in a sample of 52 subjects chosen from those sequenced for HVS-I (Richards et al., 2000). These samples (Table 2) included all the most frequent haplogroups. The phylogenetic relationships of the haplotypes (Ht) obtained by combining the high-resolution RFLP data and the HVS-I data are illustrated in Fig. 2. The topology of the phylogeny reveals three main star-like clades. One clade is defined by the site *Hae*III at np 11,718 and encompasses pre-HV, HV, and H. A second clade is characterized by the *Hin*fI site at np 12,308 and includes K and U. The third major clade is defined by the *Nla*III site at np 4216 and contains haplogroups J and T.

3.2. Y-chromosome variation

Bi-allelic markers. Y chromosomes from 139 Iraqis were examined by using 17 bi-allelic markers. These markers, which are considered to have occurred only once during human evolution, define the main haplogroups of the human Y-chromosome genealogy (The YCC, 2002). Since no significant differences among the Iraqi groups were observed (Assyrian and Kurd sample sizes are very small), all data were pooled. The results are reported in Fig. 3.

Iraqi Y-chromosomes fall into haplogroups, E, F, G, I, J, K, and R. Haplogroup E is mainly African, but its clade E-M35 is also present in Europe where it is believed to have arrived from the Middle East in Neolithic times (Hammer et al., 1998; Semino et al., 2000a). The frequency of this haplogroup in Iraq, with the exclusion of the two samples belonging to the sub-Saharan clade E-M2, is 10.8% and falls into the lower edge of the Middle Eastern population range (10–30%) (Hammer et al., 2000; Nebel et al., 2001; Quintana-Murci et al., 2001; Semino et al., 2000a; Underhill et al., 2000; Wells et al., 2001).

Haplogroups G-M201 and I-M170 are observed only in 3 and 1 subjects, respectively, whereas haplogroup J-12f2 is the most represented in the Iraqi Y chromosomes (58.3%). Haplogroup K shows a frequency of 8.6% and, interestingly, a major portion is represented by the sub-clade K-M70 that has been found at a comparable incidence in Ethiopians (Cruciani et al., 2002; Semino et al., 2002). Elsewhere, this sub-clade was mostly found as single observations and widely scattered in several continents (Semino et al., 2000a; Underhill et al., 2000).

Haplogroup R corresponds to the cluster defined by the M173 mutation. This is considered an ancient marker brought by or arisen in the first group of *Homo sapiens sapiens* who populated Eurasia. It is present at a high frequency (more than 50%) in Eurasia (Bosch et al., 2001; Karafet et al., 2001; Semino et al., 2000a; Underhill et al., 2000; Wells et al., 2001) where it has differentiated into various branches. Its two sub-clades, the R-M17 and R-P25 which has been recently found associated with the M269 mutation in almost all the European tested samples (Santachiara-Benerecetti,

Table 1
MtDNA haplogroup frequencies in Iraqi and some comparison populations

Population/ Region	Sample size	Haplogroup frequency (%)																
		pre-HV	HV	H	V	J	T	K	U	I	X	W	B	M	Asian	L1	L2	African	Others
															A–G,M,N9			L1–L3	
Iraqi	**216**	**4.2**	**10.6**	**19.9**	**0.5**	**9.3**	**8.8**	**3.2**	**19.0**	**1.9**	**2.8**	**1.9**	**0.9**	**1.4**		**1.4**	**2.8**		**11.5**
Iran[a]	451	2.4	5.5	17.1	—	13.5	8.4	7.5	21.5	2.0	2.9	2.0	nr	nr	6.2	nr	nr	2.2	8.6
Arabia[a]	389	15.2	3.6	12.9	—	20.8	4.6	3.6	10.5	0.8	1.8	1.8	nr	nr	7.7	nr	nr	10.5	6.2
Syrian[b]	69	5.8	4.3	24.6	2.9	10.1	10.1	4.3	15.9	—	—	2.9	—	1.4		2.9	2.9		11.6
Palestinian[b]	117	2.6	1.7	30.8	—	9.4	12.8	6.8	7.6	—	3.4	2.6	—	1.7		0.9	4.3		15.4
Georgian[c]	139	0.7	7.2	17.3	0.7	3.6	12.9	10.1	21.6	2.2	10.1	1.4	—	2.9		—	—		9.3
Armenian[c]	192	0.5	7.3	30.9	—	8.9	11.5	7.9	22.5	1.6	2.1	1.0	0.5	—		—	—		5.3
Anatolia[a,c]	388	2.8	3.6	25.0	—	10.9	11.9	5.9	19.3	2.3	4.4	3.9	—	4.4	0.8	nr	nr	0.3	4.5
Italian[d]	99	—	2.0	33.3	5.1	7.1	9.1	8.1	22.2	4.0	3.0	2.0	—	—		—	—		4.1
Slav[e]	324	nr	nr	41.4	3.1	10.5	12.3	3.7	19.4	2.8	0.6	0.9	—	0.9		—	—		4.4
Finno-Ugrie[e]	149	nr	nr	45.6	2.0	12.1	6.0	3.3	22.8	1.4	2.0	2.7	—	0.7		—	—		1.4
German[f]	200	—	nr	50.0	2.5	7.5	8.5	6.5	13.5	2.5	0.5	1.0	—	—		—	—		7.5
C-Asian[g]	205	nr	nr	14.0	—	2.5	3.5	0.5	8.0	1.0	—	1.0	6.8	38.5		—	—		24.2
Indian[a]	1300	0.5	0.6	2.4	—	0.8	1.1	0.2	12.0	0.6	0.2	1.5	nr	nr	66.3	—	—		13.8

nr, Not reported.
[a] Kivisild et al. (2003), in press.
[b] Richards et al. (2000).
[c] Tambets et al. (2000).
[d] Torroni et al. (1997).
[e] Kivisild et al. (1999).
[f] Deduced from the sequence data of Lutz et al. (1998) by Kivisild et al. (1999).
[g] Deduced from the sequence data of Comas et al. (1998) by Kivisild et al. (1999).

Table 2
Combined high-resolution RFLP and control-region haplotypes observed in 52 Iraqi MtDNAs

Sample ID #	Haplogroup	Haplotype[a]		Read from np/to np
		RFLP[b]	Control-region[c]	
IQ 7	**H**	**−7025a**, −9380f, **+11718e**, **+14766u**	192 362	16003/16497
IQ 46	**H**	**−7025a**, **+11718e**, **−14766u**	256 352	16003/16497
IQ 77	**H**	**−7025a**, **+11718e**, **−14766u**, −16049k, +16517e	051 172	16003/16497
IQ 130	**H**	**−7025a**, **+11718e**, **−14766u**, +16517e	291	16003/16497
IQ 140	**H**	**−7025a**, −9380f, **+11718e**, **−14766u**, +16478c	362 482	16003/16497
IQ 148	**H**	**−7025a**, **+11718e**, **−14766u**, −16310k	266 311 362	16003/16497
IQ 165	**H**	+4745k, **−7025a**, −7335l, **+11718e**, −13325k, **−14766u**, +14877g, +16517e	167 274 488	16005/16497
IQ 176	**H**	**−7025a**, **+11718e**, **−14766u**, −16303k	304	16003/16497
IQ 195	**H**	**−7025a**, −7598f, **+11718e**, **−14766u**, +16296c	104 300 325 362	16003/16497
IQ 14	**HV**	+7090g, **+11718e**, **−14766u**	129 221	16006/16497
IQ 72	**HV**	**+11718e**, **−14766u**, −16310k	248 311	16003/16497
IQ 145	**HV**	**+11718e**, **−14766u**, −16310k	172 311	16003/16497
IQ 175	**HV**	**+11718e**, **−14766u**, −16310k, +16389m/+16390j/−16390b	311 391	16007/16497
IQ 202	**HV**	**+11718e**, **−14766u**, +16517e	189 316 362	16006/16429
IQ 203	**HV**	**+11718e**, **−14766u**, +16517e	189 316	16003/16429
IQ 13	**pre-HV**	**+11718e**, +16517e	**126** 185 355 **362**	16003/16497
IQ 60	**pre-HV**	**+11718e**	**126** 224 **362**	16003/16497
IQ 131	**pre-HV**	+8249b/−8250e, **+11718e**	**126** 145 189 232A **362**	16003/16430
IQ 64	**R***	+314i, −8592j, −12560a, −13634o, −16303k, −16310k, +16517e	266 294 297 304 311 355 356	16005/16497
IQ 33	**R2**	**+4216q**, **+4769a**, **−14304a**, +16517e	**071** 325 355 357	16005/16497
IQ 17	**J1**	**+4216q**, **+10394c**, −15883e, **−16065g**, +16517e	**069 126** 145 **222 261**	16003/16497
IQ 20	**J1**	+1004h, **+4216q**, **+10394c**, **−16065g**, +16296c	**069 126** 193 300 309	16005/16497
IQ 26	**J1**	**+4216q**, −8391e, **+10394c**, **−16065g**, +16517e	**069 126 193** 212 368	16005/16497
IQ 39	**J1**	**+4216q**, −6211g, **+10394c**, −10598a, **−16065g**	**069 126** 145 **222 261** 290	16004/16497
IQ 76	**J1**	**+4216q**, **+10394c**, **−16065g**, +16517e	**069 126** 145	16003/16497
IQ 86	**J1**	−3337k, **+4216q**, **+10394c**, **−16065g**, +16517e	**069 126** 145 **261**	16003/16497
IQ 106	**J1**	**+4216q**, −8150i, **+10394c**, −13916g, **−16065g**	**069** 075 **126** 290	16004/16497
IQ 118	**J1**	**+4216q**, **+10394c**, **−16065g**	**069 126** 145 **261** 355	16003/16497
IQ 132	**J1**	**+4216q**, **+10394c**, **−16065g**	**069 126** 148	16003/16497
IQ 206	**J1**	**+4216q**, **+10394c**, +14458a, **−16065g**	**069** 093 **126** 145 **222** 223 **261** 274	16003/16497
IQ 147	**J2**	**+4216q**, **−7474a**, **+10394c**, +11001n/+11002f, +11439j, −12170g/+12171j, **−15254p**, −16049k, **−16065g**, −16310k	051 **069 126** 188 311	16003/16497
IQ 27	**T**	−2849k, **+4216q**, **+13366m/−13367b/+13367j**, **+15606a**, **−15925i**, +16517e	**126** 213 **294** 296	16003/16497
IQ 84	**T**	**+4216q**, **+13366m/−13367b/+13367j**, **+15606a**, **−15925i**, +16517e	**126 294** 296	16003/16497
IQ 126	**T**	**+4216q**, **+13366m/−13367b/+13367j**, **+15606a**, **−15925i**, +16517e	**126** 286 **294** 362	16005/16497
IQ 121	**T**	**+4216q**, +7852j/−7853o, −12629b, **+13366m/−13367b/+13367j**, **+15606a**, **−15925i**	**126 163 186 189 294**	16004/16497
IQ 138	**T**	**+4216q**, **+13366m/−13367b/+13367j**, **+15606a**, **−15925i**, −16303k, +16517e	093 **126 294** 296 **304**	16003/16497
IQ 2	**K**	−3337k, +7118b, −8515c, **−9052n/−9053f**, **+10394c**, **+12308g**, **−16310k**, +16517e	093 **224 311**	16004/16497
IQ 67	**K**	−6260e, **−9052n/−9053f**, **+10394c**, **+12308g**, **−16310k**, +16517e	**224 311**	16005/16497
IQ 190	**K**	**−9052n/−9053f**, **+10394c**, **+12308g**, **−16310k**	093 **224** 278 **311**	16005/16497

Table 2 (*continued*)

Sample ID #	Haplogroup	Haplotype[a]		Read from np/to np
		RFLP[b]	Control-region[c]	
IQ 207	**K**	+7118b, **−9052n/−9053f**, **+10394c**, **+12308g**, **−16310k**, +16517e	189 **224 311**	16004/16429
IQ 125	**U1**	**−4990a**, +9288a, **+12308g**, **+14068l**	129 **189 249** 288 362	16003/16430
IQ 6	**U2**	−7859j, **+12308g**, **+15907k**, **+13730g**, **−16049k**, +16517e	**051 129C** 189 355 362	16003/16404
IQ 192	**U2**	−984c, **+12308g**, **+15907k**, **+13730g**, **−16049k**	**051 129C** 189 235 256 362	16006/16390
IQ 134	**U3**	+2293a, **+12308g**, +16389g/−16390b, +16517e	129 189 319 **343** 390	16005/16430
IQ 167	**U3**	**+12308g**	**343**	16003/16459
IQ 236	**U3**	+2293a, +8368f, **+12308g**	193 249 **343**	16004/16497
IQ 237	**U4**	**+4643k**, +7702k, −8587k, **+11329a**, **+12308g**, +16517e	**356**	16003/16497
IQ 184	**U6**	**+3348j**, **+12308g**,**+16217l**, +16517e	134 **172 219** 278	16003/16497
IQ 89	**U7**	+8165e, **+12308g**	167 256 **318T** 368	16003/16497
IQ 129	**U7**	−8572e, **+12308g**, +16517e	126 148 213 309 **318C**	16003/16497
IQ 178	**U7**	**+12308g**, −16065g, +16223c/+16226a, +16517e	069 227 **318T** 359	16003/16497
IQ 12	**M**	−7886e, **+10394c**, **+10397a**, +16517e	037 **223** 278 465	16003/16497

[a] Motifs diagnostic for each haplogroup are shown in bold.

[b] RFLP sites are numbered from the first nucleotide of the recognition sequence. A "+" indicates the presence of a restriction site, a "−" the absence. Sites separated by a slash indicate simultaneous changes for different enzymes due to a single nucleotide substitution. The restriction enzymes used in the analysis are designated by the following single-letter codes: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; f, *Hha*I; g, *Hin*fI; h, *Hpa*I; i, *Msp*I; j, *Mbo*I; k, *Rsa*I; l, *Taq*I; m, *Bam*HI; n, *Hae*II; o, *Hin*cII; p, *Acc*I; q, *Nla*III; and u, *Mse*I.

[c] From Richards et al. (2000). Only those nucleotide positions (−16,000) that differ from the Cambridge reference sequence (CRS) (Anderson et al., 1981; Andrews et al., 1999) are shown. Mutations are transitions unless the base change is explicitly specified.

unpublished data), show an opposite gradient of frequency having R-M269 its maximum value in Western Europe and R-M17 in Central Asia/Eastern Europe (Passarino et al., 2001; Rosser et al., 2000; Semino et al., 2000a). Haplogroup R is sporadic in Africa (Underhill et al., 2000) with the exception of some populations from Northern Cameroon where it is present at a high frequency (about 40%). However, none of the Cameroonian M173 chromosomes belongs to the sub-haplogroup R-M269 or R-M17. Differently from the African sporadic findings that, harboring the M269 marker were interpreted as recent arrivals from Europe, the Cameroonian Y chromosomes were explained as a probable record of a back migration from Asia to Africa (Cruciani et al., 2002). In Iraq, haplogroup R has a frequency of 17%. None of these Y chromosomes belongs to the "African" R branch being for 62.5% R-M269 and 37.5% R-M17.

3.3. 49a,f/TaqI polymorphism

The 49a,f polymorphism (Ngo et al., 1986) is a complex system due to at least five polymorphic bands of the eighteen fragments detected by probes 49a and 49f on Southern blots of male DNAs digested with the restriction enzyme *Taq*I. In spite of the partial understanding of its molecular basis (Saxena et al., 1996), this system has provided some very informative anthropogenetic markers (Excoffier et al., 1996; Lucotte and Hazout, 1996; Poloni et al., 1997; Santachiara-Benerecetti et al., 1993; Spurdle and Jenkins, 1992; Torroni et al., 1990). Moreover, the high variability of some of its bands allows the subdivision of haplogroups into sub-haplogroups.

A total of 21 different 49a,f haplotypes were found and are illustrated in Fig. 4 as a sub-classification of the Iraqi Y-chromosome haplogroups. The most represented haplotype of haplogroup E is haplotype 5 ($A_2C_0D_0F_1I_1$). This is followed by haplotype 11 ($A_3C_0D_0F_1I_1$) at a much lower frequency. Haplotypes 5 and 11 were observed both in Africa (Lucotte et al., 2001; Passarino et al., 1998; Persichetti et al., 1992; Santachiara-Benerecetti and Semino, 1996; Spurdle and Jenkins, 1992; Torroni et al., 1990) and Eurasia (Passarino et al., 2001; Semino et al., 2000b) but in Africa they belong to haplogroup E, whereas in Eurasia, particularly in Northeastern Eurasia, they belong mainly to the haplogroup R-M17. Interestingly, the proportion of haplotypes 5 and 11 in haplogroups E and R-M17 is reversed, with haplotype 5 prevalent in haplogroup E and haplotype 11 in haplogroup R-M17. By considering that the two haplotypes differ by a single band change and their different proportion in the two lineages, it is likely that haplotype 11 is a derivative of haplotype 5 in haplogroup E and just the opposite in haplogroup R-M17. It is worth mentioning that in haplogroup E two

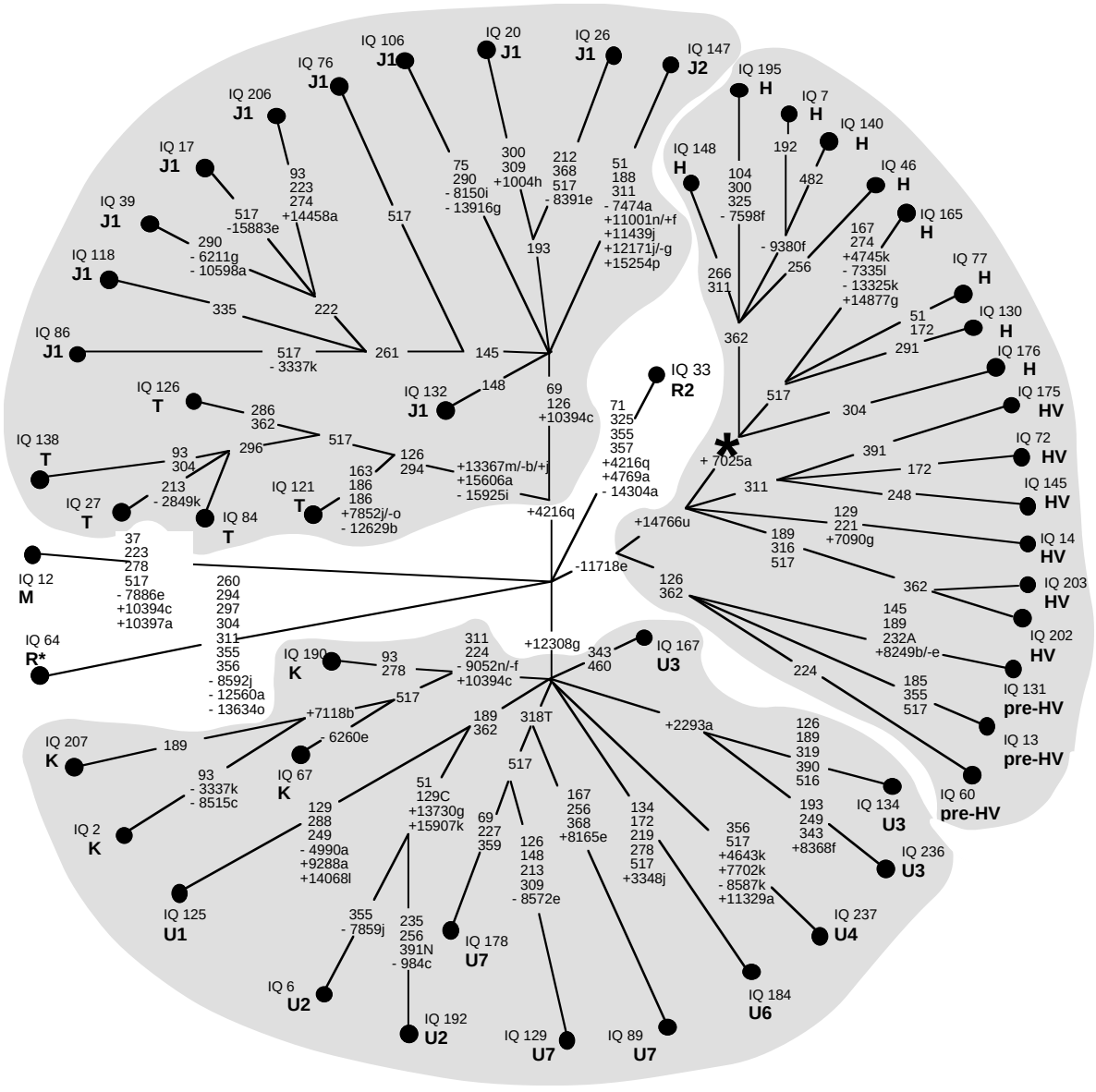
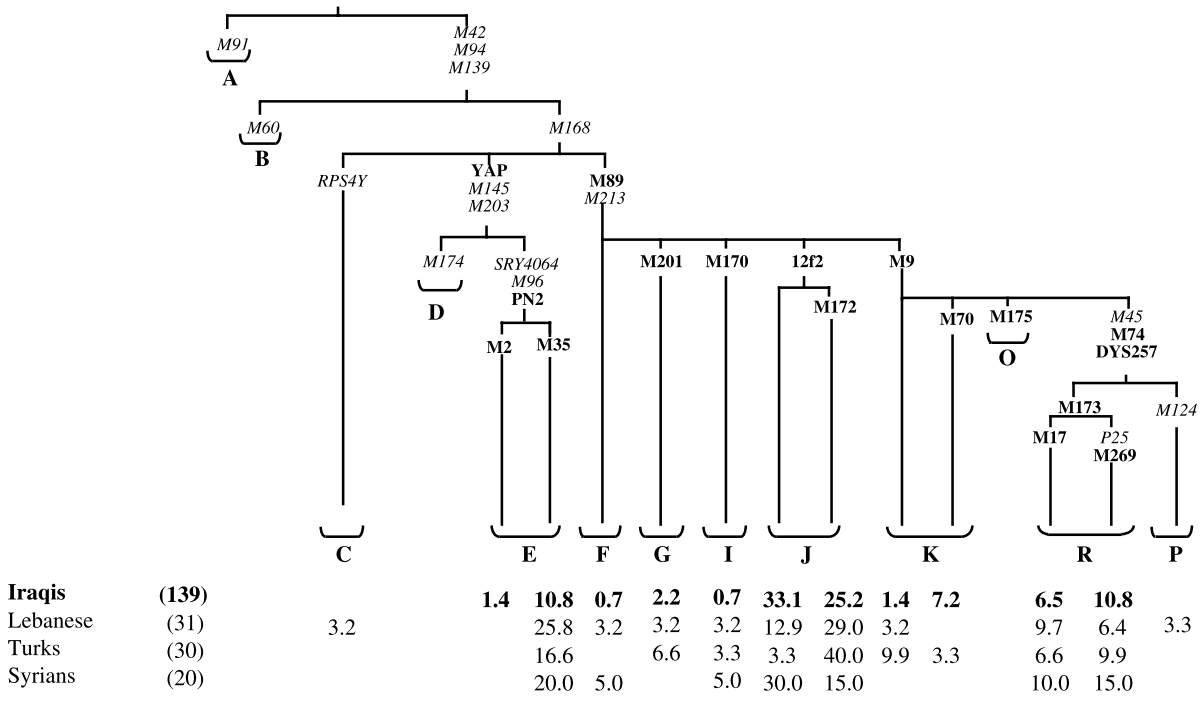


Fig. 2. Phylogeny of the mtDNA haplotypes observed in 52 Iraqi subjects. This phylogeny has been constructed by hand, taking into account the information provided by the literature, including the complete sequence data of Ingman et al. (2000), Finnila et al. (2001), and Maca-Meyer et al. (2001), and by using both high-resolution RFLP data and sequence variation of the control-region. Differences between haplotypes are represented by the mutations listed on the branches connecting them. However, to improve visualization, the lengths of the branches were drawn without an exact correspondence to the number of listed mutations. Circles indicate distinct mtDNA haplotypes. The node marked by an asterisk (*) corresponds to the CRS (Anderson et al., 1981; Andrews et al., 1999). The direction of indicated RFLP site losses (−) and gains (+) is from this node toward the periphery. The restriction enzymes are indicated by a letter: a, *Alu*I; b, *Ava*II; c, *Dde*I; e, *Hae*III; f, *Hha*I; g, *Hin*fI; h, *Hpa*I; i, *Msp*I; j, *Mbo*I; k, *Rsa*I; l, *Taq*I; m, *Bam*HI; n, *Hae*II; o, *Hin*cII; p, *Acc*I; q, *Nla*III; and u, *Mse*I. Site positions followed by a slash indicate simultaneous changes for different enzymes due to a single nucleotide substitution. Control-region mutations (from np 16,007 to 16,390) are given −16,000. These are transitions, unless specified by the base change. Capital letters (in bold) indicate the haplogroup affiliation.

subjects belong to the African specific E-M2 clade, which is very frequent in the Western and Southern part of the continent (Cruciani et al., 2002; Passarino et al., 1998; Scozzari et al., 1999; Seielstad et al., 1994; Semino et al., 2002; Underhill et al., 2000) and has been related to the Bantu expansion. These two Y chromosomes harbor haplotype 4 ($A_1C_0D_0F_1I_1$) which is also African-specific and shows the same geographic distribution (Excoffier et al., 1987; Passarino et al., 1998; Spurdle and Jenkins, 1992; Torroni et al., 1990). Within haplogroup J, haplotypes 7 ($A_2C_0D_1F_1I_0$) and 8 ($A_2C_0D_1F_1I_1$) are the most represented, but haplotype 7 is observed only in the J-M172 sub-set. This suggests that the 49a,f haplotype 7 arose on a 12f2-8Kb/M172 Y chromosome.

Out of the 12 Y chromosomes belonging to haplogroup K-M70, 10 are 49a,f haplotype 13 ($A_3C_0D_1F_1I_1$). As for haplogroup R, the R-M17 clade in Iraq is



Iraqis	**(139)**		**1.4**	**10.8**	**0.7**	**2.2**	**0.7**	**33.1**	**25.2**	**1.4**	**7.2**	**6.5**	**10.8**	
Lebanese	(31)	3.2		25.8	3.2	3.2	3.2	12.9	29.0	3.2		9.7	6.4	3.3
Turks	(30)			16.6		6.6	3.3	3.3	40.0	9.9	3.3	6.6	9.9	
Syrians	(20)			20.0	5.0		5.0	30.0	15.0			10.0	15.0	

Fig. 3. Phylogenetic tree of the Y-chromosome haplotypes and their percent frequencies in Iraq compared with the frequencies in the Lebanese, Syrians, and Turks previously reported by Semino et al. (2000a). Numbering of mutations is according to The YCC (2002): those examined in the present study are shown in boldface type; those inferred are shown in italics. Capital letters indicate haplotypes according to The YCC (2002). Haplogroup J(xM172) includes the samples previously classified as Eu10 (Semino et al., 2000a) with the exception of one Lebanese which was 12f2-10Kb and therefore belonging to haplogroup F-M89.

preferentially associated with haplotype 11 as in other Western Eurasian populations (Passarino et al., 2001) while the R-M269 clade appears almost completely associated with haplotype 35 ($A_3C_1D_0F_1I_1$). Only one R-M269 subject displayed haplotype15 ($A_3C_1D_2F_1I_1$) which is by far the most frequent haplotype of this clade in Europe, where haplotype 35 is much less represented (Santachiara-Benerecetti, unpublished data). Haplotype 39 ($A_3C_0D_0F_1I_0$-BHPR), observed in one subject, is characterized by the absence of four nonpolymorphic bands, most likely due to a single deletion event (Passarino et al., 2001).

4. Discussion

The study of the Iraqi population through the analysis of maternal and paternal markers helps to fill a gap in the picture of the Middle Eastern genetic population structure. As illustrated by the PC analysis of the mtDNA data (Fig. 5), Iraq clusters with the other Middle Eastern populations, and together with Iran, occupies a central position among Arabia, Caucasus, and Europe. The first component separates Arabia from the others, mainly from Europe and Caucasus, being the Arabia characterized by a substantial presence of African haplogroups and a high frequency of haplogroup pre-HV, whereas Europe and Caucasus lack the African components and harbor a high frequency of haplogroup H. The second component, which maintains the Iraqi and Iranian populations close to each other and well separates them from the other Middle Eastern populations, is mainly due to the distribution of haplogroup HV and, in part to that of haplogroup U.

The main determinants of this PC analysis are therefore the HV supercluster members, haplogroups H, HV, and pre-HV. Actually, whereas in the European populations haplogroup H reaches its highest frequencies, HV and pre-HV mtDNAs (when present) have a very low incidence. In this regard, it is interesting to note the high frequency of haplogroups pre-HV and HV in the Arabians (18.8%) (Kivisild et al., 2003, in press) and in the Arabian Bedouins (20.6%) (Di Rienzo and Wilson, 1991) in which the H haplogroup shows one of its lowest values. These data are in agreement with the hypothesis that haplogroup H originated in the Middle East where, although less frequent than in Europe, it displays the greatest heterogeneity (Torroni et al., 1998). Haplogroup U, which has in Iraq a frequency comparable to haplogroup H, does not play an important role in the PC analysis. This is probably due to the heterogeneous distribution of its numerous sub-clades not available for most of the populations. When the different sub-clades are separately considered (Richards et al.,

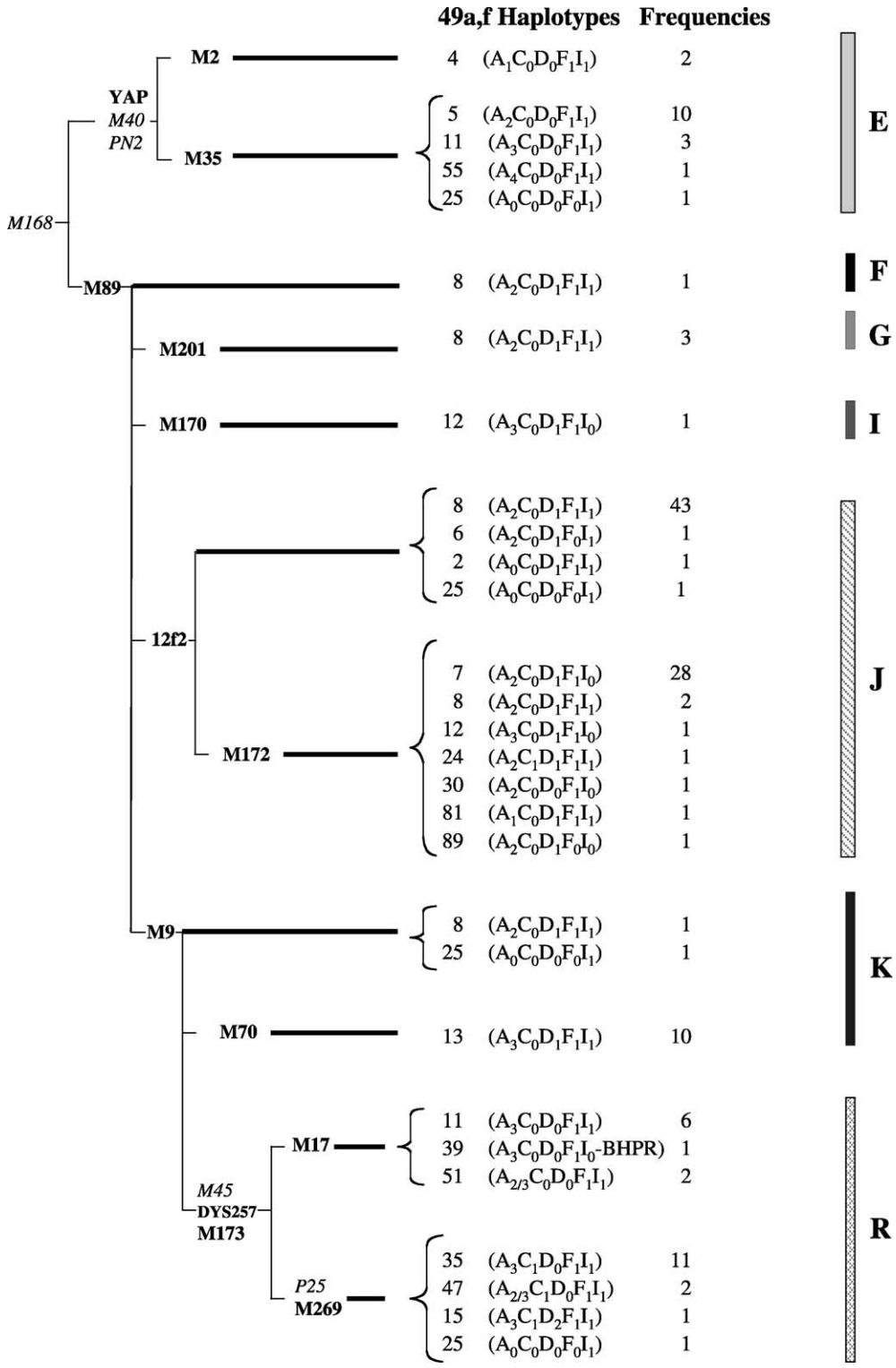


Fig. 4. The distribution of the 49a,f haplotypes in the Y-chromosome haplogroups found in Iraq. The simultaneous absence of four polymorphic bands (BHPR) in haplotype 49a,f-39 is probably due to a single deletion event (Passarino et al., 2001).

2002), U5 correlates with H and pre-V, concentrated in the European pool, while U1 with haplogroup pre-HV concentrated in the Near Eastern pool. As illustrated in Fig. 2, haplogroup U exhibits in Iraq all its most common sub-groups with the exception of the rare U8 (Finnila et al., 2001). Indeed, haplogroup U5, which does not appear in the network (constructed with 52 mtDNAs), was observed in one subject included in the larger Iraqi sample ($N = 116$) reported by Richards et al. (2000). According to these authors, the Iraqi U5

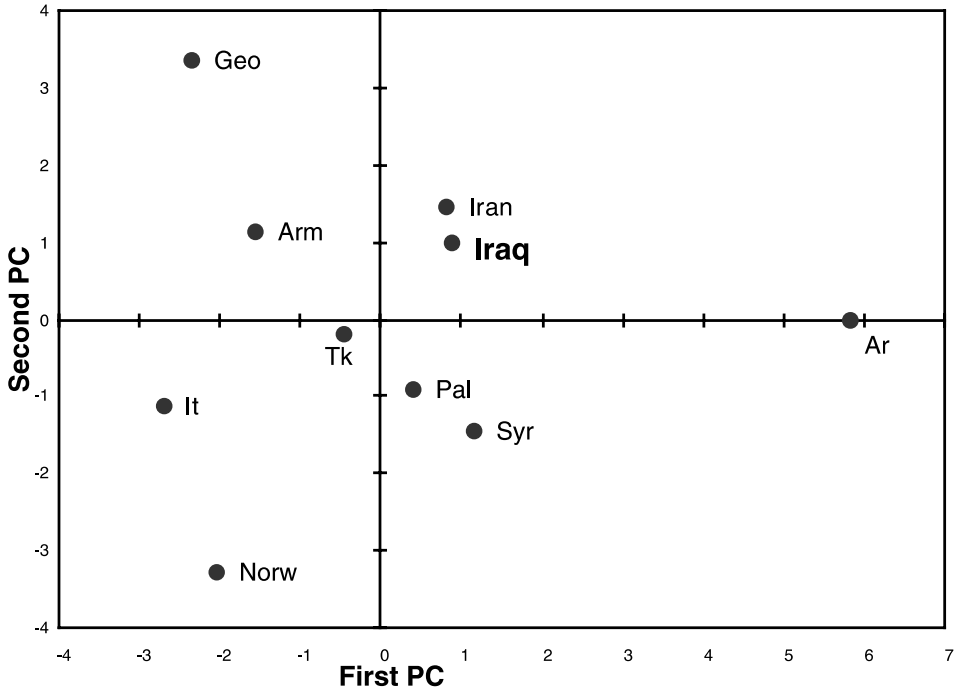


Fig. 5. PC analysis of the mtDNA haplogroups. The populations represented are: Iraqi (present study), Italian [It] (Torroni et al., 1997), Armenian [Arm], Georgian [Geo] (Tambets et al., 2000), Palestinian [Pal], Syrian [Syr] (Richards et al., 2000), Iranian, Arab [Ar], Anatolian [Tk] (Kivisild et al., 2002, in press), and Norwegian [Norw] (Passarino et al., 2002). On the whole, 67% of the total variance is represented: 43% by the first component and 24% by the second component.

mtDNA belongs to the highly derived European subcluster U5a1a, and is due to recent gene flow from Europe. The finding of the U6, which is typical of Berbers and North Africans (Macaulay et al., 1999), may be considered the result of contact with these populations. On the contrary, the presence (2.6% of 116 sub-classified mtDNAs) of haplogroup U7, which is very rare in Europe, can be regarded as indicative of a genetic connection between Southwestern Asia (Iran, Arabia, Anatolia, the two Caucasus regions, and India) and Eastern Africa (Nile Valley) where it is also found (Tambets et al., 2000).

As for the Y chromosome, the distinguishing feature of the Iraqis is the extremely high frequency (58.3%) of the haplogroup J-12f2. This, together with the values observed in Syria and in the Zagros Mountains of Iran, is the highest incidence reported so far (Hammer et al., 2000; Mitchell et al., 1993; Nebel et al., 2001; Quintana-Murci et al., 2001; Ritte et al., 1993; Santachiara-Benerecetti et al., 1993; Semino et al., 1996, 2000a; Wells et al., 2001). It has been proposed that the 12f2-8Kb mutation arose in the Fertile Crescent, and the decreasing frequency gradient from Southeastern to Northwestern Europe has been attributed to the demic diffusion of agriculture (Semino et al., 1996, 2000a). In addition, this haplogroup might have marked other migrations such as the Indo-European, the Phoenician, and the Greek as well as ancient contacts with Ethiopia (Malaspina et al., 2001; Mitchell and Hammer, 1996; Passarino et al., 1998; Santachiara-Benerecetti and Semino, 1996; Scozzari et al., 2001). As revealed by the PC analysis, haplogroup J-12f2 plays an important role both in the first and second principal components. Actually, the clade J without the mutation M172 [J(xM172)], especially frequent in the Middle East and rare in Europe, and haplogroups R-M269 and I-M170, very frequent in Europe, are the main determinants of the first principal component (Fig. 6) which clearly separates the European populations from the others. On the other hand, haplogroups J-M172 and G-M201, which show their maximum frequency in the Caucasus and Anatolia, together with haplogroups E-M35 and J(xM172), frequent in the African populations and Arabs, are the main determinants of the second principal component. This component, which does not well discriminate between the European populations, clearly distinguishes those of the Middle East. This distinction is mainly due to the ratio J-M172/J(xM172), which is 0.76 in Iraq, 0.5 in Syria, and close to 1 in some Middle Eastern populations (Malaspina et al., 2001; Nebel et al., 2001; Semino et al., 2000a), strongly increases in Europe (Malaspina et al., 2001; Semino et al., 2000a) but decreases toward the Southern Middle East (Nebel et al., 2001), North Africa, and Ethiopia (Bosch et al., 2001; Cruciani et al., 2002; Semino et al., 2002). Thus, J-M172 can be considered a marker of the Neolithic expansion in Europe, whereas the J(xM172) might have marked migrations toward the Arabian Peninsula, North Africa, and East Africa. When

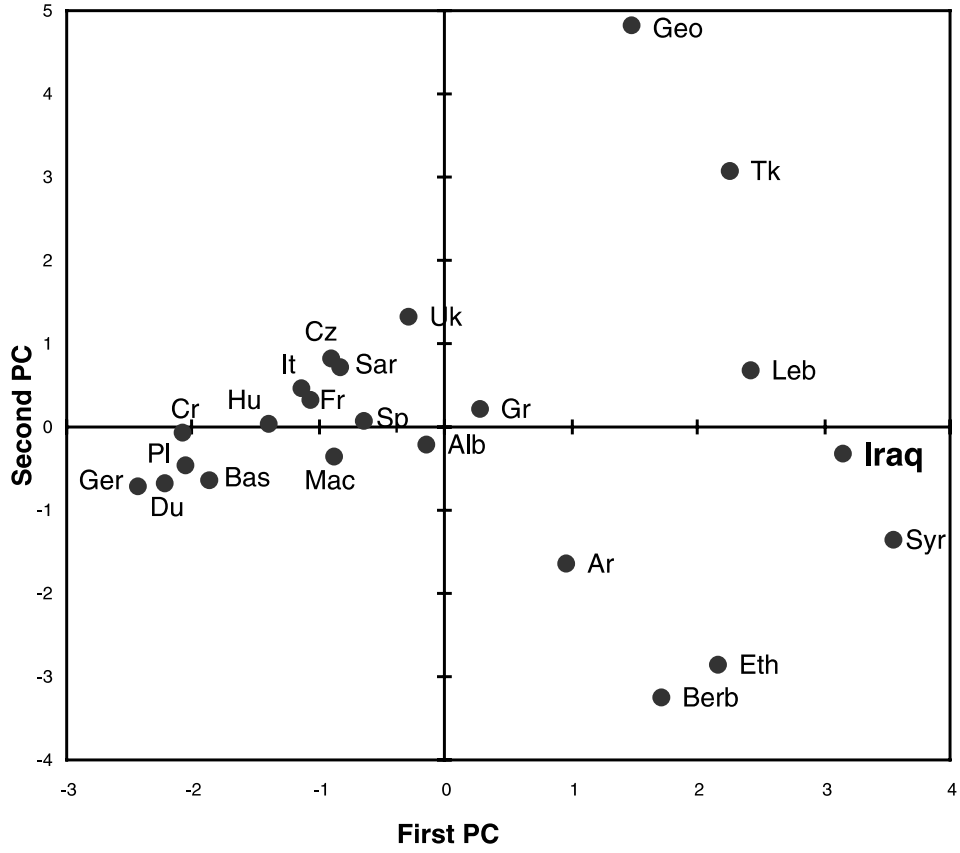


Fig. 6. PC analysis of the Y-chromosome haplogroups. The populations represented are: Iraqi (present study), Basque [Bas], Spanish [Sp], French [Fr], German [Ger], Italian [It], Sardinian [Sar], Croat [Cr], Albanian [Alb], Greek [Gr], Macedonian [Mac], Polish [Pl], Hungarian [Hu], Ukrainian [Uk], Georgian [Geo], Anatolian [Tk], Lebanese [Leb], Syrian [Syr] (Semino et al., 2000a,b), Ethiopian [Eth] (Semino et al., 2002), Arab [Ar], and Berber [Ber] (Cruciani et al., 2002). The African haplogroups A-M13, B-M60, and D-P2 which, with the exception of single observations in Sardinia and in the Berbers, were observed only in Ethiopian groups, were pooled. On the whole, 37% of the total variance is represented: 20% by the first component and 17% by the second component.

J-M172 and J(xM172) are compared, a higher variation both for the 49a,f system and for eight microsatellites (Santachiara-Benerecetti et al., unpublished data) is observed in J-M172 ($h = 0.947$ vs. 0.844), and the internal microsatellite variance of J-M172 is more than twice that of J(xM172). Thus, whatever the real age of these two lineages may be (age evaluations for the Y chromosome are always very uncertain, Brinkmann et al., 1998; Forster et al., 2000), J-M172 appears to be older than J(xM172). This consideration, which is in agreement with the data of Nebel et al. (2001) relative to other Middle Eastern populations, is apparently in contrast with the Y-chromosome phylogeny (Fig. 3) that derives J(xM172) directly from the node 12f2-8Kb, whereas J-M172 is derived through a further mutation (M172). These data suggest that all (or most) of the J(xM172) Y chromosomes might represent a monophyletic cluster defined by yet unidentified mutations. The lower variability of J(xM172) with respect to J-M172, together with the different distribution of these two lineages in the Middle East and in surrounding areas affected by Middle Eastern migrations, may suggest that they evolved in two separate human groups with a previous common ancestry and underwent expansions at different times. In particular, it seems reasonable to assume that the haplogroup J-M172 expanded first, possibly from the Northwestern part of the Fertile Crescent (Asia Minor), where agriculture actually started and triggered the development and growth of other Middle Eastern populations. On the other hand, J(xM172) seems to have had its center in the Eastern part of the Fertile Crescent (Iraq and Iran), and to have undergone a later expansion mainly affecting modern Arab populations (Cruciani et al., 2002; Nebel et al., 2001; Semino et al., 2002).

The Eurasian haplogroups, R-M269 and R-M17, represent the second most frequent component of the Iraqi Y-chromosome gene pool. The R-M17 cluster distribution and its 49a,f associations suggest that the population movements from Central Asia/Eastern Europe into the modern Iran postulated by Quintana-Murci et al. (2001) also influenced Iraq. Indeed, the R-M17 shows a frequency of 6.5% in Iraq which falls in the lower edge (3–9%) of the range observed in the Iranian populations living at the border of Iraq. As for

R-M269, its 49a,f associations do not allow a simple connection with the European representatives of this lineage. As shown in Fig. 4, haplogroup R-M269 in Iraq is mainly associated with 49a,f haplotype 35. This haplotype has a rather uniform distribution all over Europe (Santachiara-Benerecetti et al., 1993; Santachiara-Benerecetti, unpublished data; Semino et al., 2000a; Torroni et al., 1990) whereas the West–East decreasing gradient of frequency of R-M269 in Western Eurasia reflects the distribution of haplotype 49a,f 15 which was found in only one subject in the Iraqi sample. In this scenario it is possible that the 49a,f haplotype 35 characterized the lineage M173/M269 present in Western Eurasia in early times. Afterward, haplotype 15 could have derived from haplotype 35 from which it differs by only one mutation, and its extremely high incidence in Western Europe could be the result of drift, consistent with an inferred population bottleneck during the Last Glacial Maximum.

The data obtained from these analyses illustrate the complexity of the Iraqi genetic structure. For both uniparental systems, the large majority of the haplogroups observed in the Iraqi population are those (H, J, T, and U for the mtDNA, J-M172 and J(xM172) for the Y chromosome) considered to have originated in the Middle East and to have later spread mainly in Western Eurasia.

Although the level of resolution of our analyses does not allow an evaluation of the back migrations of widely distributed haplogroups, it is possible to identify some European, African, and Asian signature markers in both male and female gene pools of modern Iraqis. Overall, these external long-range input appear to represent about 9% of mtDNAs and 30% of Y-chromosomes. MtDNA gene flow is revealed by the presence of the European haplogroups V and U5 (1.4%), the North African haplogroup U6 (0.9%), the sub-Saharan African haplogroups L1 and L2 (4.2%), and the East-Asian haplogroups M and B (2.3%). As for the Y chromosome, the African input is represented by the haplogroup E (12.2%), the input from Eastern Europe/Central Asia is revealed by R-M269/49a,f haplotypes 35 with its derivative 47 and R-M17 (15.9%), while haplogroup I and R-M269/49a,f haplotype 15 Y chromosomes indicate a limited gene flow from Europe (1.4%). The different extent of male and female external contribution observed in the Iraqi gene pool is a frequent observation in this kind of studies and often finds an explanation in the different male and female social behavior (Excoffier et al., 1996; Huoponen et al., 1997; Passarino et al., 1998; Perez-Lezaun et al., 1999; Salem et al., 1996; Scozzari et al., 1997; Seielstad et al., 1998; Torroni et al., 1994). However, it has also to be taken into account that the level of resolution of the mtDNA markers used in this study could be lower than that reached by the Y-chromosome counterparts.

In conclusion, this study provides missing information on an interesting Middle Eastern region, confirms the geographical importance of the Middle East as a crossroad between the three continents, supports the Eastern side of the Fertile Crescent as the homeland of the Y-chromosome haplogroup J, and suggests that long-range gene flow to this area might have involved mainly males.

Acknowledgments

This research was supported by the Italian Ministry of the University, Progetti Ricerca Interesse Nazionale 1999 and 2002 (to A.S.S.-B. and A.T.), Progetto Finalizzato C.N.R. "Beni Culturali" (to A.S.S.-B.), as well as from Fondo d'Ateneo per la Ricerca dell'Università di Pavia (to A.S.S.-B. and A.T.). N.A.-Z. was supported by ICGBE (Trieste) and University of Pavia fellowships.

References

Akey, J.M., Sosnoski, D., Parra, E., Dios, S., Hiester, K., Su, B., Bonilla, C., Jin, L., Shriver, M.D., 2001. Melting curve analysis of SNPs (McSNP): a gel-free and inexpensive approach for SNP. Genotyping Biotech. 30, 358–367.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., Young, I.G., 1981. Sequence and organization of the human mitochondrial genome. Nature 290, 457–465.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. 23, 147.

Bandelt, H.-J., Forster, P., Sykes, B.C., Richards, M.B., 1995. Mitochondrial portraits of human populations using median networks. Genetics 141, 743–753.

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 Iberian Peninsula. Am. Hum. Genet. 68, 1019–1029.

Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., Rolf, B., 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am. J. Hum. Genet. 62, 1408–1415.

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

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A., 1994. The History and Geography of Human Genes. Princeton University Press, Princeton, NJ.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D., Bertranpetit, J., 1998. Trading genes along the silk road: mtDNA sequences and the origin of central Asian populations. Am. J. Hum. Genet. 63, 1824–1838.

Cruciani, F., Santolamazza, P., Shen, P., Macaulay, V., Moral, P., Olckers, A., Modiano, D., Holmes, S., Destro-Bisol, G., Coia, V.,

Wallace, D.C., Oefner, P.J., Torroni, A., Cavalli-Sforza, L.L., Scozzari, R., Underhill, P.A., 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. Am. J. Hum. Genet. 70, 1197–1214.

Di Rienzo, A., Wilson, A.C., 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc. Natl. Acad. Sci. USA 88, 1597–1601.

Excoffier, L., Pellegrini, B., Sanchez-Mazas, A., Simon, C., Langaney, A., 1987. Genetics and history of sub-Saharan Africa. Year Book Phys. Anthropol. 30, 151–194.

Excoffier, L., Poloni, E.S., Santachiara-Benerecetti, A.S., Semino, O., Langaney, A., 1996. The molecular diversity of the Niokholo Mandenkalu from Eastern Senegal: an insight into West Africa genetic history. In: Boyce, A.J., Mascie-Taylor, C.G.N. (Eds.), Molecular Biology and Human Diversity. Cambridge University Press, Cambridge.

Finnila, S., Lehtonen, M.S., Majamaa, K., 2001. Phylogenetic network for European mtDNA. Am. J. Hum. Genet. 68, 1475–1484.

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.

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

Hammer, M.F., 1994. A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol. Biol. Evol. 11, 749–761.

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

Hammer, M.F., Redd, A.J., Wood, E.T., Bonner, M.R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M.A., Jenkins, T., Ostrer, H., Bonne-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.

Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., Zegura, S.L., 1997. The geographic distribution of human Y chromosome variation. Genetics 145, 787–805.

Huoponen, K., Torroni, A., Wickman, P.R., Sellitto, D., Gurley, D.S., Scozzari, R., Wallace, D.C., 1997. Mitochondrial DNA and Y chromosome-specific polymorphisms in the Seminole Tribe of Florida. Eur. J. Hum. Genet. 5, 25–34.

Ingman, M., Kaessmann, H., Paabo, S., Gyllensten, U., 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408, 708–713.

Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R.S., Redd, A.J., Zegura, S.L., Hammer, M.F., 2001. Paternal population history of East Asia: sources, patterns, and microevolutionary processes. Am. J. Hum. Genet. 69, 615–628.

Kivisild, T., Kaldma, K., Metspalu, M., Parik, J., Papiha, S., Villems, R., 1999. The place of the Indian mitochondrial DNA variations in the global network of maternal lineages and the peopling of the old world. In: Papiha, S., Deka, R., Chakraborty, R. (Eds.), Genomic Diversity: Applications in Human Population Genetics. Kluwer Academic/Plenum publishers, New York.

Kivisild, T., Rootsi, S., Metspalu, M., Metspalu, E., Parik, J., Katrin, K., Usanga, E., Mastana, S., Papiha, S.S., Villems, R., 2003. Genetics of the language and farming spread in India. In: Renfrew, C., Boyle, K. (Eds.), McDonald Institute Monographs series, Cambridge, UK. Cambridge, in press.

Lucotte, G., Gerard, N., Mercier, G., 2001. North African genes in Iberia studied by Y-chromosome DNA haplotype 5. Hum. Biol. 73, 763–769.

Lucotte, G., Hazout, S., 1996. Y-chromosome DNA haplotypes in Basques. J. Mol. Evol. 42, 472–475.

Lutz, S., Weisser, H.J., Heizmann, J., Pollak, S., 1998. Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. Int. J. Legal. Med. 111, 67–77.

Maca-Meyer, N., Gonzalez, A.M., Larruga, J.M., Flores, C., Cabrera, V.M., 2001. Major genomic mitochondrial lineages delineate early human expansions. BMC Genet. 2, 13.

Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B., Torroni, A., 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. Am. J. Hum. Genet. 64, 232–249.

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

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

Mitchell, R.J., Earl, L., Williams, J.W., 1993. Two Y-chromosome specific restriction fragment length polymorphisms (DYS11 and DYZ8) in Italian and Greek migrants to Australia. Hum. Biol. 65, 387–399.

Mitchell, R.J., Hammer, M.F., 1996. Human evolution and the Y chromosome. Curr. Opin. Genet. Dev. 6, 737–742.

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.

Ngo, K.Y., Vergnaud, G., Johnsson, C., Lucotte, G., Weissenbach, J., 1986. A DNA probe detecting multiple haplotypes of the human Y-chromosome. Am. J. Hum. Genet. 38, 407–418.

Passarino, G., Semino, O., Magri, C., Al-Zahery, N., Benuzzi, G., Quintana-Murci, L., Andellnovic, S., Bullc-Jakus, F., Liu, A., Arslan, A., Santachiara-Benerecetti, A.S., 2001. The 49a,f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. Hum. Immunol. 62, 922–932.

Passarino, G., Semino, O., Quintana-Murci, L., Excoffier, L., Hammer, M., Santachiara-Benerecetti, A.S., 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-Chromosome polymorphisms. Am. J. Hum. Genet. 62, 420–434.

Passarino, G., Cavalleri, G.L., Lin, A.A., Cavalli-Sforza, L.L., Børresen-Dale, A.-L., Underhill, P.A., 2002. Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. Eur. J. Hum. Genet. 10, 521–529.

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

Persichetti, F., Blasi, P., Hammer, M., Malaspina, P., Jodice, C., Terrenato, L., Novelletto, A., 1992. Disequilibrium of multiple DNA markers on the human Y chromosome. Ann. Hum. Genet. 56, 303–310.

Poloni, E.S., Semino, O., Passarino, G., Santachiara-Benerecetti, A.S., Dupanloup, I., Langaney, A., Excoffier, L., 1997. Human genetic affinities for Y-chromosome P49a.,f/TaqI haplotypes show strong

correspondence with linguistics. Am. J. Hum. Genet. 61, 1015–1035.

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

Quintana-Murci, L., Semino, O., Poloni, E.S., Liu, A., VanGijn, G., Passarino, G., Brega, A., Nasidze, I.S., Maccioni, L., Cossu, G., Al-Zahery, N., Kidd, J.R., Kidd, K.K., Santachiara-Benerecetti, A.S., 1999b. Y-chromosome specific YCAII, DYS19 and YAP polymorphisms in human populations: a comparative study. Ann. Hum. Genet. 63, 153–166.

Rashed, I., 1997. The bone skeletons in the cave of Shanidar-Iraq. Iraqi J. Sci. 90, 31–32.

Richards, M., Macaulay, V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild, T., Villems, R., Thomas, M., Rychkov, S., Rychkov, O., Rychkov, Y., Golge, M., Dimitrov, D., Hill, E., Bradley, D., Romano, V., Cali, F., Vona, G., Demaine, A., Papiha, S., Triantaphyllidis, C., Stefanescu, G., 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. Am. J. Hum. Genet. 67, 1251–1276.

Richards, M., Macaulay, V.A., Bandelt, H.J., Sykes, B.C., 1998. Phylogeography of mitochondrial DNA in western Europe. Ann. Hum. Genet. 62, 241–260.

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.

Ritte, U., Neufeld, E., Broit, M., Shavit, D., Motro, U., 1993. The differences among Jewish communities—maternal and paternal contributions. J. Mol. Evol. 37, 435–440.

Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D.G., Brede, G., Cooper, G., Corte-Real, H.B., de Knijff, P., Decorte, R., Dubrova, Y.E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E.W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S.A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L.A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T.A., Mikelsaar, A.V., Mitchell, R.J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P.C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M.J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D.C., Saillard, J., Santos, F.R., Stefanescu, G., Sykes, B.C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M.A., 2000. Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. 67, 1526–1543.

Saillard, J., Magalhaes, P.J., Schwartz, M., Rosenberg, T., Norby, S., 2000. Mitochondrial DNA variant 11719G is a marker for the mtDNA haplogroup cluster HV. Hum. Biol. 72, 1065–1068.

Salem, A.H., Badr, F.M., Gaballah, M.F., Paabo, S., 1996. The genetic of traditional living: Y-chromosomal and mitochondrial lineages in the Sinai peninsula. Am. J. Hum. Genet. 59, 741–743.

Santachiara-Benerecetti, A.S., Semino, O., Passarino, G., Torroni, A., Brdicka, R., Fellous, M., Modiano, G., 1993. The common Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. Ann. Hum. Genet. 57, 55–64.

Santachiara-Benerecetti, A.S., Semino, O., 1996. Y-chromosome polymorphisms and history of populations. Cell. Pharmacol. 3, 199–204.

Saxena, R., Brown, L.G., Hawkins, T., Alagappan, R.K., Skaletsky, H., Reeve, M.P., Reijo, R., Rozen, S., Dinulos, M.B., Disteche, C.M., Page, D.C., 1996. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed repeatedly amplified and pruned. Nat. Genet. 14, 292–299.

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

Scozzari, R., Cruciani, F., Santolamazza, P., Sellitto, D., Cole, D.E., Rubin, L.A., Labuda, D., Marini, E., Succa, V., Vona, G., Torroni, A., 1997. mtDNA and Y chromosome-specific polymorphisms in modern Ojibwa: implications about the origin of their gene pool. Am. J. Hum. Genet. 60, 241–244.

Scozzari, R., Cruciani, F., Santolamazza, P., Malaspina, P., Torroni, A., Sellitto, D., Arredi, B., Destro-Bisol, G., De Stefano, G., Rickards, O., Martinez-Labarga, C., Modiano, D., Biondi, G., Moral, P., Olckers, A., Wallace, D.C., 1999. Combined use of biallelic and microsatellite Y-chromosome polymorphisms to infer affinities among African populations. Am. J. Hum. Genet. 65, 829–846.

Seielstad, M.T., Hebert, J.M., Lin, A.A., Underhill, P.A., Ibrahim, M., Vollrath, D., Cavalli-Sforza, L.L., 1994. Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum. Mol. Genet. 3, 2159–2161.

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., Brega, A., Fellous, M., Santachiara-Benerecetti, A.S., 1996. A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am. J. Hum. Genet. 59, 964–968.

Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhill, P.A., 2000a. The genetic legacy of Paleolithic *Homo sapiens 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, A.S., 2000b. A 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., Santachiara-Benerecetti, A.S., Falaschi, F., Cavalli-Sforza, L.L., Underhill, P.A., 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am. J. Hum. Genet. 70, 265–268.

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

Spurdle, A., Jenkins, T., 1992. Y chromosome probe p49a detects complex PvuII haplotypes and many new TaqI haplotypes in southern African populations. Am. J. Hum. Genet. 50, 107–125.

Tambets, K., Kivisild, T., Metspalu, E., Parik, J., Kaldma, K., Laos, S., Tolk, H.V., Gölge, M., Demirtas, H., Geberhiwot, T., Papiha, S.S., de Stefano, G.F., Villems, R., 2000. The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of Europe: some preliminary considerations. In: Renfrew, C., Boyle, K. (Eds.), Archaeogenetics: DNA and the Population Prehistory of Europe. Cambridge, pp. 219–235.

Torroni, A., Bandelt, H.J., D'Urbano, L., Lahermo, P., Moral, P., Sellitto, D., Rengo, C., Forster, P., Savontaus, M.L., Bonne-Tamir, B., Scozzari, R., 1998. mtDNA analysis reveals a major late Paleolithic population expansion from southwestern to northeastern Europe. Am. J. Hum. Genet. 62, 1137–1152.

Torroni, A., Bandelt, H.J., Macaulay, V., Richards, M., Cruciani, F., Rengo, C., Martinez-Cabrera, V., Villems, R., Kivisild, T., Metspalu, E., Parik, J., Tolk, H.V., Tambets, K., Forster, P., Karger, B., Francalacci, P., Rudan, P., Janicijevic, B., Rickards,

O., Savontaus, M.L., Huoponen, K., Laitinen, V., Koivumaki, S., Sykes, B., Hickey, E., Novelletto, A., Moral, P., Sellitto, D., Coppa, A., Al-Zahery, N., Santachiara-Benerecetti, A.S., Semino, O., Scozzari, R., 2001. A signal from human mtDNA of postglacial recolonization in Europe. Am. J. Hum. Genet. 69, 844–852.

Torroni, A., Cruciani, F., Rengo, C., Sellitto, D., Lopez-Bigas, N., Rabionet, R., Govea, N., Lopez De Munain, A., Sarduy, M., Romero, L., Villamar, M., del Castillo, I., Moreno, F., Estivill, X., Scozzari, R., 1999. The A1555G mutation in the 12S rRNA gene of human mtDNA: recurrent origins and founder events in families affected by sensorineural deafness. Am. J. Hum. Genet. 65, 1349–1358.

Torroni, A., Huopenen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M.L., Wallace, D.C., 1996. Classification of European mtDNAs from an analysis of three European populations. Genetics 144, 1835–1850.

Torroni, A., Lott, M.T., Cabell, M.F., Chen, Y.S., Lavergne, L., Wallace, D.C., 1994. mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication in the D-loop region. Am. J. Hum. Genet. 55, 760–776.

Torroni, A., Petrozzi, M., D'Urbano, L., Sellitto, D., Zeviani, M., Carrara, F., Carducci, C., Leuzzi, V., Carelli, V., Barboni, P., De Negri, A., Scozzari, R., 1997. Haplotype and phylogenetic analyses suggest that one European-specific mtDNA background plays a role in the expression of Leber hereditary optic neuropathy by increasing the penetrance of the primary mutations 11778 and 14484. Am. J. Hum. Genet. 60, 1107–1121.

Torroni, A., Semino, O., Scozzari, R., Sirugo, G., Spedini, G., Abbas, N., Santachiara-Benerecetti, A.S., 1990. Y chromosome DNA polymorphisms in human populations: differences between Caucasoids and Africans detected by 49a and 49f probes. Ann. Hum. Genet. 54, 287–296.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., Oefner, P.J., 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. 7, 996–1005.

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

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

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

Further Reading

Quintana-Murci, L., Semino, O., Bandelt, H.J., Passarino, G., McElreavey, K., Santachiara-Benerecetti, A.S., 1999a. Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat. Genet. 23, 437–441.

Roewer, L., Arnemann, A.J., Spurr, N.K., Grzeschik, K.H., Epplen, J.T., 1992. Simple repeated sequences on the Y chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89, 389–394.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., de Knijff, P., 1996. Analysis of molecular variance AMOVA of Y-chromosome-specific microsatellites in two closely related human populations. Hum. Mol. Genet. 5, 1029–1033.

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

Whitfield, L.S., Sulston, J.E., Goodfellow, P.N., 1995. Sequence variation on the human Y chromosome. Nature 378, 379–380.