ORIGINAL INVESTIGATION

Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood

Michael F. Hammer · Doron M. Behar · Tatiana M. Karafet · Fernando L. Mendez · Brian Hallmark · Tamar Erez · Lev A. Zhivotovsky · Saharon Rosset · Karl Skorecki

Abstract It has been known for over a decade that a majority of men who self report as members of the Jewish priesthood (Cohanim) carry a characteristic Y chromosome haplotype termed the Cohen Modal Haplotype (CMH). The CMH has since been used to trace putative Jewish ancestral origins of various populations. However, the limited number of binary and STR Y chromosome markers used previously did not provide the phylogenetic resolution needed to infer the number of independent paternal lineages that are encompassed within the Cohanim or their coalescence times. Accordingly, we have genotyped 75 binary markers and 12 Y-STRs in a sample of 215 Cohanim from diverse Jewish communities, 1,575 Jewish men from across the range of the Jewish Diaspora, and 2,099 non-Jewish men from the Near East, Europe, Central Asia, and India. While Cohanim from diverse backgrounds carry a total of 21 Y chromosome haplogroups, 5 haplogroups account for 79.5% of Cohanim Y chromosomes. The most frequent Cohanim lineage (46.1%) is marked by the recently reported P58 T->C mutation, which is prevalent in the Near East. Based on genotypes at 12 Y-STRs, we identify an extended CMH on the J-P58* background that predominates in both Ashkenazi and non-Ashkenazi Cohanim and is remarkably absent in non-Jews. The estimated divergence time of this lineage based on 17 STRs is 3,190 ± 1,090 years. Notably, the second most frequent Cohanim lineage (J-M410*, 14.4%) contains an extended modal haplotype that is also limited to Ashkenazi and non-Ashkenazi Cohanim and is estimated to be 4.2 ± 1.3 ky old. These results support the hypothesis of a common origin of the CMH in the Near East well before the dispersion of the Jewish people into separate communities, and indicate that the majority of contemporary Jewish priests descend from a limited number of paternal lineages.

M. F. Hammer and D. M. Behar contributed equally to this work.

M. F. Hammer · T. M. Karafet · B. Hallmark · T. Erez
ARL Division of Biotechnology, University of Arizona,
Tucson, AZ 85721, USA

M. F. Hammer · F. L. Mendez
Department of EEB, University of Arizona,
Tucson, AZ 85721, USA

D. M. Behar · K. Skorecki
Molecular Medicine Laboratory,
Rambam Health Care Campus, 31096 Haifa, Israel

L. A. Zhivotovsky
Institute of General Genetics,
Russian Academy of Sciences, 119991 Moscow, Russia

S. Rosset
Department of Statistics and Operations Research,
School of Mathematical Sciences, Tel Aviv University,
69978 Tel Aviv, Israel

K. Skorecki (✉)
Rappaport Faculty of Medicine and Research Institute,
Technion, Israel Institute of Technology, 31096 Haifa, Israel
e-mail: skorecki@tx.technion.ac.il

Introduction

More than a decade ago a characteristic Y chromosome haplotype was found to be associated with the Jewish priesthood, a patrilineal dynasty thought to be founded by the first Jewish priest, the biblical Aaron (Skorecki et al.

1997). The sharing of this Y chromosome lineage between Ashkenazi and non-Ashkenazi Jews pointed to a common origin of the Cohanim before the separation of Jewish communities in the Diaspora. Subsequently, it was shown that ~50% of Cohanim carry a characteristic suite of alleles at six Y-linked STRs, which defined the Cohen Modal Haplotype (CMH). Dating based on variation associated with five of these six Y-STRs suggested that contemporary CMH chromosomes trace to a common ancestral chromosome 2,100–3,250 years ago (Thomas et al. 1998). This time roughly corresponds to the period between the biblical exodus and the destruction of the first temple. However, the small number of Y-STRs typed by Thomas et al. (1998) produced a "low resolution" CMH that was shared among many non-Jewish populations (Arredi et al. 2004; Behar et al. 2004; Bonne-Tamir et al. 2003; Cadenas et al. 2008; Cinnioglu et al. 2004; Nebel et al. 2001; Robino et al. 2008; Shen et al. 2004; Zalloua et al. 2008). Similarly, the few binary markers known at the time did not provide the phylogenetic resolution needed to infer the geographic origin of the CMH lineage. Also left unresolved is the paternal ancestry of Cohanim who do not carry Y chromosomes related to the CMH, leaving open the question of the number of founding lineages of the priesthood.

The recent publication of a revised Y chromosome haplogroup tree comprised of >600 binary markers (Karafet et al. 2008) and the availability of dozens of Y-STRs provides the opportunity to determine the origin(s) of Cohanim lineages with greater phylogenetic and geographic resolution than previously possible (Kayser et al. 2004; Lim et al. 2007; Redd et al. 2002). To further elucidate the paternal ancestry of Jewish priests, we genotype 75 binary markers and 22 Y-STRs in a sample of 122 Ashkenazi and 93 non-Ashkenazi Cohanim, and compare patterns of Y chromosome variation with similar data from 3,459 Jewish and non-Jewish men from the Near East, Europe, Central Asia, and India. We define a 12-locus "*extended CMH*" and then determine the phylogenetic position and geographic distribution of this lineage. To explore the origin of this and other Cohanim haplogroups, we estimate the ages of Y-STR diversity associated with several key lineages present at frequencies >5% in our sample of Cohanim. Finally, we discuss models for the origin and maintenance of multiple paternal lineages since the founding of the Jewish priesthood.

Subjects and methods

Population samples

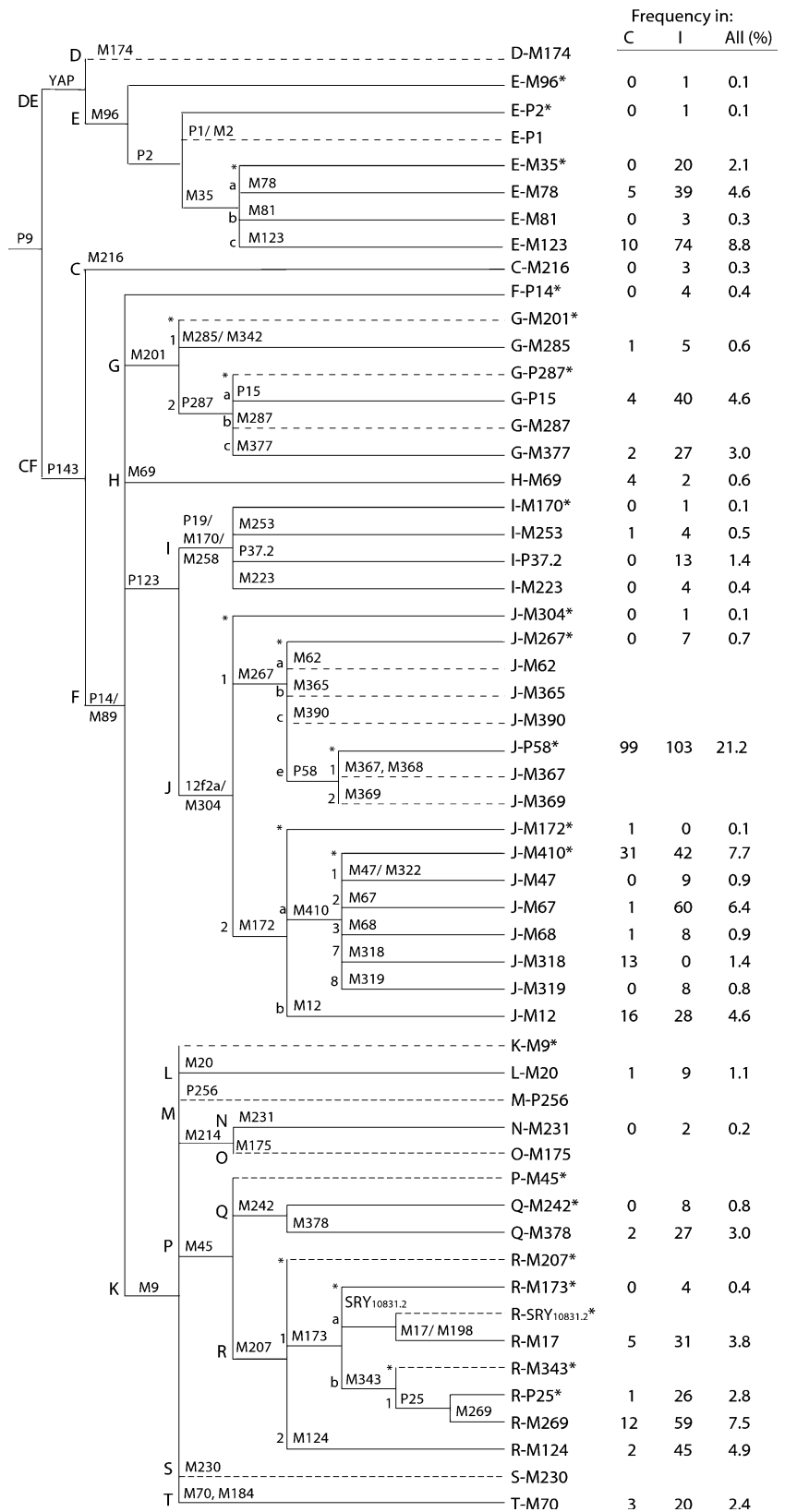Blood or buccal samples were collected from a total of 3,674 individuals according to procedures approved by the Human Subjects Committees at the University of Arizona, Rambam Medical Center and the National Laboratory for the Genetics of Israeli Populations. Volunteers reported the birthplace of their father, grandfather, and in many cases, great grandfather. Jewish volunteers were also asked to report their affiliation to one of the three Jewish castes, Cohen, Levite, or Israelite. Those who did not know their caste status were classified as "unknown". Table S1 lists the Jewish population samples genotyped, which were comprised of unrelated Jewish males representing the major Jewish communities across the Jewish Diaspora ($n = 1,575$), including 215 Cohanim, 738 Israelites, 154 Levites, and 468 of unknown caste status. Table S2 lists our surveyed samples of unrelated non-Jewish men from 30 populations representing Europe, India, the Near East, North Africa, and Central Asia ($n = 2,099$). We note that many additional markers were genotyped in samples that were previously reported, and that Cohanim samples reported here do not overlap with the original collection of Skorecki et al. (1997) and Thomas et al. (1998).

NRY marker analysis

We chose the following set of 75 binary markers to be typed hierarchically in this set of 3,674 chromosomes: Hg BT: $SRY_{10831.1}$; Hg B: M60 or (its equivalent) M181; Hg D: M174, P99, P47; Hg E: M96, P2, P1 or M2, M35, M78, M81, M123; Hg C: M216; Hg FT: P14 or M89; Hg G: M201, M285 or M342, P287, P15, M287, M377; Hg H; M69; Hg IJ: P123; Hg I: P19 or M170 or M258, M253, P37.2, M223; Hg J: 12f2a or M304, M267, M62, M365, M390, P58, M367, M368, M369, M172, M410, M47 or M322, M67, M68, M318, M319, M12; Hg KT: M9; Hg L: M20; Hg M: M256, P35 or M106; Hg NO: M214; Hg N: M231; Hg O: M175, M119, P31, M122; Hg PQR: M45; Hg Q: M242, M378; Hg R: M207, M173, $SRY_{10831.2}$, M17 or M198, M343, P25, M269, M124; Hg S: M230; Hg T: M70, M184. This set of binary markers represents a total of 64 different bifurcations on the NRY phylogeny, as 10 are phylogenetically equivalent (Fig. 1). The genotypes for these sites were determined by multiple techniques such as allele-specific PCR, TaqMan, Kaspar, and direct sequencing. The technical information for detecting these binary polymorphisms has been previously reported by Karafet et al. (2008).

For the microsatellite analysis, 12 short tandem repeats (Y-STRs): DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439) were genotyped in 2 multiplex reactions following the protocol of Redd et al. (2002). To allow more accurate coalescence estimates and better comparison ability with published databases the Cohanim samples were genotyped by an additional set of

Fig. 1 The phylogeny of NRY haplogroups inferred from the panel of 75 binary markers used herein. The polymorphic sites are shown on the *branches*. The number of chromosomes in each haplogroup (counts) for 215 Co-hanim and 737 Israelites is shown at the right of the tree, next to haplogroup frequencies (%) for the entire sample of 952 Jewish chromosomes (*All*)

Haplogroup	C	I	All (%)
D-M174			
E-M96*	0	1	0.1
E-P2*	0	1	0.1
E-P1			
E-M35*	0	20	2.1
E-M78	5	39	4.6
E-M81	0	3	0.3
E-M123	10	74	8.8
C-M216	0	3	0.3
F-P14*	0	4	0.4
G-M201*			
G-M285	1	5	0.6
G-P287*			
G-P15	4	40	4.6
G-M287			
G-M377	2	27	3.0
H-M69	4	2	0.6
I-M170*	0	1	0.1
I-M253	1	4	0.5
I-P37.2	0	13	1.4
I-M223	0	4	0.4
J-M304*	0	1	0.1
J-M267*	0	7	0.7
J-M62			
J-M365			
J-M390			
J-P58*	99	103	21.2
J-M367			
J-M369			
J-M172*	1	0	0.1
J-M410*	31	42	7.7
J-M47	0	9	0.9
J-M67	1	60	6.4
J-M68	1	8	0.9
J-M318	13	0	1.4
J-M319	0	8	0.8
J-M12	16	28	4.6
K-M9*			
L-M20	1	9	1.1
M-P256			
N-M231	0	2	0.2
O-M175			
P-M45*			
Q-M242*	0	8	0.8
Q-M378	2	27	3.0
R-M207*			
R-M173*	0	4	0.4
R-SRY10831.2*			
R-M17	5	31	3.8
R-M343*			
R-P25*	1	26	2.8
R-M269	12	59	7.5
R-M124	2	45	4.9
S-M230			
T-M70	3	20	2.4

the following ten STRs: DYS437, DYS438, DYS447, DYS448, DYS449, DYS454, DYS455, DYS458, DYS459a and DYS459b. For the duplicated microsatel-lites DYS385a,b and DYS459a,b, the short and long scores are reported according to allele size (i.e., without confirming that identical scores observed in two different

samples represent the same locus). PCR products were electrophoresed on a 3730xl Genetic Analyzer (Applied Biosystems) and fragment lengths were converted to repeat number by the use of allelic ladders. Table S3 lists all 215 Cohanim surveyed here and their allele scores for 22 Y-STRs.

Terminology

We follow the terminological conventions recommended by the Y Chromosome Consortium (Karafet et al. 2008) for naming NRY chromosomes. Capital letters A-T identify the 20 major NRY haplogroups and are followed by the names of the binary markers used to assign samples to their positions on the NRY phylogeny. When no further downstream markers in the Karafet et al. (2008) NRY phylogeny were typed, we considered the most derived marker to define a haplogroup (Fig. 1). Haplogroups not defined on the basis of a final derived character state represent interior nodes of the tree and are potentially paraphyletic. In these cases, all binary markers that were excluded by our genotyping strategy are noted within parentheses after an initial "x" symbol. We note that all Cohanim J-P58* chromosomes were found to have the ancestral state at the three downstream markers shown in Fig. 1 (i.e., M367, M368, and M369). Therefore, these chromosomes belong to paragroup J-P58(xM367, M368, M369) (or J1e*); however, we refer to this lineage as J-P58* for simplicity. The term haplotype is used to describe any combination of STRs for a given sample.

Network analysis

Compilation and organization of the data was performed using standard Excel files. Median Joining networks were created using the software NETWORK 4.5.1.0 using the following 12 STRs: DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439. Networks were constructed by the median-joining method (Bandelt et al. 1995). We weighted the STR loci according to their observed variation in our collection of J-P58* Y chromosomes, giving less weight to STRs with higher variance in repeat numbers. To allow maximal resolution we included the duplicated microsatellite DYS385a,b in our network analyses (Niederstatter et al. 2005). While it is possible that identical scores in different samples represent different loci we assume that this potential error is less likely in the case of closely related Cohanim Y chromosomes. We applied the reduced-median algorithm followed by the median-joining algorithm as described at the Fluxus Engineering Web site.

Coalescence analysis

The coalescence times of closely related clusters of haplotypes were estimated using several approaches. First, the age of the Cohanim lineages were calculated as previously reported by Zhivotovsky et al. (2004). The age of STR-variation was computed by averaging across loci the single-locus variances in repeat scores (i.e., with respect to median value at each locus), and then dividing by an average mutation rate of 0.00069 per 25 years. Standard errors were computed across loci as described by Zhivotovsky et al. (2004). The duplicated loci, DYS385ab and DYS459ab, were omitted from the analysis. In addition DYS449 was excluded as it was previously shown to be characterized by multi-repeat variation that substantially differs from all other genotyped STRs (Kayser et al. 2004). We also performed calculations based on a subset of the 17 Y-STRs used in the above analysis. To control for the use of different STRs in our study and that of Zhivotovsky et al. (2004), we calculated divergence times using the same nine Y-STRs in Zhivotovsky et al. (2004): DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, and DYS439. In addition, we used the same five Y-STRs as in the original CMH paper by Thomas et al. (1998): DYS19, DYS390, DYS391, DYS392, and DYS393.

We also estimated coalescence times for key Cohanim lineages by employing the Bayesian Analysis of Trees With Internal Node Generation (BATWING) program of Wilson et al. (2003). To do this within a population genetics framework, we constructed an Ashkenazi population by including our sample of Ashkenazi Israelites and a sub-sample of our Ashkenazi Cohanim to equal 5% of the total population. We used both Y-STR and SNP data to constrain the coalescence of lineages and estimate the TMRCA of individual haplogroups. To estimate the age of a particular Cohanim haplogroup, we excluded non-Cohanim samples carrying this haplogroup (i.e., so that only Cohanim samples carried the particular haplogroup under study). For example, in the case of J-P58*, we included all 317 Ashkenazi Israelite samples that did not carry J-P58*, and a random sample of 17 Cohanim (i.e., which could carry any Y chromosome lineage including J-P58*). We repeated this analysis four times with a different random sample of Cohanim. A phylogenetic tree of binary markers (UEPs) was considered (under option 2) using SNPs that were variable in the Ashkenazi population: M96, M123, P14, M201, M69, P123, M304, M67, P58, M172, M410, M12, M9, M45, M207, M173, M269, M17, and M70. The priors used were: gamma(1.46, 2124) for STR mutation rate, corresponding to a mean of 0.00069 and standard error of 0.00057 as in Zhivotovsky et al. (2004); gamma(1.2, 0.00016) for *N*, corresponding to a mean ancestral popula-

tion size of 7,500 and a standard deviation of 6,847; gamma(1.5, 75) for alpha, corresponding to a mean exponential growth rate of 0.02 and a standard deviation of 0.016; and gamma(1.2, 4) for beta, the time of start of expansion (in units of N generations). These priors returned posteriors of the mutation rate that were similar to the prior for mutation rate. A total of 5×10^4 samples of the program's output were taken after discarding the first 2×10^4 samples as "burn-in". Convergence was confirmed by finding that results of longer runs (i.e., 10^5 MCMC cycles) were similar to those of the shorter runs.

Simulations of the decay of paternal lineages

Simulations of the decay of paternal lineages were performed in Matlab using code written for this purpose. The standard constant population size Wright-Fisher model without mutation was used, since only the loss of lineages due to drift was of interest. Each simulation began by assigning the N individuals a number indicating haplogroup membership. In the first set of simulations, each individual was given a unique number representing N distinct founding lineages, while in the second, each individual was randomly assigned to a haplogroup (with the number of haplogroups varying from 2 to 10) (see Supplementary Material). For all initial conditions 10,000 simulations were performed and the results were averaged. Simulations were run for 160 generations corresponding to between 4 and 5,000 years. We note that a bottleneck or founder event would reduce the number of haplogroups faster on average than we see in a population with constant size, whereas the decay of lineages would be slower on average in an expanding population.

Comparative data

We conducted an extended literature search for Cohanim haplotypes identified here, by comparing allelic scores at as many Y-STRs as possible (i.e., that were typed in our study and the published literature). Allele scores at the 12 STRs that we use to define the *extended CMH* (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439) are 14-13-15-16-13-30-23-10-11-12-11-12, respectively (Table S4). The addition of the four STRs DYS437, DYS438, DYS459a, and DYS455 showed no further variation while the addition of DYS459b and DYS454 demonstrated two additional haplotypes that were each comprised of two samples one mutation step away from the *extended CMH*. The four remaining STRs DYS447, DYS448, DYS449 and DYS458 contained most of the observed variation with 26, 21, 26, and 17.2, respectively, being the most frequent scores observed for these sites. A

similar pattern was found for with two relatively frequent haplotypes that are closely related to the extended CMH at 12 Y-STRs (see "Results and Discussion"). A total of 14 out of the possible 17 STRs were compared with the YHRD database: DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448 and DYS458. The search yielded zero out of 10,243 matching haplotypes in 66 populations for the extended CMH and its two closely related haplotypes. The same STRs demonstrated no matches when compared to Cadenas et al. (2008). No matches were found when the following 12 STRs were used to the dataset reported by Arredi et al. (2004): DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS437, DYS438 and DYS439. Similarly, no matches were found when the following 14 STRs were used to compare with Robino et al. (2008) dataset: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS458. Two matches were found when the following nine STRs were screened in Cinnioglu et al. (2004): DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393 and DYS439; one of which was J1-M369 and the second was J1-M267(xM369). Three Lebanese samples from Zalloua et al. (2008), defined as J-M304(xM172) matched the extended CMH or its two closely related haplotypes when the following 11 STRs were compared: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439. In a recent survey of Hg J-M267, only five of the 282 J-M267 samples studied fell within J-P58; all of these were M367 positive and did not fit within the J-P58* lineage found among our Cohanim sample (Tofanelli et al. 2009). We compared the following 15 STRs overlapping between our paper and Tofanelli et al. (2009) and found no matches: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a, DYS385b, DYS388, DYS437, DYS438, DYS439, DYS448 and DYS458. Available datasets that did not allow comparison of more than seven STRs were not included in our comparative database because they did not allow a sufficiently high level of resolution.

Results and discussion

Genotyping with binary markers yields a total of 37 haplogroups in our Jewish database (Fig. 1). Interestingly, men self-reporting as Cohanim carry Y chromosomes that belong to 21 different haplogroups. However, most of these haplogroups are extremely rare, and a single lineage within the J1 sub-clade of haplogroup J (J-P58*) predominates in both Ashkenazi (51.6%) and non-Ashkenazi (38.7%) Coha-

Fig. 2 The distribution of haplogroup frequencies for all haplogroups present in Ashkenazi and non-Ashkenazi Israelites (*top*) and Cohanim (*bottom*) at a frequency >5%. The following haplogroups are not shown: C-M216, E-M96, E-P2, E-M81, F-P14, I-M170, I-P37.2, I-M223, I-M253, J-M304, L-M20, N-M231, Q-M242, R-M173
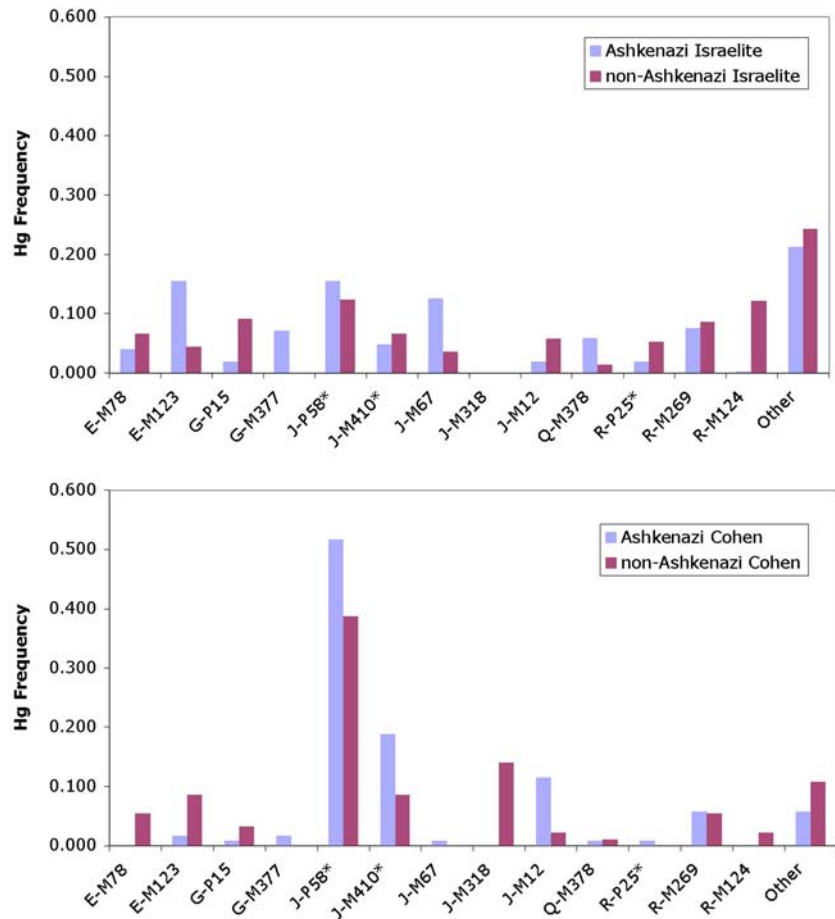
nim in the current sample set (Fig. 2b). Only four of the remaining haplogroups are found at frequencies greater than 5% in all Cohanim sampled, three within the J2 subclade of haplogroup J (J-M410*, 14.4%; J-M12, 7.4%; and J-M318, 6.1%), and one within haplogroup R (R-M269, 5.6%). In contrast, the distribution of haplogroups within Israelites is more uniform, with no single haplogroup reaching a frequency greater than 14% in our Israelite sample (Fig. 2a).

When we genotype the 6 Y-STRs that defined the *original CMH* (DYS19, DYS388, DYS390, DYS391, DYS392, DYS393) (Thomas et al. 1998) in our sample of 99 Cohanim with J-P58* chromosomes, we find that 87 carry a haplotype that is identical to the original modal haplotype and 10 carry haplotypes that are one-step removed from the *original CMH* (i.e., only 2 individuals were 2 or more steps removed). A total of 43 of the 99 chromosomes still match completely when we increase the number of Y-STRs to 12 (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439) (Table S4). We call this 12-locus modal haplotype the *extended CMH*. Figure 3 shows a median-joining network of the 29 12-locus STR haplotypes

associated with Ashkenazi and non-Ashkenazi Cohanim J-P58* chromosomes. One-step mutations at two hypermutable Y-STRs (DYS385 and DYS439) result in two relatively frequent haplotypes (i.e., present at frequencies of 10.1 and 11.1%) that are closely related to the extended CMH (Table S4). The extended CMH and the two closely related haplotypes, which are shared between Ashkenazi and non-Ashkenazi Jews, account for a total of 64.6% of the chromosomes within the Cohanim J-P58* lineage (and 29.8% of Cohanim variation).

The availability of a greater number of binary markers enables examination of the distribution of the original and extended CMH across the branches of a highly resolved Y chromosome haplogroup tree. In our dataset, the original CMH is observed in a total of 215 chromosomes, all of which belong to haplogroup J. Notably, most of these chromosomes are partitioned between the J1 and J2 subclades of the J haplogroup, specifically on the J-P58* and J-M67 lineages. A small number of original CMH chromosomes ($n = 9$) are found within other subclades of Hg J. In contrast, the *extended CMH* and its two closely related haplotypes shown in Fig. 3 are almost entirely limited to P58 chromosomes within the J1 clade: the *extended CMH*

appears outside of J-P58* only once (i.e., within J-M319), while one of its two closely related haplotypes appears within haplogroup J-M67.

A survey of our database confirms that chromosomes carrying the original CMH are not specific to either Cohanim or Jewish populations. The original CMH is present at moderate frequencies (5–8%) in the other Jewish castes (i.e., Levites and Israelites), among non-Jewish Yemenites (13%) and Jordanians (~7%), and as singletons in a number of other non-Jewish populations (Druze, Egyptians, Palestinians, Syrians, Turks, Iranians, Italians, Romanians, and Uzbeks). In contrast, the extended CMH and its two related haplotypes are observed only among Cohanim (29.8%) and Israelites (1.5%) (i.e., it is completely absent from the Levites and non-Jews surveyed here). We also performed a search of the current literature (Arredi et al. 2004; Cadenas et al. 2008; Cinnioglu et al. 2004; Robino et al. 2008; Zalloua et al. 2008) and found a similar pattern: the original CMH is present in several Near Eastern populations, while the extended CMH is extremely rare outside of Jewish populations.

To better estimate the age of the Cohanim J-P58* lineage, we genotyped an additional 10 Y-STRs (DYS437, DYS438, DYS447, DYS448, DYS449, DYS454, DYS455, DYS458, DYS459a and DYS459b) (i.e., a total of 22) in our sample of 215 Cohanim. Interestingly, 4 of the additional 10 Y-STRs (DYS437, DYS438, DYS459a and DYS455) do not further subdivide the group of 43 samples comprising the modal haplotype in Fig. 3. Using the method of Zhivotovsky et al. (2004) (and excluding the duplicated DYS385ab and DYS459ab loci, as well as DYS449, which contains a complex repeat structure) we

estimate the age of Y-STR diversity associated with Cohanim J-P58* chromosomes as $3{,}190 \pm 1{,}090$ (Table 1). To control for differences in mutation rate among loci, we also calculate divergence times using the nine loci that Zhivotovsky et al. (2004) employed to estimate the effective mutation rate of Y-STRs. Similar age estimates are returned for our set of 17 Y-STRs and Zhivotovsky et al.'s (2004) set of 9 Y-STRs in our sample of all 99 Cohanim, as well as in our sample of 63 Ashkenazi Cohanim (Table 1). We note that estimates of the age of the J-P58* lineage are lower when using the five Y-STRs that were employed in the original CMH study of Thomas et al. (1998) (Table 1). This effect is exaggerated when assuming the pedigree mutation rate of 0.0021/generation, which was implemented in the Thomas et al. (1998) calculation (i.e., we obtain a J-P58* lineage divergence time estimate of 0.4 ± 0.2 kyears). A Bayesian-based coalescence analysis using BATWING (Wilson et al. 2003) on a constructed Ashkenazi population comprised of 95% Israelites and 5% Cohanim yields an average median TMRCA for the Cohanim J-P58* lineage of 4,415 years (95% CI 1,130–21,530 years) (Table S5).

In sum, the high frequency of a closely related set of J-P58* chromosomes among Ashkenazi and non-Ashkenazi Cohanim that share a common modal haplotype, and that are estimated to have diverged from a common ancestor >2,000 years ago, is consistent with the hypothesis that the J-P58* lineage traces the Cohanim dynasty to a time before the Jewish diaspora. While the frequency of the J-P58* lineage is higher among Ashkenazi Jews (Fig. 2a), Y-STR variation associated with this haplogroup is older in the non-Ashkenazi community (e.g., we obtained divergence time estimates of 4.6 ± 1.8 and 3.5 ± 2.1 kyears for the

Fig. 3 Network of J-P58* haplotypes observed within Ashkenazi (*black*) and non-Ashkenazi Cohanim (*white*). The following STRs comprise the network: DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439. *Circle areas* are proportional to haplotype frequency with the *smallest circles* representing singletons. The branch lengths are proportional to the number of STRs separating the nodes
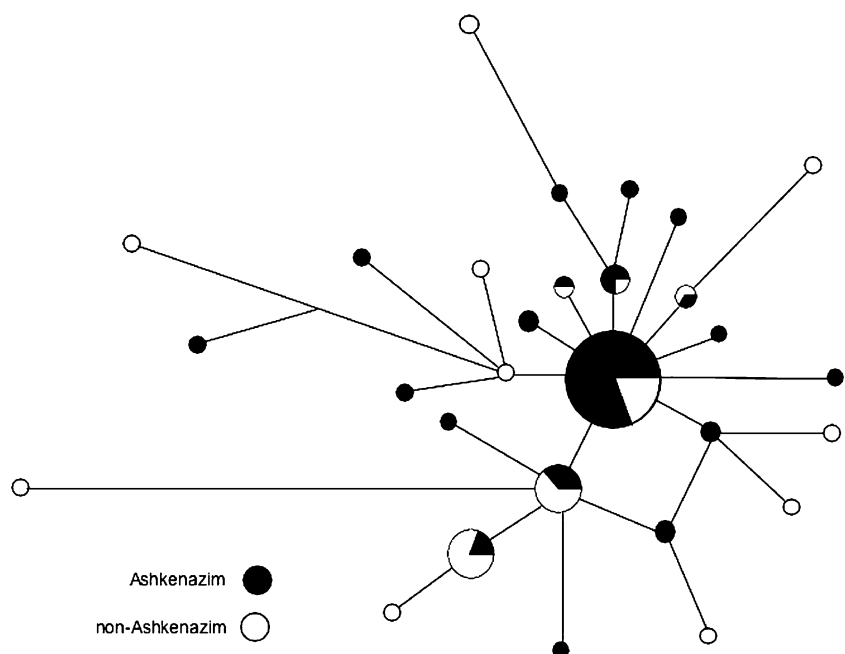
Ashkenazim ●

non-Ashkenazim ○

Table 1 Divergence time (based on method of Zhivotovsky et al. (2004)) (ky) (mean ± SE) of Cohanim lineages based on Y-STR loci

Haplogroup	Cohanim sample (N)	17 loci[a]	9 loci[b]
J-P58*	All (99)	3.2 ± 1.1	3.0 ± 1.5
	Ashkenazi (63)	2.4 ± 0.8	2.8 ± 1.2
J-M410*	All (31)	5.9 ± 2.0	4.9 ± 1.9
	All (29)[c]	4.2 ± 1.3	3.8 ± 1.4
	Ashkenazi (23)	3.8 ± 1.2	4.2 ± 1.8
J-M12	All (16)	12.1 ± 4.4	5.5 ± 1.9
	All (14)[c]	3.4 ± 1.2	4.0 ± 1.8
	Ashkenazi (14)	6.7 ± 2.5	4.3 ± 2.0
	Ashkenazi (13)[c]	3.0 ± 1.3	4.3 ± 2.0
J-M318	Non-Ashkenazi (13)	1.3 ± 0.5	1.9 ± 0.8
R-M269	All (12)	11.5 ± 2.8	14.1 ± 4.1

[a] This study, see Methods for list of loci included

[b] Zhivotovsky et al. (2004); DYS19, DYS388, DYS389I. DYS389II, DYS390, DYS391, DYS392, DYS393, DYS439

[c] Divergent haplotypes removed (J-M410, 2 non-Ashkenazi Cohanim; JM12, 1 Ashkenazi and 1 non-Ashkenazi Cohen)

17- and 9-locus datasets, respectively). In this regard, it is also worth noting that the J-P58* network topology suggests population expansion, especially within the Ashkenazim. This may be attributable to the strong founder effect previously suggested for the Ashkenazi population (Behar et al. 2004, 2006).

Our results also document a set of non-J-P58* lineages that are carried collectively by more than 50% of Cohanim. Some of these lineages are members of the J2 subclade, while others are more distantly related (i.e., within haplogroups R and E) (Figs. 1, 2). To further explore the origin of these Cohanim lineages, we estimate the ages of Y-STR diversity associated with the J-M410*, J-M318, J-M12, and R-M269 lineages (Table 1), all of which are found at frequencies >5% in the Cohanim sample groups examined here (Fig. 2b). J-M410*, which is carried by both Ashkenazi (18.9%) and non-Ashkenazi (8.6%) Cohanim, is also found in ∼5% of non-Cohanim Jews, as well as in many non-Jewish populations from the Near East (data not shown). Moreover, there is a modal Cohanim haplotype that is shared between Ashkenazi and non-Ashkenazi (North African and Sephardi) communities (Figure S1), and absent from non-Jewish populations. Divergence time estimates of Cohanim J-M410* chromosomes based on 17 and 9 Y-STRs range between 5.9 ± 2.0 and 4.9 ± 1.9 kyears, respectively. However, median-joining networks constructed from our Cohanim and non-Jewish data indicate that two Cohanim individuals carry divergent haplotypes that do not appear to descend from a common (modal) cluster of Cohanim J-M410* chromosomes (Figure S2). When we exclude these two divergent haplotypes from the analy-

sis, we obtain divergence time estimates of 4.2 ± 1.3 and 3.8 ± 1.4 kyears for 17 and 9 Y-STRs, respectively. Our BATWING analysis returns coalescence time estimates of 3.2 kyears (95% CI, 0.7–16.7 kyears) (Table S5), similar to the divergence time estimates for Ashkenazi Cohanim in Table 1. These results support the hypothesis that J-M410* represents a second major founding lineage of the Cohanim, coalescing to a point within the early history of the ancient Hebrews of the Near East.

Similar results are obtained for the less frequent J-M12 lineage, which is carried by 16 Cohanim in our survey (14 of which are of Ashkenazi descent). As in the case of J-M410*, a median-joining network suggests that 2 of 16 individuals (1 Ashkenazi and 1 non-Ashkenazi Cohen) carry divergent haplotypes that may have entered the Cohanim population recently (Figure S3). Divergence time estimates made after removing these individuals are comparable to those for the Cohanim J-P58* and J-M410* lineages (3.4 ± 1.2 and 4.0 ± 1.8 kyears for the 17- and 9-locus datasets, respectively) (Table 1). In contrast, our network and divergence time analyses suggest that R-M269 chromosomes entered the Cohanim population via several "migration" events, and do not represent a single Cohanim founding lineage. For example, divergence time estimates are much older for Cohanim R-M269 chromosomes (>10 kyears) than for the three Cohanim lineages in haplogroup J discussed above, and median-joining networks of Cohanim R-M269 chromosomes lack a modal haplotype and show many unrelated singleton haplotypes that are interspersed among Cohanim and non-Jewish samples (Figure S4).

We note that divergence times for the J-P58*, J-M410, and J-M12 lineages are not statistically significantly different from one another as a result of the large standard deviations in Table 1. Moreover, we cautiously interpret dating of lineages that are not defined by the derived state at a terminal SNP (i.e., are internal nodes on the Y chromosome tree) and those that we have not typed all known downstream SNPs (J-M410*/J-P58* and J-M12, respectively) because subsets of chromosomes within these lineages may be marked by undiscovered SNPs. Our estimate of the age of J-M318 may be more reliable because this SNP represents a terminal mutation within the M410 sub-clade of the J2 branch (Fig. 1). Originally discovered in a single Libyan Jew (Shen et al. 2004), we find the derived allele at M318 to be present in 16 individuals in this survey—13 of which are Cohanim from Tunisia/Libya or the island of Jerba (the remaining 3 samples come from Tunisian or Libyan Jews who did not have information on their Cohen, Levite, or Israelite status). The much younger estimated divergence time for the J-M318 haplogroup (1.3 ± 0.5 and 1.9 ± 0.8 kyears for the 17- and 9-locus datasets, respectively) (Table 1) suggests that either the M318 mutation (a) arose within the Cohanim population of North Africa,

(b) expanded within this community following migration of a founding J-M318 Cohen from another geographic location, or (c) became incorporated into the Cohanim patriline via conversion, adoption or non-paternity. The first of these possibilities (a) is supported by the fact that the M318 mutation occurred on the M410 background (Fig. 1), and median-joining network analysis links the cluster of Cohanim J-M318 chromosomes to that of the Cohanim J-M410* chromosomes (i.e., rather than to other J-M410* chromosomes from North Africa and the Near East) (Figure S5). The high frequency (∼60%) of the otherwise rare J-M318 haplogroup in our sample from the island of Jerba may be the result of an ancient founder effect in this Jewish isolate, which is thought to be descended from one of the earliest Diaspora communities that left the Middle East before the destruction of the second Temple in 70 A.D. (Tessler and Hawkins 1980).

Here, we discuss alternative explanations for presence of several founding lineages within the Cohanim. One possibility is that multiple males were designated as Cohanim early in the establishment of the priesthood. We performed exploratory simulations to assess the likelihood of survival of multiple paternal lineages in the history of the Ashkenazi Cohanim. The probability of survival of more than a single haplogroup depends mainly on the population size and to a lesser extent, on the number of haplogroups that are assumed to have founded the initial Cohanim group. For example, if we begin with an initial population of 50 Cohanim carrying a total of 10 haplogroups, we find that there is a very low probability of survival of more than a single haplogroup after 120 generations. We obtain a similar result if we begin the simulation with 50 males each carrying a unique haplogroup (Figure S6). However, if we begin with 100 males carrying 10 haplogroups the mean number of haplogroups surviving for 120 generations is >1 (Figure S7). Thus, there would be a reasonably high probability that more than a single Cohen haplogroup could have survived in the Ashkenazi population since the initial founding of the priesthood ∼3,000 years ago (Thomas et al. 1998) if we would be willing to accept an initial founding population size of >50 priests. However, our simulation results also suggest that it is highly unlikely that as many haplogroups as we actually observe (e.g. Fig. 2) would persist under this simple model. Another model that deserves consideration is a metapopulation (Wakeley 2004) in which semi-isolated communities maintain multiple Cohen lineages, each with a certain probability of extinction and replacement. In this model, multiple Cohanim lineages would then persist in the entire population, and new lineages would be expected to accrue among Cohanim over time. The presence of several founding lineages among the Cohanim of this survey—both shared between or specific to the Ashkenazi and non-Ashkenazi communities, as well as highly variable frequencies of these lineages among sub-populations within Ashkenazi and non-Ashkenazi communities (data not shown), may lend support to a metapopulation model. Mutation alone does not provide an explanation for the multiplicity of Cohanim haplogroups, because the ages of most of these haplogroups predate the foundation of the Jewish people (Cruciani et al. 2006; Karafet et al. 2008; Semino et al. 2004). Indeed, our divergence time estimates for the J-P58*, J-M410*, J-M12 lineages based on variation at the set of 9 STRs in our Israelite population sample are 19.0 ± 5.6, 22.6 ± 2.9, and 15.1 ± 3.1 kyears, respectively (data not shown).

In conclusion, we demonstrate that 46.1% (95% CI = 39–53%) of Cohanim carry Y chromosomes belonging to a single paternal lineage (J-P58*) that likely originated in the Near East well before the dispersal of Jewish groups in the Diaspora. Support for a Near Eastern origin of this lineage comes from its high frequency in our sample of Bedouins, Yemenis (67%), and Jordanians (55%) and its precipitous drop in frequency as one moves away from Saudi Arabia and the Near East (Fig. 4). Moreover, there is a striking contrast between the relatively high frequency of J-58* in Jewish populations (∼20%) and Cohanim (∼46%) and its vanishingly low frequency in our sample of non-Jewish populations that hosted Jewish diaspora communities outside of the Near East. An extended Cohen Modal Haplotype accounts for 64.6% of chromosomes with the J-P58* background, and 29.8% (95% CI = 23–36%) of Cohanim Y chromosomes surveyed here. These results also confirm that lineages characterized by the 6 Y-STRs used to define the original CMH are associated with two divergent sub-clades within haplogroup J and, thus, cannot be assumed to represent a single recently expanding paternal lineage. By combining information from a sufficient number of SNPs and STRs in a large sample of Jewish and non-Jewish populations we are able to resolve the phylogenetic position of the CMH, and pinpoint its geographic distribution. Our estimates of the coalescence time also lend support to the hypothesis that the extended CMH represents a unique founding lineage of the ancient Hebrews that has been paternally inherited along with the Jewish priesthood. However, the sharing of several less frequent haplogroups (and modal haplotypes within these haplogroups) between Ashkenazi and non-Ashkenazi communities, as well as evidence for the persistence of population-specific Cohanim haplogroups, supports the formulation that males from other remote lineages also contributed to the Jewish priesthood, both before and after the separation of Jewish populations in the Diaspora. Genotyping a larger sample of Cohanim Y chromosomes from other divergent haplogroups may further elucidate the complex paternal history

Fig. 4 Geographic distribution of J-P58* chromosomes for all populations listed in Tables S1 and S2. The frequency of J-P58* chromosomes for each population is indicated in *black*
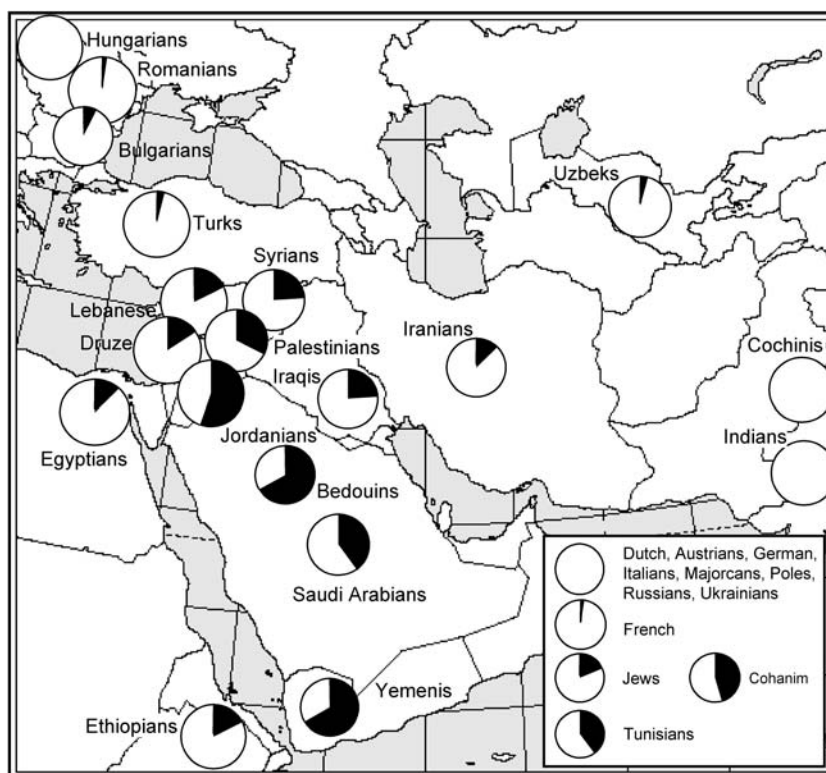
of Jewish priests, and aid in the identification of lost tribes claiming ancient Hebrew ancestry.

Web resources

http://www.fluxus-technology.com/.
http://www.yhrd.org/.

References

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, Tyler-Smith C (2004) A predominantly neolithic origin for Y-chromosomal DNA variation in North Africa. Am J Hum Genet 75:338–345

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743–753

Behar DM, Garrigan D, Kaplan ME, Mobasher Z, Rosengarten D, Karafet TM, Quintana-Murci L, Ostrer H, Skorecki K, Hammer MF (2004) Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. Hum Genet 114:354–365

Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, Skorecki K (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet 78:487–497

Bonne-Tamir B, Korostishevsky M, Redd AJ, Pel-Or Y, Kaplan ME, Hammer MF (2003) Maternal and paternal lineages of the Samaritan isolate: mutation rates and time to most recent common male ancestor. Ann Hum Genet 67:153–164

Cadenas AM, Zhivotovsky LA, Cavalli-Sforza LL, Underhill PA, Herrera RJ (2008) Y-chromosome diversity characterizes the Gulf of Oman. Eur J Hum Genet 16:374–386

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

Cruciani F, La Fratta R, Torroni A, Underhill PA, Scozzari R (2006) Molecular dissection of the Y chromosome haplogroup E-M78 (E3b1a): a posteriori evaluation of a microsatellite-network-based approach through six new biallelic markers. Hum Mutat 27:831–832

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838

Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A,

Tyler-Smith C (2004) A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183–1197

Lim SK, Xue Y, Parkin EJ, Tyler-Smith C (2007) Variation of 52 new Y-STR loci in the Y Chromosome Consortium worldwide panel of 76 diverse individuals. Int J Legal Med 121:124–127

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

Niederstatter H, Berger B, Oberacher H, Brandstatter A, Huber CG, Parson W (2005) Separate analysis of DYS385a and b versus conventional DYS385 typing: is there forensic relevance? Int J Legal Med 119:1–9

Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler JM, Hammer MF (2002) Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 130:97–111

Robino C, Crobu F, Di Gaetano C, Bekada A, Benhamamouch S, Cerutti N, Piazza A, Inturri S, Torre C (2008) Analysis of Y-chromosomal SNP haplogroups and STR haplotypes in an Algerian population sample. Int J Legal Med 122:251–255

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

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, Oefner PJ (2004) Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. Hum Mutat 24:248–260

Skorecki K, Selig S, Blazer S, Bradman R, Bradman N, Waburton PJ, Ismajlowicz M, Hammer MF (1997) Y chromosomes of Jewish priests. Nature 385:32

Tessler MA, Hawkins LL (1980) The political culture of Jews in Tunisia and Morocco. Int J Middle East Stud 11:59–86

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

Tofanelli S, Ferri G, Bulayeva K, Caciagli L, Onofri V, Taglioli L, Bulayev O, Boschi I, Alu M, Berti A, Rapone C, Beduschi G, Luiselli D, Cadenas AM, Awadelkarim KD, Mariani-Costantini R, Elwali NE, Verginelli F, Pilli E, Herrera RJ, Gusmao L, Paoli G, Capelli C (2009) J1-M267 Y lineage marks climate-driven prehistorical human displacements. Eur J Hum Genet. doi:10.1038/ejhg.2009.58 (advance online publication 15 April 2009)

Wakeley J (2004) Metapopulation models for historical inference. Mol Ecol 13:865–875

Wilson IJ, Weale ME, Balding DJ (2003) Inferences from DNA data: population histories, evolutionary processes and forensic match probabilities. J R Stat Soc Series A (Stat Soc) 166:155–188

Zalloua PA, Xue Y, Khalife J, Makhoul N, Debiane L, Platt DE, Royyuru AK, Herrera RJ, Hernanz DF, Blue-Smith J, Wells RS, Comas D, Bertranpetit J, Tyler-Smith C (2008) Y-chromosomal diversity in Lebanon is structured by recent historical events. Am J Hum Genet 82:873–882

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