Increased Resolution of Y Chromosome Haplogroup T Defines Relationships among Populations of the Near East, Europe, and Africa

Author(s): Fernando L. Mendez, Tatiana M. Karafet, Thomas Krahn, Harry Ostrer, Himla Soodyall, and Michael F. Hammer
Source: Human Biology, 83(1):39-53. 2011.
Published By: Wayne State University Press
DOI: http://dx.doi.org/10.3378/027.083.0103
URL: http://www.bioone.org/doi/full/10.3378/027.083.0103

Increased Resolution of Y Chromosome Haplogroup T Defines Relationships among Populations of the Near East, Europe, and Africa

FERNANDO L. MENDEZ,[1,2] TATIANA M. KARAFET,[1] THOMAS KRAHN,[3] HARRY OSTRER,[4] HIMLA SOODYALL,[5] AND MICHAEL F. HAMMER[1,2]

Abstract Increasing phylogenetic resolution of the Y chromosome haplogroup tree has led to finer temporal and spatial resolution for studies of human migration. Haplogroup T, initially known as K2 and defined by mutation M70, is found at variable frequencies across West Asia, Africa, and Europe. While several SNPs were recently discovered that extended the length of the branch leading to haplogroup T, only two SNPs are known to mark internal branches of haplogroup T. This low level of phylogenetic resolution has hindered studies of the origin and dispersal of this interesting haplogroup, which is found in Near Eastern non-Jewish populations, Jewish populations from several communities, and in the patrilineage of President Thomas Jefferson. Here we map 10 new SNPs that, together with the previously known SNPs, mark 11 lineages and two large subclades (T1a and T1b) of haplogroup T. We also report a new SNP that links haplogroups T and L within the major framework of Y chromosome evolution. Estimates of the timing of the branching events within haplogroup T, along with a comprehensive geographic survey of the major T subclades, suggest that this haplogroup began to diversify in the Near East ~25 kya. Our survey also points to a complex history of dispersal of this rare and informative haplogroup within the Near East and from the Near East to Europe and sub-Saharan Africa. The presence of T1a2 chromosomes in Near Eastern Jewish and non-Jewish populations may reflect early exiles between the ancient lands of Israel and Babylon. The presence of different subclades of T chromosomes in Europe may be explained by both the spread of Neolithic farmers and the later dispersal of Jews from the Near East. Finally, the moderately high frequency (~18%) of T1b* chromosomes in the Lemba of southern Africa supports the hypothesis of a Near Eastern, but not necessarily a Jewish, origin for their paternal line.

[1]ARL Division of Biotechnology, University of Arizona, Tucson, AZ 85721.
[2]Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721.
[3]FamilyTreeDNA, Houston, TX 77008.
[4]Human Genetics Program, Department Pediatrics, New York University School of Medicine, New York, NY 10016.
[5]Division of Human Genetics, University of Witwatersrand, Johannesburg, South Africa.
Corresponding Author: Michael F. Hammer, ARL Division of Biotechnology, Life Sciences South 231, University of Arizona, Tucson, AZ 85721, Tel: 520-621-9828, Fax: 520-626-8050, Email: mfh@u.arizona.edu.

Human Biology, February 2011, v. 83, no. 1, pp. 39–53

KEY WORDS: Y CHROMOSOME, NEOLITHIC, JEWISH, LEMBA.

Y chromosome markers can be informative for tracing the origin and dispersal patterns of ancestral human groups. This is particularly true for paternal lineages that are at sufficiently high frequency in geographically localized populations. Haplogroup T is a candidate for such a lineage because it is extremely rare in most global surveys and yet appears at intermediate frequencies in the Near East and East Africa (King et al. 2007). However, the lack of markers within this Y chromosome clade has hitherto hindered its use in phylogeographic studies. Originally known as K2 and defined by mutation M70 (YCC 2002), the phylogenetic status of this lineage was recently elevated to 'major haplogroup' after the discovery of several new lineage-defining mutations (Karafet et al. 2008). The lineage marked by M70 received considerable attention because it was found to be associated with the patrilineage of President Thomas Jefferson, who was hypothesized to have fathered a son with his slave, Sally Hemings (Foster et al. 1998). While the hypothesis that haplogroup T originated in the general region of the Near East is well accepted (Nogueiro et al. 2010), the specific source of European T chromosomes is still unclear. For example, did this haplogroup spread to Europe as part of the Neolithic expansion from the Near East (Balaresque et al. 2010), or through later dispersals into Europe by Jews or traders from the Near East?

To better resolve the origin and diversification of T chromosomes, we report several new markers that further delineate the internal structure of this major clade and clarify the position of haplogroup T in the larger framework of the Y chromosome phylogenetic tree. The relationship among different branches of paragroup K, to which T belongs, has been partially clarified by the finding of M526 (Chiaroni et al. 2009). While the M526 mutation is known to be derived in most major branches of the Y chromosome phylogeny (K, M, N, O, P, and S) and ancestral in haplogroups T and L, the relationship among lineages carrying the ancestral and derived states at M526 has yet to be determined. Of the 12 novel single-nucleotide polymorphisms (SNPs) reported here, one (P326) marks a higher-level branch (i.e., analogous to M526 in the Y chromosome phylogeny hierarchy) that represents the ancestor of both haplogroups T and L. Ten other SNPs define eight new sub-branches within the T clade. These new markers provide improved phylogenetic and temporal resolution for the diversification of T lineages.

Methods

Samples. Three sets of samples were genotyped in this study. The first set was used to estimate the frequency of the T haplogroup and its sub-branches in 29 populations (Table 1 and Supplemental Table 1). The second set was used to estimate population demographic parameters and is a subset of the first group (Supplemental Table 1). The restrictions of the second set come both from minimizing population structure when combining samples from different populations and only including samples that were typed successfully for all 10 short tandem repeats (Y-STRs) (see below). Finally, the third set,

Table 1. Frequency of Haplogroup T and Its Sub-Clades in the Surveyed Populations

Population	n	*T*	*M184**	*M70**	*PS21**	*P77*	*P330*	*P321**	*P317*	*L131**	*P322**	*P327*
North African												
Egyptians	150	0.067	—	—	0.060	0.007	—	—	—	—	—	—
Tunisians	34	0.000	—	—	—	—	—	—	—	—	—	—
Ethiopians	58	0.069	—	—	0.069	—	—	—	—	—	—	—
Near Eastern												
Palestinians	115	0.052	—	—	0.009	0.009	—	—	—	—	—	0.035
Bedouins	28	0.000	—	—	—	—	—	—	—	—	—	—
Druze	39	0.077	—	—	0.077	—	—	—	—	—	—	—
Jordanians	187	0.032	—	—	0.021	0.005	—	—	—	0.005	—	—
Lebanese	34	0.000	—	—	—	—	—	—	—	—	—	—
Syrians	95	0.054	0.011	—	0.021	0.011	—	0.011	—	—	—	—
Turks	284	0.011	—	—	0.004	—	—	—	—	0.007	—	—
Assyrians	31	0.129	—	—	0.097	—	0.032	—	—	—	—	—
Iraqis	36	0.056	—	—	0.028	0.028	—	—	—	—	—	—
Iranians	73	0.014	—	—	0.014	—	—	—	—	—	—	—
Saudi Arabians	33	0.000	—	—	—	—	—	—	—	—	—	—
Yemenis	18	0.000	—	—	—	—	—	—	—	—	—	—
North African Jews												
Moroccan Jews	54	0.074	—	—	0.056	0.019	—	—	—	—	—	—
Tunisian Jews	10	0.000	—	—	—	—	—	—	—	—	—	—
Ethiopian Jews	21	0.048	—	—	0.048	—	—	—	—	—	—	—
Near Eastern Jews												
Kurdish Jews	50	0.180	—	—	0.080	0.100	—	—	—	—	—	—
Iraqi Jews	32	0.219	—	0.094	—	0.125	—	—	—	—	—	—
Iranian Jews	22	0.136	—	0.091	—	0.045	—	—	—	—	—	—
Yemenite Jews	44	0.068	—	—	—	0.045	—	—	—	—	0.023	—
Central Asian Jews												
Uzbeki Jews	9	0.000	—	—	—	—	—	—	—	—	—	—
Sephardic Jews												
Bulgarian Jews	42	0.048	—	—	—	—	—	—	—	0.048	—	—
Turkish Jews	34	0.059	—	—	0.029	0.029	—	—	—	—	—	—
Other Jewish Groups												
Roman Jews	53	0.057	—	—	0.019	—	—	—	0.038	—	—	—
Ashkenazi Jews	587	0.010	—	—	0.007	—	—	—	—	0.003	—	—
Europeans												
Bulgarians	29	0.000	—	—	—	—	—	—	—	—	—	—
Sub-Saharan Africans												
Lemba	34	0.176	—	—	—	—	—	—	—	0.176	—	—

which is used in the building of a network of haplotypes and in one of two TMRCA estimation methods used below, consists of samples that belong to haplogroup T and that have been successfully genotyped for 24 Y-STRs (Supplemental Table 1). The third set is mostly a subset of the first set, except that it also incorporates seven European samples and one Jewish sample from Israel. The third set contains samples that are not present in the second set because they belong to populations that were excluded to reduce the effect of population structure. All sampling procedures were approved by the University of Arizona Human Subjects Committees. The Lemba samples were obtained with either verbal or written consent with approval from the Committee for Research on Human Subjects, University of the Witwatersrand (protocol number M980553).

Y-Chromosome Genotyping. We genotyped 98 samples previously collected for Y chromosome studies that were known to belong to paragroup K and not to any of the following haplogroups: K1, K2, K3, K4, L, M, N, O, P, and S. All samples were typed either for M70, M184, or both. Information on M70, M184, M193, M272, M320, and P77 mutations can be found in Karafet et al. (2008) and in references therein. The mutations here named PS2, PS21, and PS78 (information retrieved from dbSNP built 128) were found together with M70 and M184 in a study on structural variation in the Y chromosome (Repping et al. 2006). The authors did not provide explicit names for the mutations. These mutations were submitted to dbSNP as S00002, S00021, S00078, S00046, and S00034, respectively, with "PAGE" as handle. More recently, these five mutations together with a sixth mutation, PS129 (submitted as S00129 with the same handle) were published as part of a larger resequencing study (Rozen et al. 2009). The mutations L131, L206, P317, P321, P322, P326, P327, P328, and P330 are reported here for the first time and were discovered during the course of this study. Mapping information as well as primer sequences for all the SNPs in this paper can be found in the Supplemental Table 2. P326 was typed in all major sub-branches of the K haplogroup, as well as in two samples belonging to K, but not to any major sub-branch (K*).

We characterized the frequency of the haplogroup T and its sub-branches in 2236 samples from 29 populations (set 1 above and Table 1). For the purpose of estimating times of divergence, we surveyed variation in a subset of the previous group, with samples from the Near East, Europe, and Africa (Supplemental Table 1), for a total of 1279 samples. These samples were genotyped for 10 Y-STRs: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, and DYS439 in two multiplex reactions (Redd et al. 2002) and came from a group of 23 populations that clustered in Principal Component Decomposition analyses (i.e., after outliers were removed, see below) (Supplemental Table 1). Besides being genotyped for the abovementioned Y-STRs, samples carrying haplogroup T chromosomes were further genotyped for DYS385a, DYS385b, DYS437, DYS438, DYS442, DYS447, DYS448, DYS449, DYS454, DYS455, DYS456, DYS458, DYS460, DYS570, DYS576, DYS607, and H4 according to a published protocol (Redd et al. 2006).

Data Analyses. Reduced median-joining networks (Bandelt et al. 1999) were constructed using the NETWORK 4.5c program based on 24 microsatellites and eight SNPs. We excluded the monomorphic DYS388 and DYS426 loci, as well as DYS389II because of the occurrence of a null allele. For network calculations, STRs were inversely weighted according to their repeat number as described elsewhere (Karafet et al. 2005). The highest weight(10) was assigned to SNPs. The reduced median output was used as input for the median-joining network.

Two different methods were applied to estimate the TMRCA of different nodes of the Y chromosome tree. We used Td, which is based on the stepwise mutation model (Kimura and Ohta 1978) and assumes that the average of the

square number of steps from the ancestral type for a given STR is an unbiased estimator of the mutation rate times the TMRCA (Zhivotovsky et al. 2004). An average Y-STR mutation rate of 0.00069 per 25 yrs was used, and the median value for each STR was assumed to be the ancestral state. Point estimates were averaged and standard errors were calculated across loci from those point estimates as in Zhivotovsky et al. (2004). In addition to the 10 Y-STRs genotyped in the second set of samples, DYS438, DYS457, DYS458, DYS455, DYS454, DYS447, and DYS448 were also used in the Td analysis.

We also used the Bayesian Analysis of Trees With Internal Node Generation (BATWING) program (Wilson 2003) to estimate TMRCA values, using data from both the set of 10 Y-STRs and Y-SNPs (Supplemental Methods). The results were also combined with TMRCA estimations obtained from the distribution of SNPs in the genealogy (see below). Given the low frequency of T chromosomes, individuals from several populations were pooled under the assumption that they could be treated as coming from a single population. The populations originally considered were Egyptians, Tunisians, Ethiopians, Lemba, Turks, Assyrians, Druze, Bedouins, Palestinians, Syrians, Jordanians, Lebanese, Iraqis, Iranians, Saudi Arabians, Yemenis, Bulgarians, Bulgarian Jews, Turkish Jews, Moroccan Jews, Tunisian Jews, Yemenite Jews, Kurdish Jews, Iraqi Jews, Iranian Jews, Uzbeki Jews, Ashkenazi Jews, and Roman Jews. Ethiopian Jews were not successfully genotyped for the 10 Y-STRs and were excluded from this analysis. To comply with the assumptions of the coalescence analysis, samples were included regardless of their haplogroup affiliation. We sought to reduce the effect of population structure by sequentially performing a Principal Component Decomposition (PCD) on the population haplogroup frequencies and then excluding the population that appeared isolated in a plot of the first two principal components. After the sequential removal of Ethiopians, Iranian Jews, Uzbeki Jews, Lemba, and Tunisians, no further outliers could be seen in the PCD plot (Supplemental Figure 1). The samples successfully genotyped in these populations ($n = 1279$) were chosen for BATWING analysis.

The model used for BATWING assumed a single population of constant size N until a time T_e in the past at which the population started to grow exponentially at fixed rate α per generation (Supplemental Methods). The prior distributions used in BATWING were Γ (1.24, 1800) for the mutation rate, Γ (1.3, 0.00016) for N, Γ(1, 75) for α, and Γ (0.9, 4) for β (scaled time of the beginning of population growth). The generation time was assumed to be 25 yrs. Posterior distributions for the minimum age of the mutations are reported as TMRCA for the haplogroups. The ages of independent branches were compared by examining the distribution of the ratios of coalescent times defining those branches. Similarly, taking the age of the K haplogroup as a reference, and using flat priors for the ages of subsequent branching events, the likelihood of the joint relative ages of branching events within a lineage were estimated directly from the posterior distribution provided by BATWING.

Given a set of SNPs uniformly ascertained and mapped onto the genealogy, it is possible to extract information on the age of the different branching points

by analyzing the distribution of the SNPs in the genealogy (Karafet et al. 2008). Using six SNPs that were uniformly ascertained in resequencing studies (Repping et al. 2006, Rozen et al. 2009), the ages of three of the branching points (M184, M70, and PS21) were estimated. With this method, described in more detail in the Supplemental Methods, it is possible to estimate the likelihood of the relative branch lengths in a genealogy, given the distribution of mutations. This likelihood is combined with the likelihood obtained from BATWING to generate a new likelihood for the relative ages of the branches. Point estimates result from the values with the maximum likelihood (MLE) and the two-tailed 95% confidence intervals are calculated from the likelihood ratio test. Under the assumption of uniform ascertainment, it is also possible to estimate the age of the common ancestor of T and L, namely TL (Supplemental Methods).

Results and Discussion

Improved Phylogenetic Resolution of Haplogroup T. Karafet et al. (2008) published the first Y chromosome phylogeny with haplogroup T, a lineage marked by M70 and previously referred to as K2. While there were four phylogenetically equivalent mutations on the lineage leading to this clade (M70, M184, M193, and M272), there were only two mutations (M320 and P77) marking its internal branches (i.e., T-M320, T-P77, and T*). Here, we identify two additional mutations on the branch leading to the T clade (L206 and PS129) and map 10 SNPs (PS78, PS2, PS21, L131, P330, P321, P317, P322, P328, and P327) that define eight new branches within haplogroup T (Figure 1). Of these 12 new mutations, five were previously ascertained and seven were SNPs discovered in the course of this study. We found a single individual from Syria that carries the derived allele at M184, M272, M193, PS129, and L206 and the ancestral allele at M70 and PS78. This splits the long branch leading to haplogroup T and gives rise to a novel T* lineage.

Within the T clade, the new mutations PS2 and PS21 define T1a, which accounts for two thirds of our sampled T chromosomes. Three novel mutations fall within this subclade (P330, P321, and P317), and in combination with the two previously known mutations (P77 and M320) mark a total of six T1a lineages (Figure 1). None of our samples is positive for M320, which was originally reported in a single Druze (Shen et al. 2004) (all our Druze T chromosomes are PS21*). Thus, we assume that M320 defines a rare lineage within T1a. Most of the remaining T chromosomes in our survey belong to T1b, which is defined by the newly mapped L131. Three additional mutations fall in T1b (P322, P328, P327), marking the T1b1* and T1b1a lineages within this subclade.

During the course of this study, we discovered a novel SNP (P326) that marks a higher-level branch that unites the haplogroups T and L (Figure 1). This major revision of the Y chromosome phylogeny subdivides haplogroup K into two main clades, one containing haplogroups L and T, and the other containing all the remaining haplogroups downstream of K (i.e., M, N, O, P, Q, R, and S).

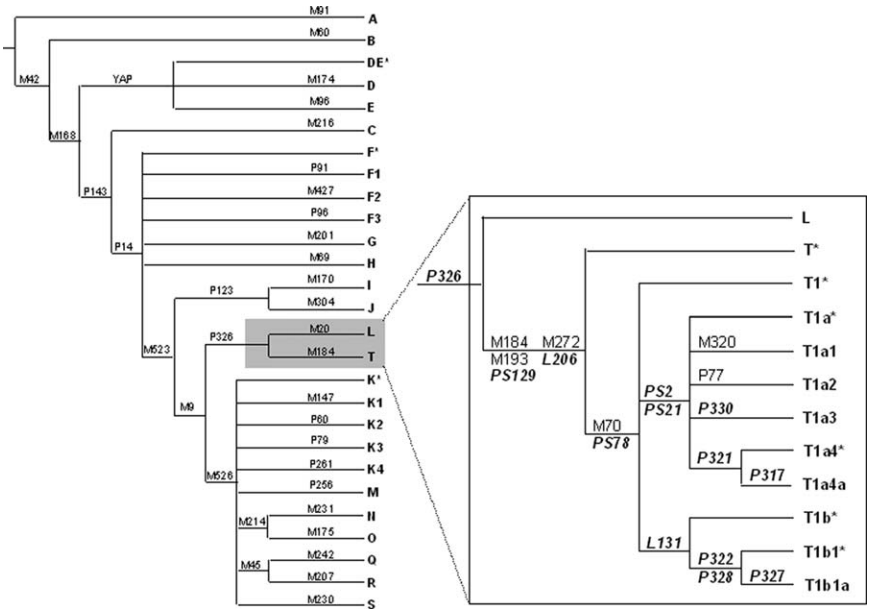

Figure 1. Y chromosome phylogenetic tree. A framework for 27 major lineages of the Y chromosome tree is shown on the left; a detailed tree for haplogroup T and its link to L is shown on the right. Haplogroup names are indicated to the right of the branches, while mutations are indicated along the branches of the tree. Mutations mapped to the tree for the first time are indicated by bold italics.

Population Demography and TMRCA Estimates. Using BATWING in a model of constant population size followed by exponential growth, we obtained estimates for model parameters. The 95% confidence interval of the growth rate is between 0.001 and 0.018 per generation (median of 0.005), with values that are roughly consistent with the inferred growth rate during the Neolithic (Carneiro and Hilse 1966). The time of the onset of growth has a median value of 21.9 Ky, predating the beginning of the Holocene. The effective population size during the constant phase has a median value of 5000 males.

We also estimated the TMRCA of different branches of the tree (T, T1, T1a, T1a2, and T1b) using a parametric (BATWING) and a nonparametric (Td) method. Based on Td and 17 Y-STRs, the estimates for the TMRCAs of the haplogroups defined by M70, PS21, L131, and P77 are 16.9 Ky, 14.0 Ky, 10.9 Ky, and 7.6 Ky, respectively (Table 2). We did not estimate the TMRCA of the entire T haplogroup using this method because the asymmetry caused by the presence of only a single M184* chromosome would likely lead to an underestimation of the age of M184. Likewise, Lemba were excluded for the estimation of the TMRCA of the haplogroup defined by L131 because their chromosomes are very similar and likely to bias the inference of the ancestral haplotype for this branch.

Table 2. TMRCA Estimates in Ky for Haplogroups T-M184, T-M70, T-S21, T-L131, and T-P77

					TMRCA K = 47.4 Ky			*TMRCA K = 48.1 Ky*		
Hg	n	*T*[b]	n	*BATWING*[c,d]	*STRs (BATWING)*[e,f]	*SNPs*[e,f]	*SNPs+STRs (BATWING)*[e,f]	*STRs (BATWING)*[e,f]	*SNPs*[e,f]	*SNPs+STRs (BATWING)*[e,f]
T-M184	NA	NA	63	29.2 (7.5–95.2)	26.1 (17.8–35.6)	37.9 (17.8–46.2)	30.8 (22.5–40.3)	26.5 (18.0–36.1)	38.5 (18.0–46.9)	31.3 (22.8–40.9)
T-M70	89	16.9 (4.8)	62	24.6 (6.3–78.3)	23.7 (16.6–30.8)	28.4 (4.7–36.7)	23.7 (19.0–33.2)	24.1 (16.8–31.3)	28.9 (4.8–37.3)	24.1 (19.2–33.7)
T-PS21	65	14.0 (3.7)	49	21.2 (5.5–67.5)	21.3 (15.4–27.3)	0.0 (0.0–14.2)	19.0 (14.2–26.1)	21.6 (15.6–27.7)	0.0 (0.0–14.4)	19.2 (14.4–26.5)
T-L131[a]	14	10.9 (3.1)	10	13.8 (3.5–45.6)	NA	NA	NA	NA	NA	NA
T-P77	21	7.6 (2.7)	17	12.8 (3.3–42.6)	NA	NA	NA	NA	NA	NA

a. Excluding Lemba.

b. TMRCA estimation based on STR variation using the Td method (Zhivotovsky 2004) with 17 STRs. Only samples successfully genotyped for 24 Y-STRs were used in this analysis.

c. All analyses on BATWING involve only samples from Jews, Near Easterners, and North Africans.

d. Median (confidence interval) from the posterior distribution.

e. Maximum likelihood estimates (confidence intervals) assuming TMRCA of K is known.

f. Joint likelihood taken on M184, M70, and S21. Only SNPs found in Rozen et al. (2010) were used in this analysis.

BATWING analysis, based on variation at 10 Y-STRs in 12 Near Eastern, one North African, one European, and nine Jewish populations (Table 1), yielded median TMRCA values of ~48.1 Ky and ~144.1 Ky for haplogroup K and for all human Y chromosomes, respectively. The former estimate is very close to the estimate of Karafet et al. (2008), which was based on an arbitrary choice of a calibration point (see below). The latter estimate is somewhat higher than previous estimates that assumed a slightly higher STR mutation rate (Pritchard et al. 1999) but is consistent with previous estimates based on SNPs (Wilder et al. 2004). When the relative values of the TMRCA of PS21 and L131 are compared, L131 is usually younger (95% confidence interval for $TMRCA_{L131}/TMRCA_{PS21}$ is [0.42–0.95]). On the other hand, the TMRCAs of L131 (13.8 Ky) and of P77 (12.8 Ky) are very similar and not statistically different (Table 2).

To combine information coming from Y-STRs and Y-SNPs, we took the joint posterior distribution from BATWING and converted it into a joint likelihood for the relative TMRCA values of the internal nodes to that of haplogroup K (Table 2). We then combined it with the joint likelihood obtained from the distribution of SNPs in the genealogy (Supplemental Methods) to achieve improved confidence intervals. This method requires as a calibration point the TMRCA of haplogroup K, for which we used two different values: 47.4 Ky (Karafet et al. 2008) and 48.1 Ky (this study). These two values differ in how they were obtained, the former coming from the distribution of SNPs in the Y chromosome tree conditioned on the age of a node (which was associated with an initial "out of Africa" dispersal ~70 kya) (Karafet et al. 2008), and the latter obtained assuming the mutation rate for STRs is known. The median values of the TMRCAs of M184, M70, and PS21 are 30.8 Ky, 23.7 Ky, and 19.0 Ky, respectively, when we use 47.4 Ky for haplogroup K, and 31.3 Ky, 24.1 Ky, and 19.2 Ky, respectively, when 48.1 Ky is used instead.

We also estimated the TMRCA of the T and L haplogroups, which share the mutation P326, using the joint likelihood method in BATWING based on Y-STRs, Y-SNPs, or both (Table 3), including also M184 and M70 in the analysis. The corresponding median values for P326, M184, and M70 are 42.7 Ky, 26.1 Ky, and 21.3 Ky when 47.4Ky is used, and 43.3 Ky, 26.5 Ky, and 21.6 Ky when 48.1 Ky is used. The analysis including P326 yields slightly younger ages for M184 and M70.

Origin and Dispersal of Haplogroup T in the Near East. Three lines of evidence support the hypothesis that haplogroup T originated in the Near East and subsequently expanded from there. First, the geographic distribution of the two sister clades, haplogroup L and haplogroup T, overlap in the Near East, although L has a more easterly epicenter in India and Pakistan (Sengupta et al. 2006). Second, almost all of M70* chromosomes surveyed here are found in the Near East, and the two main subclades (T1a and T1b) also predominate in this area. Finally, the internal structure of the T clade, with the single T* sample coming from Syria, provides evidence that the most basal haplogroup T branch is present in the

Table 3. TMRCA Estimates in Ky for Haplogroups LT, T-M184, and T-M70

	TMRCA K = 47.4 Ky			*TMRCA K = 48.1 Ky*		
Hg	*STRs (BATWING)*[a,b]	*SNPs*[a,b]	*SNPs + STRs (BATWING)*[a,b]	*STRs (BATWING)*[a,b]	*SNPs*[a,b]	*SNPs + STRs (BATWING)*[a,b]
LT-P326	36.7 (26.1–46.2)	41.5 (28.4–46.2)	42.7 (34.4–46.2)	37.3 (26.5–46.9)	42.1 (28.9–46.9)	43.3 (34.9–46.9)
T-M184	26.1 (19.0–40.3)	15.4 (4.7–30.8)	26.1 (19.0–33.2)	26.5 (19.2–40.9)	15.6 (4.8–31.3)	26.5 (19.2–33.7)
T-M70	23.7 (17.8–34.4)	4.7 (1.2–19.3)	21.3 (15.4–27.3)	24.1 (18.0–34.9)	4.8 (1.2–19.2)	21.6 (15.6–27.7)

a. Maximum likelihood estimates (confidence intervals) assuming TMRCA of K is known.
b. Joint likelihood taken on P326, M184, and M70. All known SNPs common to all M70 chromosomes were used.

Near East. To our knowledge this is the first study where both M70 and M184 (or M272) were surveyed in the same set of samples. However, several studies where M70 was typed leave no room for T*, because no K* chromosomes were reported (Cinnioglu et al. 2004; Luis et al. 2004; Semino et al. 2002).

Interestingly, Kurdish and Iraqi Jews (followed by Iranian and Yemenite Jews) have the highest frequency of the otherwise rare P77 mutation (Table 1). The former represent some of the oldest established populations of Jews, tracing their origins to the Assyrian and Babylonian exiles from the kingdoms of Israel and Judah, respectively (Levi 1999; Rejwan 1985). This suggests that P77 may have diversified in Northern Mesopotamia and expanded from there, or alternatively that it originated in Canaan and was carried to Mesopotamia during the Babylonian exile (e.g., reaching Iraqi and Iranian non-Jewish population through the assimilation of the former exiles). Chromosomes not belonging to the branches PS21 or L131 (M70*) are also found at the highest frequency in Iraqi and Iranian Jews but explain a very small fraction of chromosomes from populations located to the West.

Origins of European Haplogroup T. Using the program NETWORK with 24 Y-STRs (and including Y-SNPs) we reconstructed the relationships among 90 haplogroup T chromosomes (Figure 2). This network was constructed with the set of samples used in the nonparametric estimation of TMRCAs and with the only T-M184* sample. This set includes samples from Europeans, African, Jews, and Near Easterners, but only if they were successfully typed for 24 Y-STRs. The median-joining network reveals high levels of differentiation among haplotypes, the absence of haplotype sharing among populations, and very little geographic structure. The lack of a star-like structure suggests that this diverse set of T chromosomes did not expand recently from a single source population. Near Eastern and Jewish T chromosomes are widely distributed across the network, while European haplotypes are found associated both with Jewish and non-Jewish haplotypes. The occurrence in Europe of lineages belonging to both T1a and T1b subclades probably reflects multiple episodes of gene flow. For example, we sampled an Italian T1a-P77 haplotype that is highly derived and closely related to Jewish Y chromosomes and several other European haplotypes that appear to be more basal in the T genealogy (i.e., T1a* and T1b*) and closely related to Near Eastern non-Jewish chromosomes (i.e., Assyrians and Jordanians). This supports the hypothesis that haplogroup T chromosomes entered Europe at different time points, perhaps associated both with the diffusion of agriculture and the more recent dispersal of Jewish groups to Europe. The small sample size of European T chromosomes does not permit an accurate estimate of the timing of male gene flow to Europe; high-resolution analysis of other haplogroups associated with the spread of agriculture (Balaresque et al. 2010) or Jewish populations (Behar et al. 2004; Hammer et al. 2009) may provide better insight into this question.

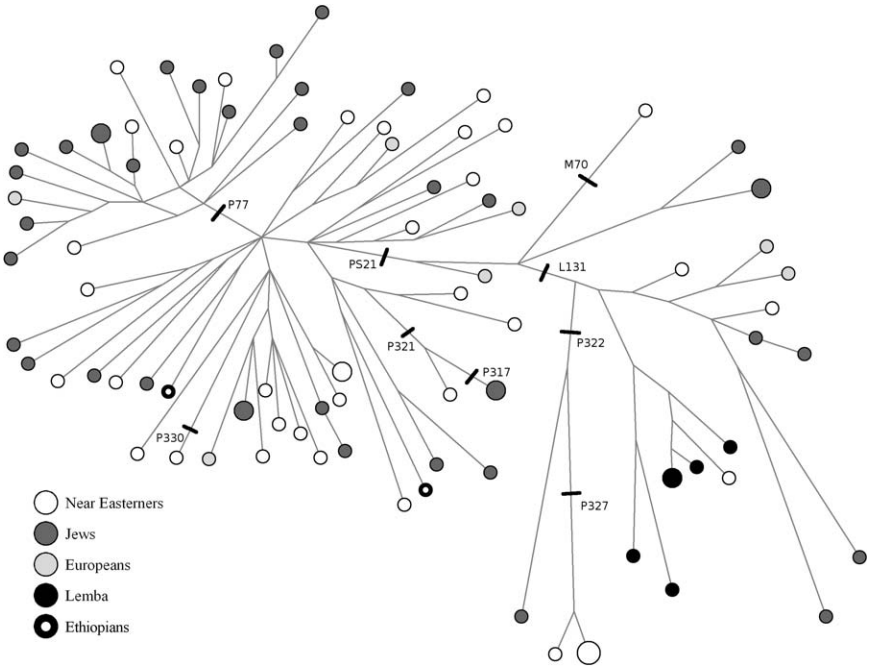

Figure 2. Median-joining haplotype network T chromosomes based on 24 Y-STRs (see text). The position of 9 Y-SNPs is denoted by cross-hatches. Y-STR haplotypes are represented by circles with an area that is proportional to the number of individuals with that haplotype. Branch lengths are proportional to the number of one-repeat mutations separating each haplotype. In this Network Egyptians are grouped within Near Easterners.

Previous analysis of the descendants of Thomas Jefferson revealed that Jefferson's Y chromosome carries the M70 mutation (King et al. 2007). A phylogenetic network based on eight Y-STR loci showed that Jefferson's haplotype was most closely related to an Egyptian haplotype. However, the observation that two of 85 unrelated British men with the Jefferson surname shared the same haplogroup with the descendants of Thomas Jefferson was considered more consistent with a scenario in which Jefferson's Y chromosome was part of a very rare European lineage. To further elucidate the paternal ancestry of Jefferson's Y chromosome we compared the Virginia-Jefferson 12-locus Y-STR haplotype with our data. No matches were found when the following loci were used: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, and DYS460. All closely-related haplotypes ($n = 18$, three to six steps removed from Jefferson's haplotype) carried the derived state at PS21 and the ancestral states at P77, P317, P321, and P330. These eighteen haplotypes were found among Egyptians ($n = 5$), Jews from different countries ($n = 7$), Europeans ($n = 2$), Turks ($n = 1$), Syrians ($n = 1$), Jordanians ($n = 1$), and a single African American (data not shown). We infer that Jefferson's Y chromosome belongs to T1a*. Similar to a

previous analysis based on eight Y-STR loci (King et al. 2007), our network links Jefferson's haplotype to two Egyptian haplotypes. However, a single German haplotype is also found in the same cluster, and as mentioned above, T1a* haplogroups in Europe likely reflect older gene flow. The affiliation of the Jefferson haplotype to T1a* and the absence of closely related haplotypes (zero to two step mutations away) in the network supports the hypothesis that this haplotype belongs to an ancient rare European Y-chromosome lineage rather than to lineages that recently migrated to Europe from the Near East.

Lemba T Chromosomes: A Jewish Link? What do our data on haplogroup T tell us about the origins of the Lemba, a southern African Bantu-speaking population claiming Jewish ancestry? Previous Y chromosome studies pointed to a possible Semitic origin for approximately 50% of Lemba Y chromosomes (Spurdle and Jenkins 1996), whereas mtDNA data showed no appreciable differentiation with neighboring Bantu-speaking populations (Soodyall 1993). This pattern was consistent with Lemba oral tradition suggesting an origin in the Near East, with subsequent intermarriage with local women during dispersal of the Lemba to their present-day locations in southern Africa. The presence of the Cohen Modal Haplotype at high frequency in a subset of Lemba men further supported the claim of Jewish ancestry (Thomas et al. 2000).

In our sample of 34 Lemba, we found six chromosomes (17.6%) that belong to T1b-L131* (Figure 2). Chromosomes in this cluster share a recent common ancestor at about 5.5 Ky based on Td, which may reflect a recent founder effect. This result is similar to the estimate of an expansion time of 3.3 kya based on a sample of T chromosomes found at a frequency of 10% in a Somalia (Sanchez et al. 2005). The hypothesis that contemporary Lemba trace to an ancient Jewish population predicts that their Y chromosomes cluster with Jewish haplotypes. However, Jewish and Lemba T chromosomes tend to fall into different subclades (T1a and T1b, respectively), and STR data show that the closest relationship of Lemba T chromosomes is with a Turk (Figure 2). Of course, it is possible that Y chromosomal lineages that became prevalent in Lemba went extinct in current Jewish populations, or are at low frequency and have not been sampled.

Conclusions

We provide a more resolved phylogeny of Y chromosome haplogroup T, which permits finer-scale resolution in phylogeographic analysis. Our results support the hypothesis that haplogroup T began to diversify in the Near East beginning ~25 kya. We posit that the current distribution of haplogroup T chromosomes was influenced by multiple demographic processes, including the spread of agriculture from the Near East, the Assyrian and Babylonian exiles, and later dispersals associated with the Jewish diaspora. Finally, our results suggest that a separate migration process involving the ancestors of the Lemba led to the

dispersal of haplogroup T from the Near East to South Africa. However, despite oral tradition, we do not find a stronger link between Lemba and Jewish Y chromosome than between Lemba and non-Jewish Y chromosomes from the Near East.

Acknowledgments We thank Sterling Udom and Daniel Teberg for technical assistance. This work was supported in part by the Salus Mundi Foundation to T.M.K.

Received 20 July 2010; revision accepted for publication 17 September 2010.

Literature Cited

Balaresque, P., G. R. Bowden, S. M. Adams et al. 2010. A predominantly neolithic origin for European paternal lineages. *PLoS Biol.* 8:e1000285.

Bandelt, H. J., P. Forster, and A. Rohl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.

Behar, D. M., D. Garrigan, M. E. Kaplan et al. 2004. Contrasting patterns of Y chromosome variation in Ashkenazi Jewish and host non-Jewish European populations. *Hum. Genet.* 114:354–365.

Carneiro, R. L., and D. F. Hilse. 1966. Determining probable rate of population growth during Neolithic. *Am. Anthropol.* 68:177–181.

Chiaroni, J., P. A. Underhill, and L. L. Cavalli-Sforza. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. *Proc. Natl. Acad. Sci. USA.* 106:20174–20179.

Cinnioglu, C., R. King, T. Kivisild et al. 2004. Excavating Y-chromosome haplotype strata in Anatolia. *Hum. Genet.* 114:127–148.

Foster, E. A., M. A. Jobling, P. G. Taylor et al. 1998. Jefferson fathered slave's last child. *Nature.* 396:27–28.

Hammer, M. F., D. M. Behar, T. M. Karafet et al. 2009. Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum. Genet.* 126:707–717.

Karafet, T. M., J. S. Lansing, A. J. Redd et al. 2005. Balinese Y-chromosome perspective on the peopling of Indonesia: Genetic contributions from pre-neolithic hunter-gatherers, Austronesian farmers, and Indian traders. *Hum. Biol.* 77:93–114.

Karafet, T. M., F. L. Mendez, M. B. Meilerman et al. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18:830–838.

Kimura, M., and T. Ohta. 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci. USA.* 75:2868–2872.

King, T. E., G. R. Bowden, P. L. Balaresque et al. 2007. Thomas Jefferson's Y chromosome belongs to a rare European lineage. *Am. J. Phys. Anthropol.* 132:584–589.

Levi, H. 1999. *Comprehensive History of the Jews of Iran: The Outset of the Diaspora.* Costa Mesa, CA: Mazda Publishers.

Luis, J. R., D. J. Rowold, M. Regueiro et al. 2004. The Levant versus the Horn of Africa: Evidence for bidirectional corridors of human migrations. *Am. J. Hum. Genet.* 74:532–544.

Nogueiro, I., L. Manco, V. Gomes et al. 2010. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. *Am. J. Phys. Anthropol.* 141:373–381.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun et al. 1999. Population growth of human Y chromosomes: A study of Y chromosome microsatellites. *Mol. Biol. Evol.* 16:1791–1798.

Redd, A. J., A. B. Agellon, V. A. Kearney et al. 2002. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int.* 130:97–111.

Redd, A. J., V. F. Chamberlain, V. F. Kearney et al. 2006. Genetic structure among 38 populations from the United States based on 11 U.S. core Y chromosome STRs. *J. Forensic. Sci.* 51:580–585.

Rejwan, N. 1985. *The Jews of Iraq: 3000 Years of History and Culture.* Boulder, CO: Westview Press.

Repping, S., S. K. van Daalen, L. G. Brown et al. 2006. High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat. Genet.* 38:463–467.

Rozen, S., J. D. Marszalek, R. K. Alagappan et al. 2009. Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. *Am. J. Hum. Genet.* 85:923–928.

Sanchez, J. J., C. Hallenberg, C. Borsting et al. 2005. High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males. *Eur. J. Hum. Genet.* 13:856–866.

Semino, O., A. S. Santachiara-Benerecetti, F. Falaschi et al. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* 70:265–268.

Sengupta, S., L. A. Zhivotovsky, R. King et al. 2006. Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists. *Am. J. Hum. Genet.* 78:202–221.

Shen, P., T. Lavi, T. Kivisild et al. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-chromosome and mitochondrial DNA sequence variation. *Hum. Mut.* 24:248–260.

Soodyall, H. 1993. *Mitochondrial DNA Polymorphisms in Southern African Populations*. Johannesburg, South Africa: University of the Witwatersrand.

Spurdle, A. B., and T. Jenkins. 1996. The origins of the Lemba "Black Jews" of southern Africa: Evidence from p12F2 and other Y-chromosome markers. *Am. J. Hum. Genet.* 59:1126–1133.

Thomas, M. G., T. Parfitt, D. A. Weiss et al. 2000. Y chromosomes traveling south: The Cohen modal haplotype and the origins of the Lemba–the "Black Jews of Southern Africa." *Am. J. Hum. Genet.* 66:674–686.

Wilder, J. A., Z. Mobasher, and M. F. Hammer. 2004. Genetic evidence for unequal effective population sizes of human females and males. *Mol. Biol. Evol.* 21:2047–2057.

Wilson, I. J., M. E. Weale, and D. J. Balding. 2003. Inferences from DNA data: Population histories, evolutionary processes and forensic match probabilities. *J. R. Stat. Soc.: Series A (Statistics in Society).* 166:155–188.

YCC. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12:339–348.

Zhivotovsky, L. A., P. A. Underhill, C. Cinnioglu et al. 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.