

Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes

M. F. Hammer^{*†‡}, A. J. Redd^{*†}, E. T. Wood^{*†}, M. R. Bonner^{*}, H. Jarjanazi^{*}, T. Karafet^{*}, S. Santachiara-Benerecetti[¶], A. Oppenheim^{||}, M. A. Jobling^{**}, T. Jenkins^{††}, H. Ostrer^{‡‡}, and B. Bonn -Tamir[§]

^{*}Laboratory of Molecular Systematics and Evolution, University of Arizona, Tucson, AZ 85721; [¶]Department of Genetics, Universit  degli Studi di Pavia, Pavia 27100, Italy; ^{||}Hadassah Medical School, Hebrew University of Jerusalem, Jerusalem 91120, Israel; ^{**}Department of Genetics, University of Leicester, Leicester LE1 7RH, England; ^{††}SAMIR, University of Witwatersrand, Johannesburg 2000, South Africa; ^{‡‡}Department of Pediatrics, New York University Medical Center, New York, NY 10016; and [§]Department of Human Genetics, Sackler School of Medicine, Ramat Aviv 69978, Israel

Communicated by Arno G. Motulsky, University of Washington, Seattle, WA, March 15, 2000 (received for review November 17, 1999)

Haplotypes constructed from Y-chromosome markers were used to trace the paternal origins of the Jewish Diaspora. A set of 18 biallelic polymorphisms was genotyped in 1,371 males from 29 populations, including 7 Jewish (Ashkenazi, Roman, North African, Kurdish, Near Eastern, Yemenite, and Ethiopian) and 16 non-Jewish groups from similar geographic locations. The Jewish populations were characterized by a diverse set of 13 haplotypes that were also present in non-Jewish populations from Africa, Asia, and Europe. A series of analyses was performed to address whether modern Jewish Y-chromosome diversity derives mainly from a common Middle Eastern source population or from admixture with neighboring non-Jewish populations during and after the Diaspora. Despite their long-term residence in different countries and isolation from one another, most Jewish populations were not significantly different from one another at the genetic level. Admixture estimates suggested low levels of European Y-chromosome gene flow into Ashkenazi and Roman Jewish communities. A multidimensional scaling plot placed six of the seven Jewish populations in a relatively tight cluster that was interspersed with Middle Eastern non-Jewish populations, including Palestinians and Syrians. Pairwise differentiation tests further indicated that these Jewish and Middle Eastern non-Jewish populations were not statistically different. The results support the hypothesis that the paternal gene pools of Jewish communities from Europe, North Africa, and the Middle East descended from a common Middle Eastern ancestral population, and suggest that most Jewish communities have remained relatively isolated from neighboring non-Jewish communities during and after the Diaspora.

Jewish religion and culture can be traced back to Semitic tribes that lived in the Middle East approximately 4,000 years ago. The Babylonian exile in 586 B.C. marked the beginning of major dispersals of Jewish populations from the Middle East and the development of various Jewish communities outside of present-day Israel (1). Today, Jews belong to several communities that can be classified according to the location where each community developed. Among others, these include the Middle Eastern communities of former Babylonia and Palestine, the Jewish communities of North Africa and the Mediterranean Basin, and Ashkenazi communities of central and eastern Europe. The history of the Jewish Diaspora—the numerous migrations of Jewish populations and their subsequent residence in various countries in Europe, North Africa, and West Asia—has resulted in a complex set of genetic relationships among Jewish populations and their non-Jewish neighbors. Several studies have attempted to describe these genetic relationships and to unravel the numerous evolutionary factors that have come into play during the Diaspora (2–11). Some of the key arguments in the literature concern the relative contributions of common ancestry, genetic drift, natural selection, and admixture leading to the observed similarities and differences among Jewish and non-Jewish communities.

Given the complex history of migration, can Jews be traced to a single Middle Eastern ancestry, or are present-day Jewish communities more closely related to non-Jewish populations from the same geographic area? Some genetic studies suggest that Jewish populations show substantial non-Jewish admixture and the occurrence of mass conversion of non-Jews to Judaism (2, 3, 10, 12). In contrast, other research points to considerably greater genetic similarity among Jewish communities with only slight gene flow from their respective host populations (5, 7, 9, 11, 13). Furthermore, it has been demonstrated that the degree of genetic similarity among Jewish communities and between Jewish and non-Jewish populations depends on the particular locus that is being investigated (4, 8, 11). This observation raises the possibility that variation associated with a given locus has been influenced by natural selection.

All of the aforementioned investigations used “classical” genetic markers such as blood groups, enzymes, and serum proteins, as well as immunoglobulins and the HLA system. More recently, restriction fragment length polymorphism studies were initiated by using mitochondrial DNA (mtDNA), the nonrecombining portion of the Y chromosome (NRY), and other nuclear loci (14–20). An advantage of nucleotide-level studies is that they circumvent some of the complications associated with selection; however, these studies have not fully resolved many of the key issues in the earlier literature.

Analyses of mtDNA and the NRY are especially relevant to studies of Jewish origins because, according to ancient Jewish law, Jewish religious affiliation is assigned maternally (1). In particular, studies of paternally inherited variation provide the opportunity to assess the genetic contribution of non-Jewish males to present-day Jewish genetic diversity. This research represents one of the first comparisons of biallelic variation on the NRY in Jewish and non-Jewish populations from similar geographic areas. We surveyed 18 biallelic polymorphisms in 7 Jewish and 22 non-Jewish populations from Europe, the Middle East, and Africa to assess the relative contributions of common ancestry, gene flow, and genetic drift in shaping patterns of NRY variation in populations of the Jewish Diaspora.

Abbreviations: mtDNA, mitochondrial DNA; NRY, nonrecombining portion of the Y chromosome; MDS, multidimensional scaling; AMOVA, analyses of molecular variance.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF257063 and AF257064).

[†]M.F.H., A.J.R., and E.T.W. contributed equally to this work.

[‡]To whom reprint requests should be addressed at: Laboratory of Molecular Systematics and Evolution, Biosciences West Room 239, University of Arizona, Tucson, AZ 85721. E-mail: mhammer@u.arizona.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Article published online before print: *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.100115997. Article and publication date are at www.pnas.org/cgi/doi/10.1073/pnas.100115997

Table 1. Y-chromosome haplotype frequencies, %, in 29 populations

Population	Haplotype									
	N	4S	1R	Med	1Ha	1U	1C	1D	1L	Other
Jews	336	20	5	36	5	8	7	5	10	4
Ashkenazi	115	24	12	31	2	3	4	9	10	3
Roman	44	16		45	5	7		5	20	2
North African	45	20	2	42	18	7	4	4	2	
Near Eastern	32	13	3	28	6	16	31		3	
Kurdish	50	8	4	44	6	18	8	4	8	
Yemenite	30	17		43		7	10	3	17	3
Ethiopian	20	45		10		5				40
Mid-East non-Jews	230	15	5	50	4	10	2	6	6	3
Palestinians	73	19	5	51	3	7	1		8	5
Syrians	91	10	3	57	3	8	1	9	9	
Lebanese	24	29	13	46		4		4		4
Druze	21	19		38	19	19	5			
Saudi Arabians	21	5	5	33	5	24	5	19		5
Europeans*	337	15	16	11	6	2	1	11	37	1
North Africans*	137	50	1	26		4	1	1	4	12
Sub-Saharan Africans*	199	6		1		1				93
Other										
Turks	98	6	12	26	9	12	3	5	20	6
Lemba	34	6	6	26		32				29

Only haplotypes common in Jewish populations are shown.

*See *Subjects and Methods* for list of populations in each group.

Subjects and Methods

DNA Samples. We analyzed a total of 1,371 males from 29 populations. These populations were categorized into five major divisions based on a combination of geographic, religious, linguistic, and ethnohistorical criteria: Jews, Middle Eastern non-Jews, Europeans, North Africans, and sub-Saharan Africans (Table 1). The Jewish samples included 115 Ashkenazim (Ash), 44 Roman Jews (Rom) (21), 45 North African Jews (Naf) (25 Moroccans, 15 Libyans, 1 Tunisian, 1 Algerian, and 3 from unspecified North African countries), 32 Near Eastern Jews (Nea) (18 Iraqis and 14 Iranians), 50 Kurdish Jews (Kur) (22), 30 Yemenite Jews (Yem) (23), and 20 Ethiopian Jews (EtJ) (23). The non-Jewish Middle Eastern samples included 73 Palestinians (Pal), 91 Syrians (Syr), 23 Lebanese (Leb), 21 Israeli Druze (Dru), and 21 Saudi Arabians (Sar). The remaining sample composition was as follows: Europeans: 31 Russians (Rus), 44 British (Bri), 33 Germans (Ger), 40 Austrians (Aus), 81 Italians (Ita), 23 Spanish (Spa), 85 Greeks (Gre); North Africans: 31 Tunisians (Tun), 58 Egyptians (Egy), 48 Ethiopians (Eth); sub-Saharan Africans: 49 Gambians (Gam), 31 Biaka (Bia), 26 Bagandans (Bag), 63 San (San), and 30 Zulu (Zul). We also analyzed a sample of 98 Turks (Tur) and 34 unrelated males from the Lemba (Lem), a Bantu (Venda)-speaking population from southern Africa who claim Jewish paternal ancestry (24). All sampling protocols were approved by the Human Subjects Committee at the University of Arizona.

Mutation Detection. Mutation detection analysis was performed by using single-stranded conformation polymorphism (SSCP) (25) and denaturing high-performance liquid chromatography (DHPLC) (26). The *DYS211* (GenBank accession no. G11997) and *DYS221* (GenBank accession no. G12000) sequence-tagged sites were amplified by using the conditions and primers reported by Vollrath *et al.* (27). The Y-specific clones 4-1 (*DYS188*) and 3-8 (*DYS194*) (28) were sequenced by primer walking (GenBank accession nos. AF257063 and AF257064), and the sequence information was used to design primers to amplify shorter fragments for DHPLC analysis. DNA sequencing was performed by standard procedures to identify mutations that caused altered

electrophoretic mobility on SSCP gels or altered DHPLC chromatograms.

Allele-Specific Genotyping Assays. Variation at the *DYS188*₇₉₂ and *DYS194*₄₆₉ polymorphic sites was genotyped by using allele-specific PCR (29). Genotyping of the *DYS221*₁₃₆ C→T transition and *DYS211*₁₀₅ A→T transversion was performed by site-specific oligonucleotide hybridization (30). The PCR conditions and primer sequences used in these allele-specific genotyping assays were deposited in the National Center for Biotechnology Information dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP>). The p12f2 polymorphism at *DYS11* was genotyped by using the PCR assay developed by M. Shlumukova, M. E. Hurler, and M.A.J. (unpublished results). *DYS287* (YAP) was scored according to the method of Hammer and Horai (31). All other single-nucleotide polymorphisms and variation in the length of the polydeoxyadenylate tract (polyA tail) at the 3' end of the YAP element were genotyped according to methods reported by Hammer *et al.* (32) and Karafet *et al.* (33).

Statistical Analyses. Genetic distances based on haplotype frequencies were estimated by using the PHYLIP package (34). We also used ARLEQUIN (35) to test the hypothesis of a random distribution of haplotypes among population groups and to perform analyses of molecular variance (AMOVA). AMOVA produces estimates of variance components and F-statistic analogs (Φ -statistics) reflecting the correlation of haplotypic diversity at different levels of hierarchical subdivision (36). Multidimensional scaling (MDS) (37) was performed with maximum likelihood estimation by using the program MULTISCALE (Ramsay, <ftp://ego.psych.mcgill.ca/pub/ramsay/multiscl>). Genetic distances among populations were regressed on geographic distances for all pairs of populations. Because these data points were not independent, the significance of each regression was assessed by a Mantel test (38). The mean and standard deviation of the time to the most recent common ancestral NRY sequence as well as the ages of each of the mutations in our cladogram were estimated by using the program GENETREE (39), as described in Hammer *et al.* (32). Simulations were based on five million replicate runs and θ values of 3.0, 3.5, 4.0, and 4.5, where θ

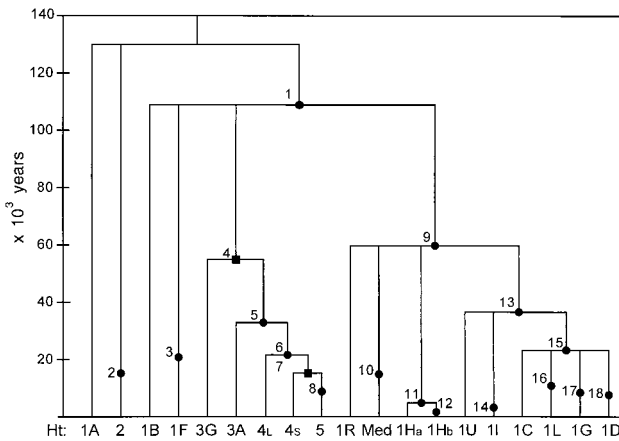


Fig. 1. Time-scaled gene tree for 19 biallelic Y-chromosome haplotypes. Each mutational event at one of 18 biallelic sites on the nonrecombining portion of the Y chromosome is represented by a dot (single-nucleotide polymorphism) or a square (small insertion/deletion). Mutational ages (in 10^3 yr) are indicated on the y axis, and the haplotypes defined by these mutational events are shown on the x axis. 1 = SRY_{10831.1}, 2 = PN3, 3 = RPS4Y₇₁₁, 4 = YAP, 5 = SRY₄₀₆₄, 6 = PN2, 7 = p(A)L→S, 8 = PN1, 9 = DYS188₇₉₂, 10 = p12f2, 11 = DYS221₁₃₆, 12 = DYS211₁₀₅, 13 = M9, 14 = Tat, 15 = DYS257, 16 = DYS194₄₆₉, 17 = DYS199, 18 = SRY_{10831.2}.

is an estimate of the average number of polymorphic sites in the length of NRY sequence screened for variation.

By using the computer program ADMIX1_0 (40), we estimated admixture proportions (m) and their standard deviations based on 1,000 bootstrap runs. To infer the Y-chromosome-haplotype frequencies of the Jewish parental population (P1), we averaged the haplotype frequencies of North African, Near Eastern, Yemenite, and Kurdish Jewish samples. To obtain the Y-chromosome-haplotype frequencies of the parental European population (P2), the haplotype frequencies from German, Austrian, and Russian samples were averaged. To estimate P2 for the cases of the Roman Jews and the Lemba, we used the frequencies of Y-chromosome-haplotypes in our Italian sample, and a sample of 86 South African East Bantu speakers (30), respectively. The approach of Shriver *et al.* (41) was followed in selecting the haplotypes exhibiting the highest frequency differential (δ) between the two parental populations for use in the admixture analyses.

Results

Haplotype Tree. During the course of this research, four Y-specific polymorphisms were discovered: a C→T transition at position 792 of DYS188 (DYS188₇₉₂ C→T), a C→A transversion at position 469 of DYS194 (DYS194₄₆₉ C→A), a C→T transition at position 136 of DYS221 (DYS221₁₃₆ C→T) and an A→T transversion at position 105 of DYS211 (DYS211₁₀₅ A→T). Comparisons with the homologous NRY sequences from great apes allowed us to infer the ancestral states at these sites. The character states at all 18 polymorphic sites gave rise to 19 Y-chromosome haplotypes in worldwide populations, 17 of which were present in the 1,371 Y-chromosomes sampled in this survey. Fig. 1 shows the evolutionary relationships among these 19 Y-chromosome haplotypes. We refer to the most basal haplotype defined by the DYS188₇₉₂ C→T mutation as haplotype 1R. Three of the polymorphisms described here mark lineages that are descended from haplotype 1R: the p12f2 8-kb allele (Med haplotype), the DYS221₁₃₆-T allele (haplotype 1Ha), and the DYS211₁₀₅-T allele (haplotype 1Hb). Haplotype 1R is also the ancestor of a set of lineages defined by M9, an ancient C→G transversion polymorphism (26). The DYS194₄₆₉-T allele defines a lineage (haplotype 1L) that is a member of the 1C clade, itself defined by the DYS257₁₈₂ G→A transition (32). In previous haplotype trees (32), YAP⁺ haplotype 4 was differentiated from

haplotype 3A by two mutational events. An intermediate haplotype with the PN2-T allele and the poly(A)-L allele provides evidence that the PN2 C→T transition occurred before the poly(A) L→S deletion. This haplotype, called YAP⁺ haplotype 4L (Fig. 1), was found only in seven Ethiopian males (Jewish and non-Jewish).

Coalescence Analysis. Fig. 1 also presents the mean age estimates for the 18 mutational events in the Y-chromosome-haplotype tree based on a θ value of 3.5. This value of θ produced results that were in excellent agreement with runs based on θ values between 3.0 and 4.5. In particular, the estimated ages of mutations <60,000 years old did not change appreciably as θ values varied (data not shown). The estimated ages of the polymorphisms reported here were $\approx 60,000 \pm 23,000$ years for the DYS188₇₉₂ C→T transition; $22,600 \pm 8,300$ years the poly(A) L→S deletion; $14,800 \pm 9,700$ years for the p12f2 8-kb deletion; $10,000 \pm 5,100$ years for the DYS194₄₆₉ C→A transversion; and $6,500 \pm 7,900$ years for the DYS221₁₃₆ C→T transition.

Geographic Distribution of Y-Chromosome Haplotypes. Of the 13 Y-chromosome haplotypes present in Jewish populations, 8 were found at frequencies of $\geq 5\%$ (Table 1). Nearly all Jewish populations were characterized by relatively high frequencies of two haplotypes (Med and YAP⁺ 4S). Ethiopian Jews were distinguished from other Jewish populations by the presence of moderately high frequencies of haplotypes 1A and 4L (20% each), as well as by the absence of haplotypes 1C, 1D, and 1L.

There was a remarkable similarity in Y-chromosome haplotype composition and average frequency in Jewish and non-Jewish Middle Eastern populations (Table 1). Many of the same haplotypes present in Jewish and Middle Eastern populations were also present in samples from Europe, although at varying frequencies. For example, although the Med haplotype was found at significantly lower frequencies, haplotype 1L was extremely common, especially in northern Europe. Interestingly, the Med and YAP⁺ 4S haplotypes accounted for $\approx 76\%$ of North African Y chromosomes. In contrast, sub-Saharan African populations were characterized by an almost completely different set of haplotypes.

Genetic and Geographic Distances Among Populations. The “among populations” variance component (Φ_{ST}) for the Ashkenazi, Roman, North African, Near Eastern, Kurdish, and Yemenite Jews (the lowest Φ_{ST} value of the five population groups analyzed in Table 2) indicated that these Jewish populations were not significantly different from one another. A series of pairwise differentiation tests in which 13 of 15 Jewish population pairs were not statistically different confirmed this result (data not shown). Furthermore, the mean Jewish interpopulation Chord (42) distance value was lower than that for any other population group (data not shown). It is of particular interest that the level of divergence among Jewish populations was low despite their

Table 2. Φ_{ST} distances within (on diagonal) and between (below diagonal) population groups

	J	MEA	EUR	NAF	SAF
J ⁺	0.013	— ⁺	+	+	+
MEA	0.008	0.033	+	+	+
EUR	0.071	0.125	0.138	+	+
NAF	0.192	0.215	0.285	0.018	+
SAF	0.359	0.388	0.421	0.117	0.281

⁺Only Ashkenazi, North African, Roman, Near Eastern, Yemenite, and Kurdish Jews were included. J, Jews; MEA, Mid-East non-Jews; EUR, Europeans; NAF, North Africans; SAF, sub-Saharan Africans.

⁺Results of population differentiation tests between groups are shown above the diagonal: “—,” not significantly different ($P > 0.03$), and “+,” significantly different ($P \leq 10^{-5}$).

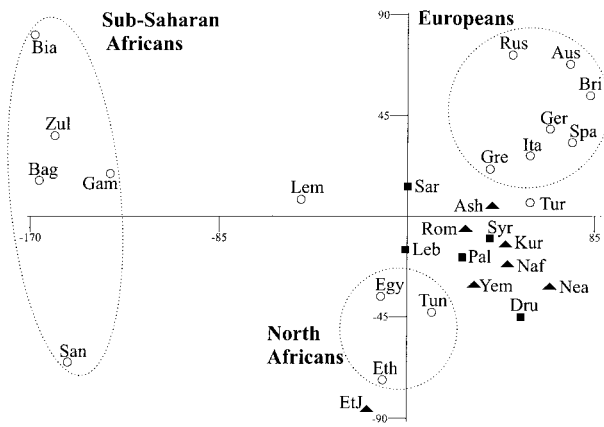


Fig. 2. MDS plot of populations based on Y-chromosome haplotype data. MDS was performed on a matrix of Chord values estimated on the basis of the frequencies of 18 Y-chromosome haplotypes in 29 populations. The three-letter population codes are defined in *Subjects and Methods*. Solid triangles represent Jewish populations, solid squares represent Middle Eastern populations, and open circles represent all other populations.

high degree of geographic dispersion. The mean geographic distance among these six Jewish populations was $\approx 3,000$ km. This value was greater than the mean geographic distances of the Middle Eastern (≈ 600 km) and European ($\approx 1,700$ km) groups and was comparable to that for the North African group ($\approx 2,900$ km). In fact, these Jewish populations had the lowest ratio of genetic-to-geographic distance of all groups in this study.

We then tested for correlations between genetic and geographic distances. There was a linear relationship between the Chord distance matrix and the geographic distance matrix of the non-Jewish populations from Europe, West Asia, and North Africa ($r = 0.620$, $P = 0.0003$). In contrast, the genetic and geographic matrices of the six Jewish populations analyzed above were not significantly correlated ($r = 0.081$, $P = 0.394$). To test whether this difference was the result of lower power caused by the smaller number of Jewish samples compared with non-Jewish samples ($n = 16$), we repeated the test by using six matched non-Jewish populations (Germans, Russians, Tunisians, Palestinians, Saudi Arabians, and Italians), which best represented the geographical locations of our Jewish samples. The correlation between the genetic and geographic distance matrices was still significantly positive ($r = 0.469$, $P = 0.047$).

Genetic Affinities Among Populations. Fig. 2 shows the results of multidimensional scaling based on Chord genetic distances. The correlation between the original Chord distance matrix and a Euclidean distance matrix derived from the two-dimensional plot was very high ($r = 0.971$). Sub-Saharan African, North African, and European populations formed three distinct clusters. The Ashkenazi, Roman, North African, Near Eastern, Kurdish, and Yemenite Jewish populations formed a fairly compact cluster between the North African and European groups. This Jewish cluster was interspersed with the Palestinian and Syrian populations, whereas the other Middle Eastern non-Jewish populations (Saudi Arabians, Lebanese, and Druze) closely surrounded it. Of the Jewish populations in this cluster, the Ashkenazim were closest to South European populations (specifically the Greeks) and also to the Turks. The Ethiopian Jews were placed close to the non-Jewish Ethiopians. The Lemba were located roughly halfway between the sub-Saharan African and Jewish clusters.

The close genetic affinity of Jewish and Middle Eastern non-Jewish populations was confirmed in population differen-

Table 3. Among-group components of variance

Population groupings*	% Variance†	P value
A. Four groups		
1. J/MEA/EUR/NAF	11.9	**
B. Pairwise analysis		
1. J/MEA	0.3	ns
2. J/EUR	5.6	*
3. MEA/EUR	10.4	*
4. J/NAF	18.9	*
5. MEA/NAF	20.8	*
6. EUR/NAF	26.5	**
C. Geographic analysis		
1. MEA/EUR/NAF	9.2	*
2. MEA/EUR	3.3	*
3. MEA/NAF	16.2	*
4. EUR/NAF	15.2	*

*Codes for population groups are the same as in Table 2, except in "Geographic analysis" (see text); †, $\Phi_{CT} = \% \text{ variance}/100$; **, $P < 0.01$; *, $P < 0.05$; ns, $P > 0.05$. ns, no significant difference.

tiation tests (Table 2). Pairwise comparisons between population groups indicated that only 0.8% of the total genetic variance in Jewish and Middle Eastern non-Jewish populations was attributable to between-group differences. This was, by far, the lowest Φ_{ST} value of any of the 10 comparisons in Table 2, and the only value that was not statistically significant.

Population Structure. When AMOVA was performed on populations grouped according to a combination of geographical and religious criteria (e.g., on Jews, Middle Eastern non-Jews, Europeans, and North Africans), 11.9% of the total genetic variance was attributable to differences among groups, 82.5% was attributable to differences within populations, and only 5.6% was partitioned among populations within groups (Table 3, part A). A series of analyses was carried out to identify the extent of among-group variation in pairwise groupings (Table 3, part B). Pairwise comparisons with the North African group tended to produce the highest between-group (Φ_{CT}) values, indicating greater population differentiation between North African and non-African populations. The among-group variance component was statistically significant in all pairwise comparisons except one: only 0.3% ($P = 0.318$) of the total variance was partitioned between Jewish and Middle Eastern non-Jewish groups.

When AMOVA was performed on populations grouped according to a strict geographic criterion (e.g., we categorized Ashkenazim and Roman Jews with Europeans, North African Jews with North Africans, and Near Eastern, Kurdish, and Yemenite Jews with Middle Eastern non-Jews), there was a considerable decrease in the amount of variation partitioned among groups (Table 3, part C). For example, Φ_{CT} values decreased by $\approx 23\%$ in the joint analysis of the Middle Eastern, European, and North African groups (in Table 3, compare part A1 with part C1); and in the pairwise comparisons of Middle Eastern/European, Middle Eastern/North African, and European/North African groups the Φ_{CT} values decreased by 68% (Table 3, part B3 vs. part C2), 22% (part B5 vs. part C3), and 43% (part B6 vs. part C4), respectively. These results indicate that religious affiliation is a better predictor of the genetic affinity among most Jewish populations in our survey than their present-day geographic locations.

Admixture. Table 4 shows the haplotypes with the highest frequency differential between the parental populations and a summary of the admixture estimates for the Ashkenazim, Roman Jews, and the Lemba. Among the Ashkenazim, haplotypes Med and 1L were the most diagnostic for distinguishing the parental Jewish (P1) and

Table 4. Estimated admixture proportions (*m*) and parental haplotype frequency differences (δ)

Population	P1	P2	Diagnostic haplotypes	δ , %	<i>m</i> (diagnostic haplotypes)	<i>m</i> (all haplotypes)
Ashkenazim	*	†	Med	36.3	0.130 ± 0.099	0.227 ± 0.078
			1L	34.3		
Roman Jews	*	‡	1L	30.0	0.289 ± 0.183	0.204 ± 0.206
			Med	24.1		
Lemba	*	§	YAP ⁺ 5	55.9	0.465 ± 0.123	0.356 ± 0.098
			Med	40.1		

*, Non-European Jewish populations; †, Russians, Germans, and Austrians (see text); ‡, Italians; §, South African East Bantu speakers (30).

parental European (P2) population components. All other haplotypes had δ values below 20% (data not shown). The *m* values based on haplotypes Med and 1L were $\approx 13\% \pm 10\%$, suggesting a rather small European contribution to the Ashkenazi paternal gene pool. When all haplotypes were included in the analysis, *m* increased to $23\% \pm 7\%$. This value was similar to the estimated Italian contribution to the Roman Jewish paternal gene pool. Our admixture estimates for the Lemba were consistent with Spurdle and Jenkins' (24) conclusion that $\approx 40\%$ of Lemba Y chromosomes are of sub-Saharan ancestry.

Discussion

Jewish Y-Chromosome Haplotypes. The present research was aimed at comparing the composition of Y-chromosome biallelic haplotypes of Jewish communities with patterns of variation in non-Jews from Africa, the Middle East, and Europe. The Jewish communities surveyed here contained a number of Y-chromosome haplotypes that were shared with non-Jewish populations from a wide geographic region. The Med haplotype, the most frequent haplotype in Jewish communities, was also common in circum-Mediterranean populations. Its widespread distribution and relatively recent age suggest high rates of male gene flow around the Mediterranean and into Europe, possibly via the Neolithic demic diffusion of farmers (43) and/or more recent migrations of sea-going peoples such as the Phoenicians (44).

The second most frequent Jewish haplotype, YAP⁺ haplotype 4, was common in Middle Eastern and southern European populations and reached its highest frequency in North Africa. The discovery of its precursor (YAP⁺ haplotype 4L) in seven Ethiopian males supports the hypothesis that the YAP⁺ haplotype 4S originated on a YAP⁺ 4L chromosome in Ethiopia ($\approx 20,000$ years ago), where it likely increased in frequency before spreading down the Nile River toward Egypt and the Levant (32). This hypothesis is consistent with mtDNA evidence indicating south-to-north gene flow down the Nile (45).

The presence of three haplotypes at very low frequencies (0.3–1.5%) in Jewish and Middle Eastern non-Jewish populations (1A, 3A, and YAP⁺ 5) may be explained by low levels of gene flow from sub-Saharan African populations. This conclusion is consistent with the observed presence of low frequencies of African mtDNA haplotypes in Jewish populations (16). Two haplotypes (1U and 1C) that are common in Asian populations (33) were present at low frequencies in Jewish and Middle Eastern non-Jewish populations (Table 1). Continued surveys of West and Central Asian populations are needed to test the hypothesis of gene flow between Asian and Middle Eastern populations.

Evidence for Common Jewish Origins. Several lines of evidence support the hypothesis that Diaspora Jews from Europe, Northwest Africa, and the Near East resemble each other more closely than they resemble their non-Jewish neighbors. First, six of the seven Jewish populations analyzed here formed a relatively tight cluster in the MDS analysis (Fig. 2). The only exception was the

Ethiopian Jews, who were affiliated more closely with non-Jewish Ethiopians and other North Africans. Our results are consistent with other studies of Ethiopian Jews based on a variety of markers (16, 23, 46). However, as in other studies where Ethiopian Jews exhibited markers that are characteristic of both African and Middle Eastern populations, they had Y-chromosome haplotypes (e.g., haplotypes Med and YAP⁺ 4S) that were common in other Jewish populations.

Second, despite their high degree of geographic dispersion, Jewish populations from Europe, North Africa, and the Near East were less diverged genetically from each other than any other group of populations in this study (Table 2). The statistically significant correlation between genetic and geographic distances in our non-Jewish populations from Europe, the Middle East, and North Africa is suggestive of spatial differentiation, whereas the lack of such a correlation for Jewish populations is more compatible with a model of recent dispersal and subsequent isolation during and after the Diaspora.

To address the degree to which paternal gene flow may have affected the Jewish gene pool, we estimated approximate admixture levels in our Jewish samples from Europe. This question remains unresolved in particular for the Ashkenazi community. Our results indicated a relatively minor contribution of European Y chromosomes to the Ashkenazim. If we assume 80 generations since the founding of the Ashkenazi population, then the rate of admixture would be $<0.5\%$ per generation (47). Interestingly, our total admixture estimate is very similar to Motulsky's (8) average estimate of 12.5% based on 18 classical genetic markers. However, the 18 markers in Motulsky's (8) study fell into two classes: a low admixture class and a high admixture class. Similarly, Cavalli-Sforza and Carmelli (48) found significant heterogeneity of admixture rates for different loci in the Ashkenazim. Because admixture should affect all loci to the same degree, there are two possible explanations for the heterogeneity: (i) admixture levels are actually low, and some loci are affected by convergent selection (e.g., in a common environment), or (ii) admixture levels are actually high, and some loci are experiencing stabilizing selection. Motulsky (8) interpreted the bimodal distribution of admixture values in terms of the former model. Because the NRY has few functional genes and is not likely to have been affected by recent selective sweeps (49, 50), our admixture results support the low admixture model.

Middle Eastern Affinities. A Middle Eastern origin of the Jewish gene pool is generally assumed because of the detailed documentation of Jewish history and religion. There are not many genetic studies that have attempted to infer the genetic relationships among Diaspora Jews and non-Jewish Middle Eastern populations. A number of earlier studies found evidence for Middle Eastern affinities of Jewish genes (4, 5, 7, 51); however, results have depended to a great extent on which loci were being compared, possibly because of the confounding effects of selection (4). Although the NRY tends to behave as a single genetic locus (52), the DNA results presented here are less likely to be biased by selective

effects. The extremely close affinity of Jewish and non-Jewish Middle Eastern populations observed here (Tables 2 and 3) supports the hypothesis of a common Middle Eastern origin. Of the Middle Eastern populations included in this study, only the Syrian and Palestinian samples mapped within the central cluster of Jewish populations (Fig. 2). Continued studies of variation in larger samples, additional populations, and at other loci are needed to confirm our inferences as well as to clarify the affinities of Jewish and Middle Eastern Arab populations.

Evolutionary Processes. What do these results tell us about the evolutionary processes that have shaped the structure of the Jewish paternal gene pool? At the most basic level, the genetic distances observed among Jewish and non-Jewish populations can be interpreted as reflecting common ancestry, genetic drift, and gene flow. The latter two processes will tend to increase genetic distances among Jewish populations, whereas admixture will also have the effect of decreasing genetic distances between Jewish and non-Jewish populations. Our results suggest that common ancestry is the major determinant of the genetic distances observed among Jewish communities, with admixture playing a secondary role.

A somewhat surprising result is the small effect that genetic drift appears to have had on genetic distances among Jewish populations. The effects of genetic drift are expected to be greater for Jewish populations because of their reduced effective size during dispersal and as isolated subpopulations in the Diaspora. For the NRY, the effects of drift are expected to be even greater because of its haploid mode of transmission and smaller effective size relative to other nuclear genes. The exaggerated effects of drift were not readily apparent, possibly because large numbers of males participated in the numerous migrations of Jewish populations from the Middle East and within various countries in Europe, North Africa, and West Asia. If we accept this possibility, then the higher relative degree of scatter observed among Jewish populations in discriminant analyses of classical genetic markers (4, 53) may be explained by a combination of drift and selection. At least

two alternative hypotheses should be considered: (i) high rates of recurrent Jewish male gene flow around the Mediterranean, Europe, and the Near East, and/or (ii) higher rates of female gene flow introducing non-Jewish variation into the Jewish gene pool (54). Although some mtDNA studies suggest close affinities of Jewish and Middle Eastern populations (14, 16), comprehensive comparisons of mtDNA variation in Jewish and neighboring non-Jewish populations are not yet available. However, the existing mtDNA data do suggest that contemporary Jewish populations are composed of a diverse set of maternal lineages that diverged $\geq 10,000$ years ago (55). This is similar to our inference that most of the 13 Jewish Y-chromosome haplotypes in Fig. 1 predate the origin of Jewish populations.

In summary, the combined results suggest that a major portion of NRY biallelic diversity present in most of the contemporary Jewish communities surveyed here traces to a common Middle Eastern source population several thousand years ago. The implication is that this source population included a large number of distinct paternal and maternal lineages, reflecting genetic variation established in the Middle East at that time. In turn, this source diversity has been maintained within Jewish communities, despite numerous migrations during the Diaspora and long-term residence as isolated subpopulations in numerous geographic locations outside of the Middle East.

We gratefully acknowledge the excellent technical assistance of Matthew Kaplan, Agnish Chakravarti, Todd Tuggle, Arani Rasanayagam, Christine Ponder, and Jared Ragland. We also thank Laurie Ozelius, Laura Zahn, Giuseppe Passarino, Ornella Semino, and the National Laboratory of Israeli Populations for samples, Robert Griffiths for help with the coalescence analysis, and Stephen Zegura and Karl Skorecki for helpful comments on the manuscript. This publication was made possible by Grant GM-53566 from the National Institute of General Medical Sciences (to M.F.H.). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. M.A.J. is a Wellcome Trust Senior Research Fellow (Grant 057559).

- Goodman, R. M. (1979) *Genetic Disorders Among the Jewish People* (Johns Hopkins Univ. Press, Baltimore).
- Patai, R. & Patai-Wing, J. (1975) *The Myth of the Jewish Race* (Scribner, New York).
- Mourant, A. E., Kopec, A. C. & Domaniewska-Sobczak, K. (1978) *The Genetics of the Jews* (Clarendon, Oxford).
- Carmelli, D. & Cavalli-Sforza, L. L. (1979) *Hum. Biol.* **51**, 41–61.
- Bonné-Tamir, B., Ashbel, S. & Kenett, R. (1979) in *Genetic Diseases Among Ashkenazi Jews*, eds. Goodman, R. M. & Motulsky, A. G. (Raven, New York), pp. 59–76.
- Bonné-Tamir, B., Karlin, S. & Kenett, R. (1979) *Am. J. Hum. Genet.* **31**, 324–340.
- Karlin, S., Kenett, R. & Bonné-Tamir, B. (1979) *Am. J. Hum. Genet.* **31**, 341–365.
- Motulsky, A. G. (1980) in *Population Structure and Disorders*, eds. Eriksson, A. W., Forsius, H. R., Nezanlinna, H. R., Workman, P. L. & Norio, R. K. (Academic, New York), pp. 353–365.
- Kobyliansky, E., Micle, S., Goldschmidt-Nathan, M., Arensburg, B. & Nathan, H. (1982) *Ann. Hum. Biol.* **9**, 1–34.
- Morton, N. E., Yee, S. & Lew, R. (1982) *Curr. Anthropol.* **23**, 157–167.
- Livshits, G., Sokal, R. R. & Kobyliansky, E. (1991) *Am. J. Hum. Genet.* **49**, 131–146.
- Mobini, N., Yunis, E. J., Alper, C. A., Yunis, J. C., Delgado, J. C., Yunis, D. E., Firooz, A., Dowlati, Y., Bahar, K., Gregersen, P. K. & Ahmed, A. R. (1997) *Hum. Immunol.* **57**, 62–67.
- Bonné-Tamir, B. (1985) *Indian Anthropologist* **1**, 153–170.
- Bonné-Tamir, B., Johnson, M. J., Natali, A., Wallace, D. C. & Cavalli-Sforza, L. L. (1986) *Am. J. Hum. Genet.* **38**, 341–351.
- Bonné-Tamir, B., Zoosman-Diskin, A. & Ticher, A. (1992) in *Genetic Diversity Among Jews*, eds. Bonné-Tamir, B. & Adam, A. (Oxford Univ. Press, New York), vol. Ch. 7, pp. 80–94.
- Ritte, U., Neufeld, E., Prager, E. M., Gross, M., Hakim, I., Khatib, A. & Bonne-Tamir, B. (1993) *Hum. Biol.* **65**, 359–385.
- Ritte, U., Neufeld, E., Broit, M., Shavit, D. & Motro, U. (1993) *J. Mol. Evol.* **37**, 435–440.
- Santachiara Benerecetti, A. S., Semino, O., Passarino, G., Torroni, A., Brdicka, R., Fellous, M. & Modiano, G. (1993) *Ann. Hum. Genet.* **57**, 55–64.
- Lucotte, G., Smets, P. & Ruffie, J. (1993) *Hum. Biol.* **65**, 835–840.
- Filon, D., Oron, V., Krichevski, S., Shaag, A., Shaag, Y., Warren, T. C., Goldfarb, A., Shneur, Y., Koren, A., Aker, M., et al. (1994) *Am. J. Hum. Genet.* **54**, 836–843.
- Oddoux, C., Guillen-Navarro, E., DiTivoli, C., DiCave, E., Clayton, C. M., Nelson, H., Sarafoglou, K., McCain, N., Peretz, H., Seligsohn, U., et al. (1999) *J. Clin. Endocrinol. Metab.* **84**, 4405–4409.
- Oppenheim, A., Jury, C. L., Rund, D., Vulliamy, T. J. & Luzzatto, L. (1993) *Hum. Genet.* **91**, 293–294.
- Hakim, I., Gross, M. & Bonné-Tamir, B. (1990) in *Pluridisciplinary Approach of Human Isolates*, eds. Chaventré, A. & Roberts, D. F. (INED, Paris), pp. 43–57.
- Spurdle, A. B. & Jenkins, T. (1996) *Am. J. Hum. Genet.* **59**, 1126–1133.
- Sheffield, V. C., Beck, J. S., Kwitek, A. E., Sandstrom, D. W. & Stone, E. M. (1993) *Genomics* **16**, 325–332.
- Underhill, P. A., Jin, L., Lin, A. A., Mehdi, S. O., Jenkins, T., Vollrath, D., Davis, R. W., Cavalli-Sforza, L. L. & Oefner, P. J. (1997) *Genome Res.* **7**, 996–1005.
- Vollrath, D., Foote, S., Hilton, A., Brown, L. G., Beer-Romero, P., Bogan, J. S. & Page, D. C. (1992) *Science* **258**, 52–59.
- Allen, B. S. & Ostrer, H. (1994) *J. Mol. Evol.* **39**, 13–21.
- Sommer, S. S., Spurdle, A. B. & Bottema, C. D. (1992) *BioTechniques* **12**, 82–87.
- 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., et al. (1997) *Genetics* **145**, 787–805.
- Hammer, M. F. & Horai, S. (1995) *Am. J. Hum. Genet.* **56**, 951–962.
- 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) *Mol. Biol. Evol.* **15**, 427–441.
- Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., et al. (1999) *Am. J. Hum. Genet.* **64**, 817–831.
- Felsenstein, J. (1993) PHYLIP (Univ. of Washington, Seattle).
- Schneider, S., Kueffer, J.-M., Roessli, D. & Excoffier, L. (1997) ARLEQUIN (Univ. of Geneva, Geneva).
- Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) *Genetics* **131**, 479–491.
- Kruskal, J. B. (1964) *Psychometrika* **29**, 1–27.
- Mantel, N. (1967) *Cancer Res.* **27**, 209–220.
- Griffiths, R. C. & Tavare, S. (1994) *Statistical Sci.* **9**, 307–319.
- Bertorelle, G. & Excoffier, L. (1998) *Mol. Biol. Evol.* **15**, 1298–1311.
- Shriver, M. D., Smith, M. W., Jin, L., Marcini, A., Akey, J. M., Deka, R. & Ferrell, R. E. (1997) *Am. J. Hum. Genet.* **60**, 957–964.
- Cavalli-Sforza, L. L. & Edwards, A. W. F. (1967) *Am. J. Hum. Genet.* **19**, 223–257.
- Semino, O., Passarino, G., Brega, A., Fellous, M. & Santachiara-Benerecetti, A. S. (1996) *Am. J. Hum. Genet.* **59**, 964–968.
- Mitchell, R. J. & Hammer, M. F. (1996) *Curr. Opin. Genet. Dev.* **6**, 737–742.
- Kings, M., Sale, A., Bauer, K., Geisert, H., Malek, A. K., Chaix, L., Simon, C., Welsby, D., Di Rienzo, A., Utermann, G., et al. (1999) *Am. J. Hum. Genet.* **64**, 1166–1176.
- Zoosmann-Diskin, A., Ticher, A., Hakim, I., Goldwitsch, Z., Rubinstein, A. & Bonné-Tamir, B. (1991) *Isr. J. Med. Sci.* **27**, 245–251.
- Jorde, L. B. (1992) in *Genetic Diversity Among Jews*, eds. Bonne-Tamir, B. & Adam, A. (Oxford Univ. Press, New York), pp. 305–312.
- Cavalli-Sforza, L. L. & Carmelli, D. (1979) in *Genetic Diseases Among Ashkenazi Jews*, eds. Goodman, R. M. & Motulsky, A. G. (Raven, New York), pp. 93–104.
- Goldstein, D. B., Zhivotovsky, L. A., Nayar, K., Linares, A. R., Cavalli-Sforza, L. L. & Feldman, M. W. (1996) *Mol. Biol. Evol.* **13**, 1213–1218.
- Nachman, M. W. (1998) *Mol. Biol. Evol.* **15**, 1744–1750.
- Szeinberg, A. (1979) in *Genetic Diseases Among Ashkenazi Jews*, eds. Goodman, R. M. & Motulsky, A. G. (Raven, New York), pp. 77–92.
- Hammer, M. F. & Zegura, S. L. (1996) *Evol. Anthropol.* **5**, 116–134.
- Picorell, A., Castro, J. A. & Misericordia Ramon, M. (1997) *Hum. Biol.* **69**, 313–328.
- Sheba, C. (1971) *Isr. J. Med. Sci.* **7**, 1333–1341.
- Tikochinski, Y., Ritte, U., Gross, S. R., Prager, E. M. & Wilson, A. C. (1991) *Am. J. Hum. Genet.* **48**, 129–136.