Am. J. Hum. Genet. 73:768–779, 2003

M O A l a L :YC E M B Na Ea a E a A M

Doron M. Behar,[1,*] Mark G. Thomas,[2,*] Karl Skorecki,[1] Michael F. Hammer,[4] Ekaterina Bulygina,[2] Dror Rosengarten,[1] Abigail L. Jones,[2] Karen Held,[2] Vivian Moses,[2] David Goldstein,[3] Neil Bradman,[2] and Michael E. Weale[2]

[1]Bruce Rappaport Faculty of Medicine and Research Institute, Technion and Rambam Medical Center, Haifa, Israel; [2]The Centre for Genetic Anthropology and [3]The Centre for Population Genetics and Human Health, Department of Biology, University College London, London; and

S>{ŭT∂% of the

n et al. 1999).

In addition to classication by caste, Jews, on the basis of their ancestry and religious practice, can be assigned to one or other of a few long-standing, geographically separated Jewish communities, the most numerous of which are the Ashkenazi and Sephardi groupings (Reif 1993). The term "Ashkenaz" describes a relatively compact area of Jewish settlement in northwestern Europe, including northeastern France and northern Germany, where Jewish settlement is documented dating back to at least the 6th century C.E. From the 10th century, Ashkenazi Jews spoke a common lan-

II. There is uncertainty concerning the relative contributions to Ashkenazi Jewry of, on one hand, western versus eastern immigration of Jews and, on the other hand, internally generated population growth versus conversion to Judaism. In particular, it has been suggested that subjects of the Khazar Empire (located to the northeast of the Black Sea), who had adopted Judaism in the last quarter of the first millennium C.E., were an important constituent of the nascent Ashkenazi community (Encyclopaedia Judaica 1972).

The term "Sephardi" originally described Jews descended from the communities that existed in Spain prior to the expulsion in 1492 C.E. However, current usage applies this designation to all descendants of the communities of North Africa and the Near East who follow the Sephardi rite of worship and cultural traditions. It is thought that, prior to the middle of the 20th century, gene flow between the Ashkenazi and non-Ashkenazi groups was relatively restricted.

The purported different modes of transmission of Levite and Cohen versus Israelite status provide a priori expectations about patterns of genetic variation on the paternally inherited nonrecombining region of the Y chromosome (NRY). In particular, because of recent shared ancestry, Cohanim and Levites would be expected to display lower gene diversity of NRY haplotypes than would Israelites. In addition, the distribution of haplotype frequencies, in the absence of drift, should be similar (*a*) in Cohanim and (*b*) in Levites across the Ashkenazi-Sephardi division, given that this division occurred after the founding of these two groups.

In fact, previous studies have indeed shown that the NRYs of Ashkenazi and Sephardi Cohanim are genetically more closely related to each other than they are to the NRYs of Israelites or non-Jews (Skorecki et al. 1997; Thomas et al. 1998). This pattern arises primarily from differences in the frequency of a particular NRY haplotype (the Cohen Modal Haplotype [CMH], defined by six rapidly mutating microsatellites [Thomas et al. 1998]), and a cluster of closely related haplotypes within a single haplogroup (defined by slowly mutating unique event polymorphisms [UEPs]). Chromosomes belonging to this haplotype and its related cluster were found at high frequency among Cohanim but at a much lower frequency among Israelites. Furthermore, the pattern of diversity within the cluster was found to be consistent with descent from a common ancestor who lived between 2,100 and 3,900 years ago. The CMH is also found, at lower frequency, in non-Jewish populations in the Near East, which would be consistent with its origin in this geographic region. However, the same studies found high frequencies of multiple haplogroups in the Levites, indicating that no single recent origin could be inferred for the majority of this group, despite an oral tradition of a patrilineal descent similar to that of the Cohanim. Moreover, a cluster of closely related NRY haplotypes was identified within a distinctive deep-rooting NRY clade that was found at much higher frequency among Ashkenazi Levites than in either Sephardi Levites or any other Jewish group. However, the reasons for this difference in the Ashkenazi Levites were not explored.

Given the importance of the paternally defined Levite caste in Jewish history and tradition, the multiple theories of the ethnogenesis of the Ashkenazi Jewish community, and a suggestion that Yiddish is a relexified Slavic tongue (Wexler 1993), we undertook a detailed investigation of the paternal genetic history of Ashkenazi Levites and compared the results with matching data from neighboring populations among which the Ashkenazi community lived during its formation and subsequent demographic expansion. By analyzing NRY haplotypes, we have revealed (*a*) a plausible historical explanation for the multiple paternal histories of the Levite caste, (*b*) the probable period during which a European introgression into the caste took place, and (*c*) the likely demographic scale of the event.

S ' M a M

Sampling

We analyzed NRY variation in a set of 988 unrelated males from Ashkenazi Jews, Sephardi Jews, and four non-Jewish European populations. Populations were categorized using a combination of geographic, religious, and ethnohistorical criteria and individual affiliation within a category was determined according to self-designation by the participant providing the sample. Three of the four non-Jewish groups (88 Germans, 112 Sorbians, and 306 Belarusians) were chosen because of their geographic location relative to the ancestral European communities of the Ashkenazi Jews, and also because of prior knowledge of the geographic areas known to have high frequencies of the NRY haplogroup within which the distinctive Ashkenazi Levite high-frequency cluster was previously reported. This haplogroup is designated "R1a1" according to the Y Chromosome Consortium (2002). An additional non-Jewish group consisting of 83 samples previously collected in Norway (Weale et al. 2002) was chosen, to represent a geographic location that excluded Jewish entry until the middle of the 19th century and is known to be outside the area in which Ashkenazi communities originated.

The Jewish samples comprised 236 Ashkenazi Jews (AJ) and 163 Sephardi Jews, who were further divided into 100 Ashkenazi Israelites (AI), 76 Ashkenazi Cohanim (AC), 60 Ashkenazi Levites (AL), 63 Sephardi Israelites (SI), 69 Sephardi Cohanim (SC), and 31 Sephardi Levites (SL). The 60 AL samples were collected

from Ashkenazi males who identified themselves as Levites, with a paternal ancestry from one of the following nine Ashkenazi Jewish communities; Austria-Hungary (10), Belarus (4), France (6), Germany (10), Lithuania (8), Netherlands (5), Poland (7), Romania (4), and Russia (6). Current political borders (including the current borders of Austria and Hungary) were used to define geographic origin. When the Ashkenazi Levite sample was split into western (France, Netherlands, and Germany) versus eastern (all others) communities, no significant differences were found in haplogroup or haplotype frequencies (using the exact test of Raymond and Rousset [1995]), but we note that the sample size is too small to have power to detect anything other than very large differences in frequency. Information regarding paternal lineage was collected from each of the participants. Paternal population affiliation was determined by the oldest male-line ancestor that the participants identified, which extended back to at least the grandparental level. The participants were not related through their

(Bonne-Tamir et al. 2003), as did a study based on mutational analysis of single sperm cells (Holtkemper et al. 2001). In contrast, some other studies based on dating deep-rooting Y chromosome clades have suggested lower "evolutionary" mutation rates (Caglia et al. 1997; Forster et al. 2000). These considerations introduce additional uncertainty in the true value of $\bar{\mu}$ for our loci (Zhivotovsky et al. 2001). We allowed for some uncertainty in $\bar{\mu}$ when calculating confidence limits for the TMRCA, and we note that the length-dependent model described below may also accommodate some locus-specific variation in mutation rate (Carvalho-Silva et al. 1999). Two microsatellite loci typed in our study,

Ta 1

Ha F M aM S· J▼ a F N -J▼ Sa S

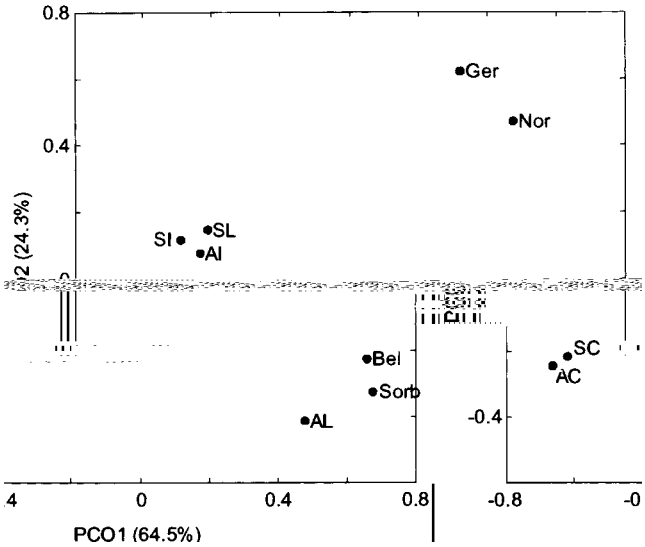
	Frequency in Sample Set[a]									
Haplogroup	AC (n = 76)	SC (n = 69)	AL (n = 60)	SL (n = 31)	AI (n = 100)	SI (n = 63)	Ger (n = 88)	Nor (n = 83)	Sorb (n = 112)	Bel (n = 306)
Y*(xBR,A3b2)	…	…	…	…	…	.0159	…	…	…	…
BR*(xDE,JR)	.0263	.0290	.0333	.1935	.1400	.0794	.3295	.4458	.1786	.2582
E*(xE3a)	.0395	.0435	.2000	.0968	.2200	.1905	.0341	.0120	.0625	.0458
J	.8684	.7536	.1000	.3226	.3700	.3651	.0341	.0241	.0268	.0327
K*(xL,N3,O2b,P)	.0263	.0290	.0333	.2258	.0800	.1270	…	…	…	.0033
L	…	…	…	.0323	…	…	…	…	…	…
N3	…	…	…	…	…	…	.0114	.0361	.0089	.1046
P*(xR1a)	.0263	.0870	.1167	.0968	.1500	.2063	.4659	.2651	.0893	.0425
R1a*	…	…	…	…	…	…	…	…	…	.0033
R1a1	.0132	.0580	.5167	.0323	.0400	.0159	.1250	.2169	.6339	.5098

[a] Population sample labels: AC = Ashkenazi Cohen; SC = Sephardi Cohen; AL = Ashkenazi Levite; SL = Sephardi Levite; AI = Askenazi Israelite; SI = Sephardi Israelite; Ger = German; Sorb = Sorb; Bel = Belarusian; Nor = Norwegian.

The other Jewish data sets separate according to whether they are Cohanim or non-Cohanim, and, within these two groups, there are no significant differences between Ashkenazi and Sephardi data sets (all tests for population differentiation performed using the exact test of Raymond and Rousset (1995). The genetic similarities among these five Jewish data sets range from 0.79 to 1.0. In contrast, the Ashkenazi Levites cluster more with the Slavonic data sets than they do with the other Jewish data sets. The genetic similarities with the other Jewish data sets range from 0.22 to 0.47, whereas the I values with the Sorbian and Belarusian data sets are 0.95 and 0.88, respectively. When a bootstrap test is used, the I value for Ashkenazi Levites with Sorbians and Belarusians is, in both cases, significantly higher than the I value for Ashkenazi Levites with Sephardi Israelites, the most similar Jewish data set to the Ashkenazi Levites (P = .004 for Sorbians and .008 for Belarusians).

At the haplotype level (combining UEP information with the six microsatellite loci), significant differences in haplotype frequency emerge between all data sets (table 3), but the same underlying patterns of similarity remain (fig. 2). Full haplotype information for all data sets is available in supplemental tables A and B (online only). The genetic similarities between the Ashkenazi Levites and the other Jewish data sets range from 0.04 to 0.18, whereas the I values with the Sorbian and Belarusian data sets are 0.52 and 0.49 respectively. Using a bootstrap test, the I value for Ashkenazi Levites with Sorbians and Belarusians is again significantly higher in both cases than the corresponding I value for Ashkenazi Levites with their most similar Jewish group, the Ashkenazi Israelites (P = .006 for Sorbians and .004 for Belarusians).

Within haplogroup R1a1, the microsatellite haplotypes found in the AL data set are tightly clustered around a modal haplotype (16-12-25-10-11-13) that comprises 74% of Ashkenazi Levites within this haplogroup, and 38% of Ashkenazi Levites overall (table 4). This modal haplotype is evenly distributed across the geographically defined communities from which the Ashkenazi Levite sample was taken (see the "Subjects and Methods" section and table A), so that clues to its origin could not be found from these data. The very high frequency of this modal haplotype makes the genetic diversity of Ashkenazi Levite NRYs within the R1a1 haplogroup much lower than in the non-Jewish comparative data sets in which R1a1 is found at high frequency (table 5). Bootstrap tests indicate that, within R1a1, h is significantly lower in AL than in the Sorbian, Belarusian, and Norwegian data sets ($P < .001$ in each case). The P value for the comparison with the German data set is of borderline significance (P = .064), but the number of R1a1 chromosomes within this data set is small (n = 11), resulting in a larger sampling variance. The additional six microsatellites genotyped for Ashkenazi Levites only reduce the modal haplotype frequency within the R1a1 haplogroup to 58%, also confirming the high degree of haplotype homogeneity (table 6). As is explained in the "Discussion" section, this arrangement of R1a1 haplotypes within the Ashkenazi Levite data set is consistent with common descent from a recent ancestor and is unlikely to have resulted from a large number of founding lineages. By assuming that the most recent common ancestor (MRCA) possessed the modal haplotype and by using a male intergeneration time of 25 years, we estimated a mean TMRCA of 663 years before present under the Simple Stepwise Mutation Model and a mean time of 1,000 years before present under the Linear Length-Dependent Stepwise Mutation Model (see the "Subjects and Methods" section). These dates coincide with the historically estimated time when compact settlements of Jews in northwest Europe began and the important religious



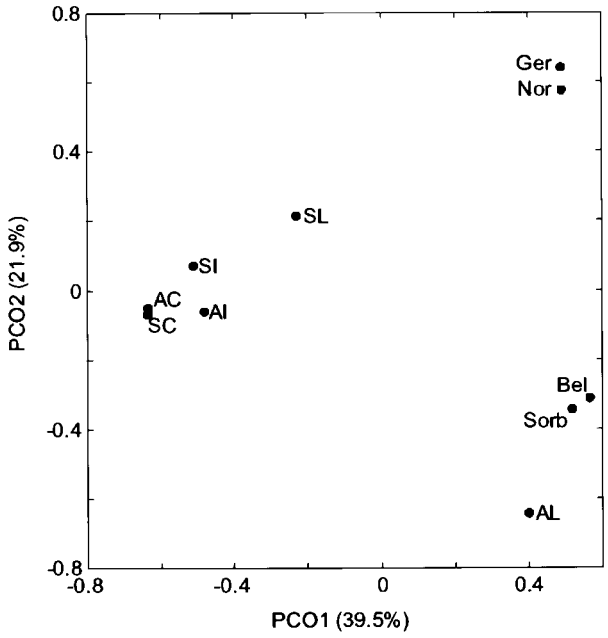
F- **1** Principal coordinates plot of the genetic identity values (haplogroup level) shown in table 2. Axis labels indicate the percentage explained by the first two principal axes.

Ta **2**

Pa-w C a- a Ha L a S- J-w a F N -J-w Sa S

	AC	SC	AL	SL	AI	SI	Ger	Nor	Sorb	Bel
AC	…	.407	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
SC	1.002	…	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
AL	.216	.293	…	<.001	<.001	<.001	<.001	<.001	<.001	<.001
SL	.793	.821	.381	…	.111	.200	<.001	<.001	<.001	<.001
AI	.821	.856	.468	.977	…	.548	<.001	<.001	<.001	<.001
SI	.819	.860	.425	.983	1.021	…	<.001	<.001	<.001	<.001
Ger	.106	.192	.426	.507	.510	.523	…	.048	<.001	<.001
Nor	.089	.159	.515	.531	.461	.398	.927	…	<.001	<.001
Sorb	.071	.144	.953	.273	.276	.205	.475	.656	…	.002
Bel	.089	.152	.875	.342	.307	.212	.510	.745	.975	…

NOTE.—Lower left triangle = Nei's Genetic Identity, *I*; upper right triangle = *P* values for



F 2 Principal coordinates plot of the genetic identity values (haplotype level) shown in table 4. Axis labels indicate the percentage explained by the first two principal axes.

Ta 3

Pa w C a a Ha L a S J w a F N -J w Sa S

	AC	SC	AL	SL	AI	SI	Ger	Nor	Sorb	Bel
AC	…									

Ta 4

M M a Ha F M Ha R1a1 a M S J a F N -J Sa S

Microsatellite Haplotype[a]	Frequency in Sample Set									
	AC	SC	AL	SL	AI	SI	Ger	Nor	Sorb	Bel
14 12 25 10 11 13	...	...	...	...	...	...	...	...	...	.0033
15 12 24 10 11 12	...	...	...	...	...	...	...	...	...	.0033
15 12 24 10 11 13	...	...	.0333	...	...	...	...	...	...	.0098
15 12 24 11 11 13	...	...	...	...	...	...	...	.0120	.0179	.0098
15 12 25 10 11 13	.0132	...	.0333	.0323	.0100	...	.0114	.0482	.0536	.0294
15 12 25 10 12 13	...	...	...	...	...	...	...	...	...	.0065
15 12 25 11 11 13	...	...	...	...	...	...	...	.0241	.0089	.0458
15 12 25 11 11 14	...	...	...	...	...	...	.0114	.0120	...	...
15 12 26 10 11 13	...	...	...	...	...	...	...	...	...	.0065
15 12 26 11 11 13	...	...	...	...	...	...	...	.0241	...	.0065
15 13 25 10 13 13	...	...	...	...	...	...	...	...	...	.0033
15 13 25 11 11 13	...	...	...	...	...	...	...	...	.0089	...
15 14 25 11 11 13	...	...	...	...	...	...	...	.0120	...	...
16 11 24 11 11 14	...	...	...	...	...	...	...	...	...	.0033
16 12 24 10 11 12	...	...	.0167	...	...	...	...	...	...	...
16 12 24 10 11 13	...	...	.0167	...	...	.0159	...	.0120	.0893	.0131
16 12 24 10 11 14	...	...	...	...	...	...	...	...	...	.0065
16 12 24 10 12 13	...	...	...	...	...	...	...	...	...	.0033
16 12 24 11 11 13	...	.0145	...	...	...	...	...	...	...	.0229
16 12 25 10 11 12	...	...	...	...	...	...	.0114	...	...	...
16 12 25 10 11 13	...	.0145	.3833	...	.0100	...	...	.0120	.1071	.0850
16 12 25 11 11 12	...	...	...	...	...	...	...	...	...	.0065
16 12 25 11 11 13	...	.0290	...	...	.0200	...	.0568	.0361	.0179	.1209
16 12 25 11 11 14	...	...	...	...	...	...	...	...	...	.0033
16 12 25 11 12 13	...	...	...	...	...	...	...	...	...	.0065
16 12 25 12 11 13	...	...	...	...	...	...	...	...	...	.0033
16 12 26 10 11 13	...	...	...	...	...	...	...	...	.0089	...
16 12 26 11 11 13	...	...	...	...	...	...	...	...	.0089	.0131
16 13 25 10 11 13	...	...	...	...	...	...	...	...	...	.0163
16 13 25 10 11 15	...	...	...	...	...	...	...	...	...	.0033
16 13 25 11 11 14	...	...	...	...	...	...	...	...	...	.0033
17 10 24 10 11 13	...	...	...	...	...	...	...	.0120	...	...
17 12 23 11 11 14	...	...	...	...	...	...	...	.0120	...	...
17 12 24 10 11 13	...	...	.0167	...	...	...	.0114	...	.0268	.0033
17 12 24 10 11 14	...	...	...	...	...	...	...	...	.0089	...
17 12 24 11 11 13	...	...	...	...	...	...	...	...	.0089	...
17 12 25 10 11 12	...	...	...	...	...	...	...	...	.0089	...
17 12 25 10 11 13	...	...	.0167	...	...	...	.0114	...	.1786	.0425
17 12 25 10 11 14	...	...	...	...	...	...	...	...	...	.0033
17 12 25 10 12 13	...	...	...	...	...	...	...	...	.0089	...
17 12 25 11 11 13	...	...	...	...	...	...	...	...	.0357	.0131
17 12 26 10 11 13	...	...	...	...	...	...	.0114	...	.0268	.0033
17 12 26 10 12 13	...	...	...	...	...	...	...	...	...	.0033
17 13 25 10 11 13	...	...	...	...	...	...	...	...	...	.0033
17 13 25 11 11 13	...	...	...	...	...	...	...	...	...	.0033
17 13 26 11 11 13	...	...	...	...	...	...	...	...	.0089	...

NOTE.—Population sample labels are the same as in table 1.

[a] Composed of the repeat sizes of microsatellite loci, in the following order: DYS19, DYS388, DYS390, DYS391, DYS392, DYS393.

the Cohen Modal Haplotype, for example, which belongs to a haplogroup that is more likely to be of Near

Ta 5

Sa S a M M a a D Ha R1a1

	Value in Sample Set									
Variable	AC	SC	AL	SL	AI	SI	Ger	Nor	Sorb	Bel
No. of Subjects in R1a1 Haplogroup	1	4	31	1	4	1	11	18	71	156
Haplotype Diversity[a]	…	…	.4516	…	…	…	.8182	.9281	.86721	.896
SE for Haplotype Diversity	…	…	.1097	…	…	…	.1191	.0401	.0245	.0142

[a] By use of the unbiased gene diversity formula (Nei 1987), for cases where $n > 10$.

cific to the Ashkenazi Levites. This is borne out by the fact that the haplogroup is virtually absent from the Sephardi Levites, which would indicate an event that occurred after the separate formation of the Ashkenazi and Sephardi groupings, and also by the very recent inferred date for the common ancestor of the NRYs

Ta 6

M M a Ha F M w Ha R1a1 12 M M a
T A a L Sa

6-Locus[a] and Constituent 12-Locus[b] Haplotypes	Count	Frequency in Haplogroup	Frequency in Sample
15 12 24 10 11 13:			
[15 11 14 12 13 17 24 10 11 13 12 14]	2	.0645	.0333
15 12 25 10 11 13:			
[15 11 14 12 13 17 25 10 11 13 12 14]	1	.0323	.0167
[15 11 14 12 13 16 25 10 11 13 12 14]	1	.0323	.0167
16 12 24 10 11 12:			
[16 11 14 12 13 18 24 10 11 12 12 14]	1	.0323	.0167
16 12 24 10 11 13:			
[16 11 14 12 13 17 24 10 11 13 12 14]	1	.0323	.0167
16 12 25 10 11 13:			
[16 11 14 12 13 16 25 10 11 13 12 14]	4	.1290	.0667
[16 11 14 12 13 18 25 10 11 13 12 14]	1	.0323	.0167
[16 11 14 12 13 17 25 10 11 13 12 14]	18	.5806	.3000
17 12 24 10 11 13:			
[17 11 14 12 13 17 24 10 11 13 12 14]	1	.0323	.0167
17 12 25 10 11 13:			
[17 11 14 12 13 17 25 10 11 13 12 14]	1	.0323	.0167

[a] Composed of the repeat sizes of microsatellite loci, in the following order: DYS19, DYS388, DYS390, DYS391, DYS392, DYS393.

[b] Shown in brackets and composed of the repeat sizes of microsatellite loci, in the following order: DYS19, DYS385a, DYS385b (arbitrarily given larger repeat size over DYS385a), DYS388, DYS389-I, DYS389-II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS439.

but not at high frequency elsewhere. An alternative ex-

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Kruger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Keys D (1999) Catastrophe: an investigation into the origins of the modern world. Ballantine Books, New York

Nebel A, Filon D, Weiss DA, Weale M, Faerman M, Oppenheim A, Thomas MG (2000) High-resolution Y chromosome haplotypes of Israeli and Palestinian Arabs reveal geographic substructure and substantial overlap with haplotypes of Jews. Hum Genet 107:630–641

Nei M (1987) Molecular evolutionaty genetics. Columbia University Press,