

ARTICLE

# MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population

Doron M Behar<sup>1</sup>, Michael F Hammer<sup>\*,2</sup>, Daniel Garrigan<sup>2</sup>, Richard Villems<sup>3</sup>, Batsheva Bonne-Tamir<sup>4</sup>, Martin Richards<sup>5</sup>, David Gurwitz<sup>4</sup>, Dror Rosengarten<sup>1</sup>, Matthew Kaplan<sup>2</sup>, Sergio Della Pergola<sup>6</sup>, Lluís Quintana-Murci<sup>7</sup> and Karl Skorecki<sup>1</sup>

<sup>1</sup>Bruce Rappaport Faculty of Medicine and Research Institute, Technion and Rambam Medical Center, Haifa, Israel; <sup>2</sup>Division of Biotechnology, University of Arizona, Tucson, AZ, USA; <sup>3</sup>Tartu University & Estonian Biocenter, Tartu, Estonia; <sup>4</sup>Department of Human Genetics, Sackler School of Medicine, Ramat Aviv, Israel; <sup>5</sup>Department of Chemical and Biological Sciences, University of Huddersfield, Huddersfield, UK; <sup>6</sup>Avraham Harman Institute of Contemporary Jewry, Hebrew University, Jerusalem, Israel; <sup>7</sup>CNRS URA1961, Institut Pasteur, Paris, France

The relative roles of natural selection and accentuated genetic drift as explanations for the high frequency of more than 20 Ashkenazi Jewish disease alleles remain controversial. To test for the effects of a maternal bottleneck on the Ashkenazi Jewish population, we performed an extensive analysis of mitochondrial DNA (mtDNA) hypervariable segment 1 (HVS-1) sequence and restriction site polymorphisms in 565 Ashkenazi Jews from different parts of Europe. These patterns of variation were compared with those of five Near Eastern ( $n = 327$ ) and 10 host European ( $n = 849$ ) non-Jewish populations. Only four mtDNA haplogroups (Hgs) (defined on the basis of diagnostic coding region RFLPs and HVS-1 sequence variants) account for ~70% of Ashkenazi mtDNA variation. While several Ashkenazi Jewish mtDNA Hgs appear to derive from the Near East, there is also evidence for a low level of introgression from host European non-Jewish populations. HVS-1 sequence analysis revealed increased frequencies of Ashkenazi Jewish haplotypes that are rare or absent in other populations, and a reduced number of singletons in the Ashkenazi Jewish sample. These diversity patterns provide evidence for a prolonged period of low effective size in the history of the Ashkenazi population. The data best fit a model of an early bottleneck (~100 generations ago), perhaps corresponding to initial migrations of ancestral Ashkenazim in the Near East or to Europe. A genetic bottleneck followed by the recent phenomenon of rapid population growth are likely to have produced the conditions that led to the high frequency of many genetic disease alleles in the Ashkenazi population.

*European Journal of Human Genetics* (2004) 12, 355–364. doi:10.1038/sj.ejhg.5201156

Published online 14 January 2004

**Keywords:** Ashkenazi; bottleneck; mtDNA

## Introduction

The term 'Ashkenazi' refers to Jewish people of recent European ancestry, with an historical separation from

other major Jewish populations in North Africa and the Middle East.<sup>1</sup> The contemporary Ashkenazi gene pool is thought to have originated from a founding deme that migrated from the Near East within the last two millennia.<sup>2</sup> After moving through Italy and the Rhine Valley, the Ashkenazi population presumably experienced a complex demographic history characterized by numerous migrations and fluctuations in population size. During the past

\*Correspondence: Dr M Hammer, Department of EEB, Biosciences West, University of Arizona, Tucson, AZ 85721, USA. Tel: + 1 520 621 9828; Fax: + 1 520 621 9247; E-mail: mfh@u.arizona.edu  
Received 17 June 2003; revised 28 October 2003; accepted 20 November 2003

~500 years, there was a period of rapid growth<sup>3</sup> culminating in an estimated population size of ~8 million Ashkenazi Jews at the outbreak of the Second World War.<sup>4</sup>

The relatively high frequency of more than 20 known recessive disease alleles in Ashkenazi populations has fueled a lively debate regarding the comparative contribution of heterozygote advantage,<sup>5</sup> as well as both recent<sup>6</sup> and more ancient founder effects.<sup>7</sup> Genetic studies focusing on the demographic history of Ashkenazi Jewish communities have the potential to address questions regarding the relative contribution of population level effects on the origins and frequency of disease alleles. In particular, haploid regions of the genome such as mitochondrial DNA (mtDNA) and the nonrecombining portion of the Y chromosome (NRY) that are unusually sensitive to genetic drift should be useful for detecting the effects of bottlenecks on Ashkenazi populations (eg, Fay and Wu<sup>8</sup>). One challenge in the use of these haploid regions is that they behave as single locus systems and are subject to large stochastic effects.

Studies of the paternally inherited NRY supported the hypothesis of a common Near Eastern origin of many Jewish populations, including Ashkenazi Jews, with low rates of admixture with host non-Jewish populations.<sup>9–11</sup> In a recent study based on mtDNA sequence variation at a portion of the hypervariable segment 1 (HVS-1), the authors inferred separate maternal founding events for several Jewish populations, with limited subsequent gene flow from surrounding host populations.<sup>12</sup> Interestingly, the Ashkenazi Jewish sample in this study appeared to be an exception to this pattern, showing no strong signal of a founding event at the level of HVS-1 sequence variation. To address the question of whether mtDNA from Ashkenazi populations exhibit signs of a genetic bottleneck, we perform a more extensive analysis of mtDNA genetic variation at HVS-1, as well as at haplogroup(Hg)-defining coding region mutations, in a sample of 565 Jews from 15 different Ashkenazi communities originating in western and eastern Europe (Table 1), and compare these patterns of variation with those of neighboring non-Jewish popula-

**Table 1** MtDNA HVS sequence variation in Ashkenazi Jews and European and Near Eastern non-Jews

Community	$n^a$	$k^b$	$h^c \pm SD$	$\pi^d \pm SD$	$F_s^e$	$p^f$
<i>Ashkenazi Jews</i> <sup>g</sup>	528	118	0.957 ± 0.003	0.014 ± 0.007		
French (Fr)	46	21	0.925 ± 0.021	0.015 ± 0.001	-5.889	0.027
German (Ge)	32	16	0.925 ± 0.027	0.013 ± 0.001	-4.450	0.039
Dutch (Du)	20	13	0.953 ± 0.028	0.013 ± 0.007	-4.137	0.027
Austro-Hungarian (Au)	29	19	0.966 ± 0.018	0.013 ± 0.001	-8.922	<0.001
Lithuanian (Li)	30	17	0.947 ± 0.022	0.012 ± 0.007	-6.296	0.007
Polish (Po)	165	58	0.952 ± 0.008	0.015 ± 0.008	-25.182	<0.001
Romanian (Ro)	93	40	0.947 ± 0.011	0.014 ± 0.008	-24.318	<0.001
Russian (Ru)	59	39	0.981 ± 0.007	0.015 ± 0.008	-25.366	<0.001
Ukrainian (Uk)	54	29	0.973 ± 0.008	0.012 ± 0.007	-18.448	<0.001
<i>Near East NJ</i>	327	221	0.992 ± 0.002	0.017 ± 0.009		
Bedouin (Bed)	29	27	0.995 ± 0.011	0.021 ± 0.002	-20.433	<0.001
Druze (Dru)	45	25	0.946 ± 0.019	0.013 ± 0.001	-13.211	<0.001
Egyptian (Eg)	67	57	0.991 ± 0.006	0.019 ± 0.001	-25.061	<0.001
Palestinian (Pal)	117	101	0.995 ± 0.003	0.016 ± 0.001	-25.154	<0.001
Syrian (Syr)	69	57	0.991 ± 0.006	0.016 ± 0.001	-25.368	<0.001
<i>European NJ</i>	849	350	0.975 ± 0.003	0.011 ± 0.006		
French (Fre)	71	52	0.968 ± 0.014	0.012 ± 0.001	-25.839	<0.001
German (Ger)	174	111	0.973 ± 0.007	0.011 ± 0.001	-25.685	<0.001
Austrian (Aus)	99	66	0.957 ± 0.015	0.012 ± 0.001	-25.747	<0.001
Polish (Pol)	37	24	0.961 ± 0.018	0.009 ± 0.001	-19.501	<0.001
Romanian (Rom)	92	57	0.982 ± 0.005	0.012 ± 0.001	-25.744	<0.001
Russian (Rus)	25	21	0.987 ± 0.015	0.012 ± 0.001	-16.311	<0.001
Bavarian (Bav)	49	35	0.981 ± 0.091	0.011 ± 0.006	-25.886	<0.001
Swiss (Swi)	70	43	0.962 ± 0.013	0.009 ± 0.005	-26.235	<0.001
Czech (Cze)	83	59	0.987 ± 0.005	0.012 ± 0.007	-25.788	<0.001
Estonia (Est)	149	80	0.974 ± 0.007	0.012 ± 0.007	-25.588	<0.001

<sup>a</sup>Sample size.

<sup>b</sup>Number of haplotypes.

<sup>c</sup>Haplotype diversity.

<sup>d</sup>Nucleotide diversity.

<sup>e</sup>The statistic of  $F_u$ <sup>19</sup>.

<sup>f</sup>One-tailed probability of the observed  $F_s$  occurring in a constant-sized population.

<sup>g</sup>Only samples with sizes  $\geq 20$  are listed.

tions. In our analysis, we take advantage of the ability to infer evidence for maternal population bottlenecks on the basis of comparative estimates of mtDNA sequence diversity.<sup>13</sup>

## Subjects and methods

### Samples

Blood or buccal swab samples were collected with informed consent from 565 unrelated individuals (404 male and 161 female subjects) of Ashkenazi Jewish origin according to procedures approved by the University of Arizona and Rambam Medical Center Human Subjects Committees. Each of the volunteers reported the birthplace of their mother, grandmother, and in most cases, great grandmother. The samples were then classified into the following geographic regions: western Europe,  $n=101$  (French Rhine Valley,  $n=46$ ; Germany,  $n=32$ ; The Netherlands,  $n=20$ ; and the Swiss Rhine Valley,  $n=3$ ), and eastern Europe,  $n=464$  (Austro-Hungary,  $n=29$ ; Belarus,  $n=9$ ; Czechoslovakia,  $n=15$ ; Estonia,  $n=1$ ; Lithuania,  $n=30$ ; Latvia,  $n=8$ ; Moldavia,  $n=1$ ; Poland,  $n=165$ ; Romania,  $n=93$ ; Russia,  $n=59$ ; and the Ukraine,  $n=54$ ). MtDNA variation in these populations was compared with published data from relevant European and Near Eastern non-Jewish populations.<sup>14</sup> The comparison populations included western Europeans,  $n=463$  (Austrians,  $n=99$ ; Bavarians,  $n=49$ ; French,  $n=71$ ; Germans,  $n=174$ ; Swiss,  $n=70$ ), eastern Europeans,  $n=386$  (Czechs,  $n=83$ ; Estonians,  $n=149$ ; Poles,  $n=37$ ; Romanians,  $n=92$ ; Russians,  $n=25$ ), and Near Easterners,  $n=327$  (Bedouin,  $n=29$ ; Druze,  $n=45$ ; Egyptians,  $n=67$ ; Syrians,  $n=69$ ; and Palestinians,  $n=117$ ). Several of the statistical analyses were restricted to Ashkenazi and European non-Jewish populations that had sample sizes of  $n \geq 20$ , and that came from similar geographic regions in Europe (Table 1).

### Genotyping

Sequences of HVS-1 of the control region were determined from position 16024–16383, by use of the ABI Prism Dye Terminator cycle-sequencing protocols developed by Applied Biosystems (Perkin-Elmer). Only the last three positions of each site position (ie, 024–383) are listed to identify the HVS-1 haplotype or sequence motif. HVS-1 sequence data were used to define 'haplotypes' within 'Hgs'; the latter were defined on the basis of coding region RFLPs and control region polymorphisms. RFLP typing of coding-region sites that are diagnostic for the major mtDNA Hgs was performed hierarchically beginning with deep-rooting markers of the mtDNA phylogeny.<sup>14–16</sup> The RFLP markers tested in this study together with the corresponding Hgs they define are shown in the supplemental data.

## Statistical analyses and computer simulations

Diversity indices were calculated using the ARLEQUIN 2.000 software package.<sup>17</sup> ARLEQUIN was also used to calculate the summary statistic  $F_s$ <sup>18</sup> and the distribution of pairwise nucleotide differences between sequences (mismatch distribution)<sup>19</sup> to test the deviation of the observed data from neutral predictions expected in a constant-sized population.  $F_s$  is based on the probability of recovering a number of haplotypes greater than or equal to the observed number in a sample drawn from a stationary population with the same mean number of pairwise differences as the observed sample. A significantly negative value of  $F_s$  indicates an excess of mutations in the terminal branches of the gene genealogy, a pattern commonly attributed to rapid population growth.<sup>18</sup> The one-tailed probability of the observed  $F_s$  statistic under the null hypothesis of constant population size was determined with the coalescent simulation methodology implemented in the ARLEQUIN software package.<sup>17</sup> Confidence intervals for the observed frequency of the  $i$ th class ( $f_i$ ) in the mismatch distribution were obtained with 500 bootstrap replicates in the computer package ARLEQUIN. Median-joining networks for HVS-1 sequence types were constructed with the use of the computer application Network 3.0.<sup>20</sup> We performed nonmetric multidimensional scaling (MDS)<sup>21</sup> on Hg frequencies using the software package NTSYS.<sup>22</sup>

The observed HVS-1 sequence data were used to estimate two demographic parameters in a population bottleneck model: the number of generations since the onset of the putative bottleneck ( $t_0$ ) and the magnitude of the reduction in population size ( $r$ ). We examined values of  $t_0$  ranging from 25 to 250 generations and values of  $r = \{1, 20, 50, 100, 200, 400\}$ . In the latter case,  $1/r$  is the fraction of the ancestral Jewish population that founded the extant Ashkenazi population. After the instantaneous reduction in population size, the simulated population again grew exponentially to its original size. The null hypothesis of no bottleneck is represented by the case of  $r=1$ . For each combination of  $t_0$  and  $r$ , 1000 coalescent simulations were performed, each consisting of 300 sequences of 360 bp in length. In all simulations, the ancestral population experienced a 1000-fold increase in population size 2000 generations ago.<sup>23</sup> A mutation rate of  $1.2 \times 10^{-3}$  per sequence per generation was assumed. As  $f_0$  and  $f_1$  have been previously shown to be especially sensitive to population bottlenecks,<sup>24</sup> we recorded the probability of obtaining simulated values of  $f_0$  and  $f_1$  that were higher than those observed in the Ashkenazi data, for each simulation bout with parameters  $t_0$  and  $r$ .

## Results

### Ashkenazi Hgs and haplotypes

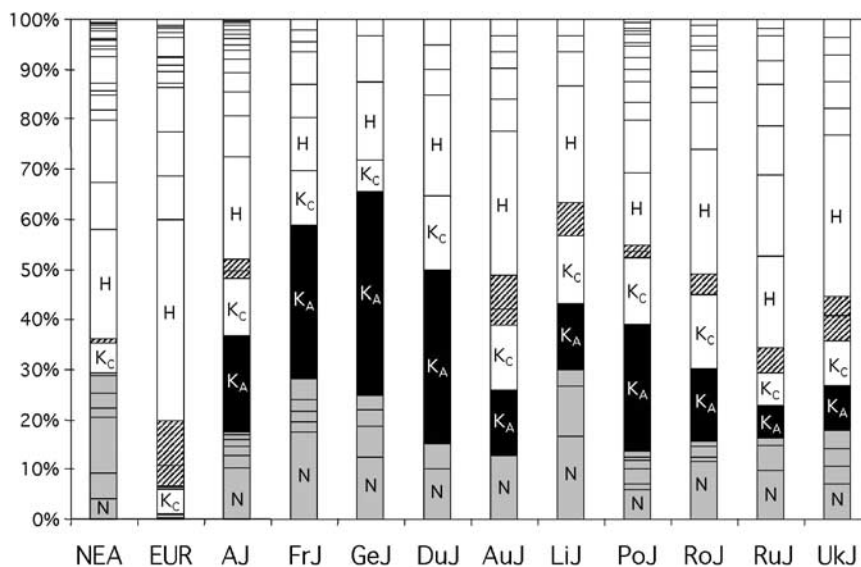
A total of 56 mtDNA Hgs is observed in the entire Jewish and non-Jewish data set ( $n=1741$ ), 37 of which are found

in Ashkenazi Jews. To facilitate visual inspection of Hg frequency differences among populations, these 56 Hgs are combined into 28 Hgs as indicated in Figure 1. The most prevalent Ashkenazi Hgs were K (32%), H (21%), N1b (10%), and J1 (7%), followed by other Hgs at minor frequencies (supplementary material). Haplogroup diversity in the combined Ashkenazi sample, as well as for each community, is lower than that for either the European or Near Eastern non-Jewish populations (data not shown).

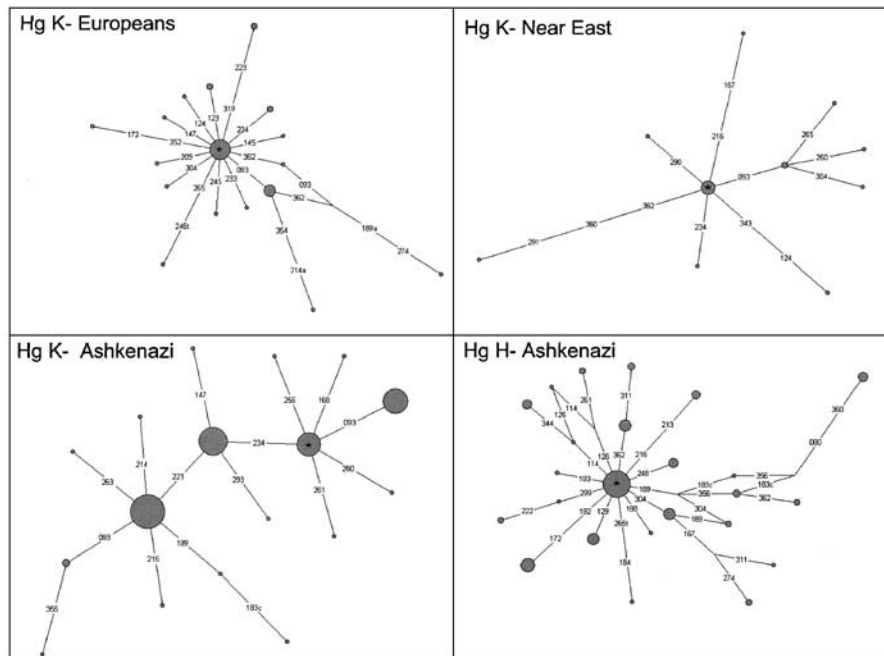
The high frequency of Hg K among the Ashkenazi Jews contrasts with its much lower frequency (~6%) reported in both Near Eastern and European non-Jews.<sup>14</sup> A notable exception among Near Eastern non-Jews is the Druze, with a Hg K frequency of 16%.<sup>14,15</sup> Based on HVS-1 sequences, there are four major subtypes of Hg K in the Ashkenazi Jewish sample. These subtypes and their frequencies within Hg K are: 223-224-234-311 (33%); 224-234-311 (24%); 093-224-311 (19%); and 224-311 (16%). Like the ancestral 224-311 haplotype of Hg K, its one-step derivative haplotype 093-224-311 is widespread (at low to moderate frequency) in both Near Eastern and European non-Jews. In contrast, the 224-234-311 haplotype is rare among European and Near Eastern non-Jewish populations. A single direct match was found in a sample of Palestinian Arabs.<sup>14</sup> However, HVS-1 sequence matches occur elsewhere at very low frequency in both Europe and the Near East, and therefore, its provenance remains unclear. Interestingly, the most

frequent Ashkenazi haplotype, 223-224-234-311 (present at 10.6% in our Ashkenazi sample), is virtually absent in European and Near Eastern non-Jewish populations.<sup>14</sup> In Figure 1, we have subdivided these four Hg K haplotypes into two classes: K-Common includes those Hg K types that are geographically widespread (224-311 and 093-224-311), while K-Ashkenazi includes the two types (224-234-311 and 223-224-234-311) that are almost entirely restricted to Ashkenazi communities. The high frequency of haplotypes in Ashkenazi populations that are rare or absent in other populations may be an indication of the effects of strong genetic drift acting on the Ashkenazi population.

Evidence of the effects of accentuated genetic drift is also visible in the shape of gene trees and patterns of diversity within individual Ashkenazi mtDNA Hgs. When we perform median-joining network analysis<sup>20</sup> using HVS-1 sequence variation, the topologies of both the European and Near Eastern non-Jewish Hg K HVS-1 networks are star like (with a single most frequent haplotype), consistent with a history of population expansion from small initial size (Figure 2a, b, respectively). In contrast, the topology of Ashkenazi Hg K haplotypes is not star like, with the aforementioned four haplotypes occurring at intermediate frequency (Figure 2c). In addition, the second most frequent Hg found in Ashkenazi populations, Hg H (Figure 1), exhibits elevated frequencies of the derived nonsingleton tip nodes (Figure 2d), in contrast to the usual



**Figure 1** MtDNA Hg frequencies in Ashkenazi Jewish (AJ) and non-Jewish populations from Europe (EUR) and the Near East (NEA). Bar charts are shaded according to the geographic distribution of Hgs: those that are present mainly in Near Eastern populations are represented as solid gray bars, those present mainly in European populations (ie, U5 and V) are represented as striped bars, those that appear to be limited to Ashkenazi populations are represented as solid black bars, and those that are present at moderate frequency in both European and Near Eastern non-Jewish populations are represented as open bars. Three letter codes are as defined in Table 1. Hgs were combined as follows (from bottom to top): N (N\*, N1a, N1b, N1c, N9a), pHV, L (L1a, L1b, L1e, L2, L3, L3b, L3d, L3e), U7, M1, U1 (U1a, U1b), K<sub>A</sub> and K<sub>C</sub> (see text), V, U5 (U5\*, U5a\*, U5a1\*, U5b\*, U5b1), H, J (J\*, J1, J2), T (T\*, T1-T5), HV1, U6 (U6a\*, U6a1, U6b), HV\*, W, X, I, M\*, U4, U1a/U1b, U2/ U2e, U3, R (R\*, R1, R2), A, B, D, U\*.



**Figure 2** Median-joining HVS-1 haplotype networks for Hgs K and H. Circle sizes are proportional to the frequency of HVS-1 sequence haplotypes within each Hg, with the smallest circle representing a singleton. Numbers along lines indicate the HVS-1 site that has undergone a mutation. An asterisk denotes haplotype 224–311 in Hg K networks, and CRS in the Hg H network.

‘star’ pattern of singletons at derived tip nodes seen in the Hg H networks of European and Near Eastern non-Jewish populations (data not shown).

In examining HVS-1 sequences associated with the third most common Hg in our Ashkenazi sample, N1b (shown as part of N in Figure 1), we note a remarkable lack of haplotype diversity: 54 of the 57 Ashkenazi N1b haplotypes have the same HVS-1 sequence motif (145-176A-223), with a diversity index of 0.06. This is the second most common HVS-1 haplotype (9.6%) in our Ashkenazi sample. Despite its much lower frequency in Europeans and Near Eastern non-Jews, Hg N1b has an overall diversity of 0.3 in these groups.<sup>14</sup> It is also important to note that the Ashkenazi N1b haplotype differs by a single transition from what has been considered the ancestral N1b haplotype (145-176G-223), and has not been previously reported. Since Hg N1b is almost entirely restricted to the Near East,<sup>14,15</sup> we infer a Near Eastern origin for the derived Ashkenazi N1b haplotype. Similarly, several other Hgs (L2, pre-HV, U7, M1, and U1b) found at varying frequencies in Ashkenazi populations also appear to have Near Eastern/Mediterranean/African origins (Figure 1).

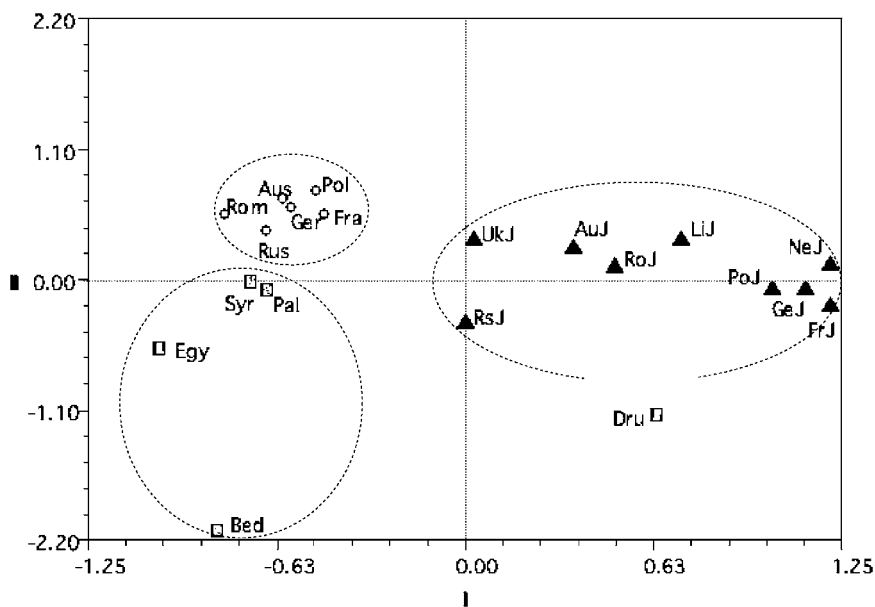
#### Multidimensional scaling (MDS) and genetic distances between populations

Results of MDS based on mtDNA Hg frequencies (as in Figure 1) are shown in Figure 3 (for which the stress value was 0.08). In contrast to both the Near Eastern non-Jews

and the Ashkenazi Jews, European non-Jews form a tight cluster. The Ashkenazi communities form a linear array along the right central portion of the plot, with some eastern communities closer to the center and all western communities closer to the right-hand side of the plot. The increased dispersion among Ashkenazi subpopulations may be the result of strong genetic drift<sup>24,25</sup> and/or admixture between eastern Ashkenazim and their Eastern European non-Jewish hosts. Support for the latter hypothesis comes from the finding of higher frequencies of the mtDNA Hgs U5 and V in eastern European *versus* western European Ashkenazim (Figure 1). The higher dispersion of Near Eastern relative to European populations in the plot could be the result of stronger genetic drift in the Bedouins and Druze, two populations with small effective sizes.<sup>26</sup> Although not readily apparent in the MDS plot, AMOVA analysis (results not shown) indicated that Ashkenazi populations as a whole are genetically more similar to Near Eastern non-Jewish populations than to European non-Jewish populations.

#### HVS-1 sequence diversity and frequency distribution

Table 1 reports HVS-1 haplotype diversity ( $h$ ) and nucleotide diversity ( $\pi$ ) values for our Ashkenazi samples, as well as for 10 European non-Jewish populations from similar geographic regions and five Near Eastern non-Jewish populations. The results indicate many fewer Ashkenazi mtDNA haplotypes (after correcting for sample size)



**Figure 3** Multidimensional scaling plot for nine Ashkenazi (triangles), six European (open circles), and five Near Eastern (squares) non-Jewish populations based on mtDNA Hg frequencies. The 'stress' value (ie, the goodness of fit between the distances in the graphic configuration and the monotonic function of the original distances) is 0.08. For three letter population codes, see Table 1.

compared with both European and Near Eastern non-Jewish populations. The use of a two-tailed *t*-test revealed that the *h* diversity measure for all Ashkenazi Jewish mtDNAs is significantly lower than the corresponding diversity measures for European and Near Eastern non-Jewish mtDNAs ( $P < 0.0001$ ). Nucleotide diversity ( $\pi$ ) does not differ significantly among the groups (Table 1). This discrepancy results from a reduced number of HVS-1 haplotypes within Hgs that are fairly differentiated at the DNA sequence level.

The frequency distribution of polymorphisms also differs between Ashkenazi Jewish and non-Jewish populations. Table 1 demonstrates that nearly all of the Near Eastern and European non-Jewish populations yield significantly negative  $F_S$  values, consistent with an excess of rare mutations. Interestingly, all three western Ashkenazi  $F_S$  values are not statistically significant (ie, at the  $\alpha = 0.02$  level), and two of the six Eastern Ashkenazi  $F_S$  values are not nearly as negative as those of other populations in Table 1.

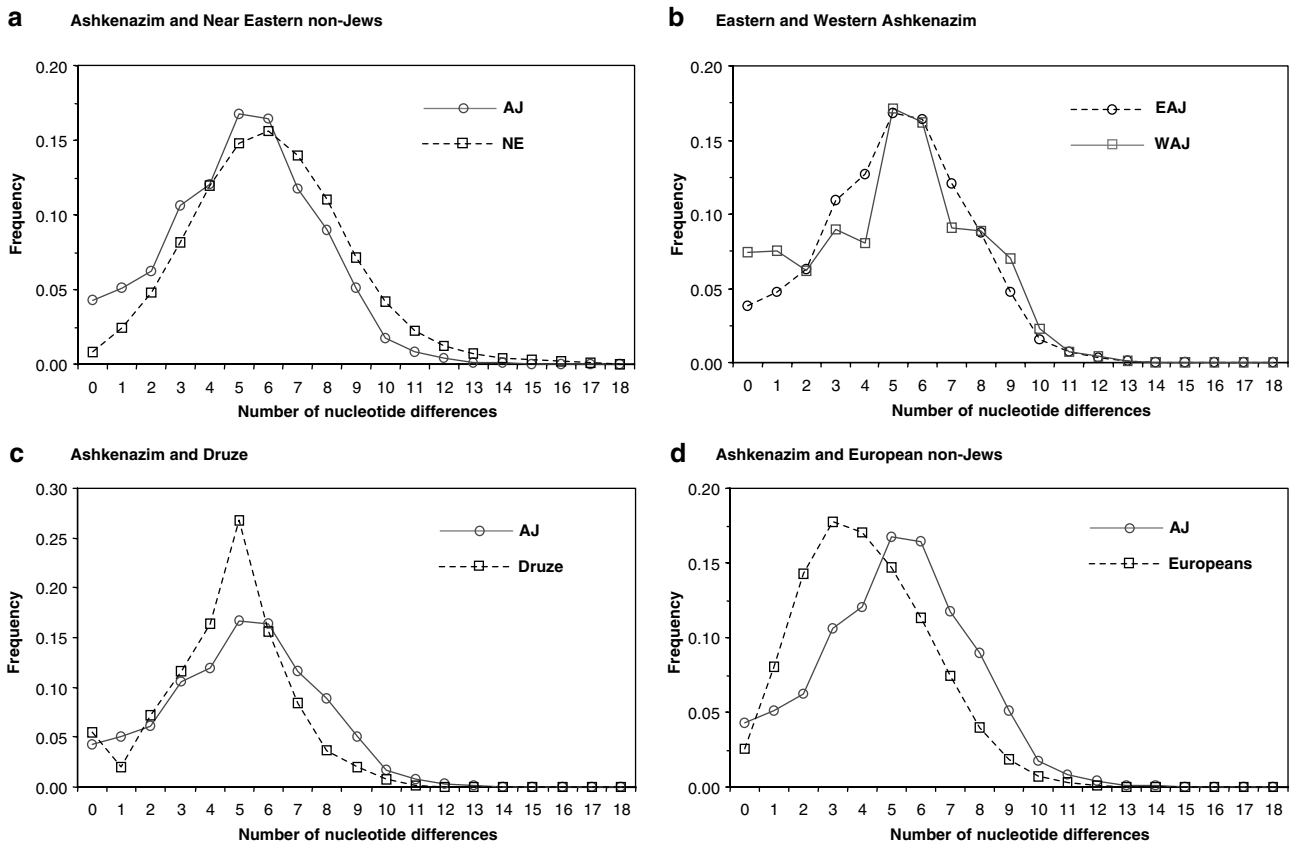
### Mismatch distributions

Visual inspection of the distribution of pairwise nucleotide differences reveals considerable insight into the recent demographic history of the Ashkenazi Jewish population. Figure 4a contrasts the mismatch distributions of large Near Eastern populations (Egyptians, Palestinians, and Syrians) with that of the Ashkenazi Jews. The means of both distributions ( $\tau$ ) are nearly identical ( $\tau = 5.5$  for Ashkenazi Jews and  $\tau = 6.2$  for Near Easterners), which

reflects a shared period of ancient population growth (approximately 46 000–52 000 years ago). The Near Eastern estimate of  $\tau$  was used in the coalescent simulations to generate  $f_0$  and  $f_1$  values under the null hypothesis of no population bottleneck and Pleistocene population growth. Despite the similarity of the  $\tau$  estimates for the Ashkenazim and Near Eastern populations, the Ashkenazi genealogy harbors an increased rate of recent coalescent events, as evidenced by an increase in  $f_0$  and  $f_1$  of the mismatch distribution.<sup>14</sup> The 95% confidence interval for  $f_0$  in the Near Eastern dataset is 0.008–0.020, while the observed Ashkenazi Jewish value is 0.043. Similarly, the 95% confidence interval for the Near Eastern dataset  $f_1$  is 0.007–0.049, the Ashkenazi Jewish  $f_1 = 0.051$ .

Figure 4b shows the mismatch distributions for both Eastern and Western Ashkenazi Jewish populations, which are identical, except for elevated  $f_0$  and  $f_1$  in the West. Figure 4c contrasts the Ashkenazi Jewish mismatch distribution with that of the Druze, a Near Eastern population with low effective size.<sup>26</sup> The Druze  $\tau$  of 4.6 is similar to that of the Ashkenazi Jews, again suggesting a period of shared growth. However, the Druze show an increase in  $f_0$ , but not  $f_1$ , as do the Ashkenazi Jewish populations. Lastly, Figure 4d juxtaposes the Ashkenazi Jewish mismatch distribution with that of the European non-Jewish population. For European non-Jews,  $\tau = 3.4$ , indicating a much more recent onset of population growth (approximately 20 000–25 000 years ago).

Coalescent simulations performed under the null hypothesis of no population bottleneck ( $r = 1$ ) did not



**Figure 4** Mitochondrial mismatch distributions for (a) Ashkenazi Jewish and Near Eastern (Egyptian, Palestinian, and Syrian) populations, (b) Eastern and Western Ashkenazi Jewish populations, (c) Ashkenazi Jewish and Druze populations, and (d) Ashkenazi Jewish and non-Jewish European populations. The Ashkenazi Jewish population shares a period of rapid expansion with both the Near Eastern and Druze populations, while the non-Jewish European population appears to have a more recent onset of population growth.

produce a single replicate with values of  $f_0$  higher than that observed in the Ashkenazi data ( $P < 0.001$ ). Similarly, simulations of a constant size model only produced 19 replicates with a value of  $f_1$  higher than that observed for the Ashkenazi data ( $P = 0.019$ ). Based upon the assumption that Jewish and Near Eastern populations share a common history of Pleistocene population growth, the null hypothesis of a constant-sized Ashkenazi population is clearly rejected on the basis of the observed increase in  $f_0$  and  $f_1$ . The probability of the observed Ashkenazi values of  $f_0$  only exceeded 0.05 when  $100 \leq r \leq 400$  and  $t_0 \geq 100$  generations. The parameter combination with the highest probability of generating the observed data was  $t_0 = 150$  generations and  $r = 200$ . This corresponds to a period of time prior to the Jewish Diaspora out of the Near East.

## Discussion

The results presented here portray a pattern of highly reduced mtDNA diversity for the Ashkenazi population (Table 1), an unusually large proportion of mtDNA

haplotypes that are unique to the Ashkenazi gene pool, and a reduction in frequency of rare haplotypes and singleton sites compared with Near Eastern populations. For example, the three most frequent Ashkenazi haplotypes (223-224-234-311, 145-176A-223, 224-234-311) account for 27.8% of total mtDNA repertoire in our Ashkenazi sample. These Ashkenazi mtDNA haplotypes are virtually absent from surrounding non-Jewish populations and therefore provide a genetic signature of the Ashkenazi maternal gene pool, and bear witness to the strong effects of genetic drift acting on this population. This contrasts with the situation in both Near Eastern and European non-Jewish populations, where only a single haplotype (CRS) was found at elevated frequencies (ie, above 5%).

There are several periods in the history of Jewish populations when bottlenecks may have occurred, for example: (1) in the Near East before the initial migration to Europe (eg, > 1500 years ago), (2) during the migrations of Jews from the Near East to Italy after the 1st century A.D., (3) upon establishment of small communities in the Rhine

Valley in the 8th century A.D., and (4) in the 12th century A.D., when migrations took place from western to eastern Europe. In addition, endogamy in combination with >100-fold population growth in the last 500 years<sup>27</sup> undoubtedly played a role in shaping patterns of variation in the Ashkenazi gene pool. While several authors posited that the high frequency of genetic conditions, such as Tay-Sachs disease, is the result of heterozygote advantage,<sup>5,28–30</sup> others have argued for an important role of genetic drift.<sup>6,26</sup> For example, Risch *et al.*<sup>6</sup> proposed that founder effects resulting from the dynamics of population growth in the 16–19th centuries, especially in the northern Jewish Pale of Settlement (Lithuania and Belarus), explain most, if not all of the genetic diseases observed at high frequency in the Ashkenazi population today. This hypothesis was supported by the inference of a recent age of the single founder mutation (~350 years) that causes early-onset idiopathic torsion dystonia.<sup>6</sup> The much older estimated age of the factor XI type II mutation (~3000 years), which has a high frequency in both Ashkenazi and Iraqi Jewish populations, implies that its frequency is largely independent of the recent demographic upheavals particular to the Ashkenazi population.<sup>7</sup> This raises the possibility of either an ancient bottleneck in a population ancestral to the major Jewish groups, and/or positive selection on heterozygotes.<sup>7</sup> Comparisons of ‘neutral’ genealogies are particularly useful for drawing inferences regarding demographic history, and DNA variation at the uniparentally inherited regions of the genome are particularly sensitive to bottlenecks.<sup>8</sup> While mtDNA and the NRY cannot provide a complete picture of population history by themselves, the large database of mtDNA HVS-1 sequences from dozens of human populations and a large body of theoretical work on the effects of population expansion/contraction on mtDNA sequence variation<sup>24,31</sup> provide a comparative framework for inferring the demographic history of the Ashkenazi population. Population expansions and bottlenecks are known to shape patterns of variation in the genome in specific ways. After a period of time, rapid population growth from a small initial size is expected to produce a preponderance of recently derived low-frequency polymorphisms that often result in statistically significant negative values of the  $F_S$ <sup>18</sup> statistic and unimodal peaks (waves) in mtDNA mismatch distributions with a near absence of zero and one mismatches among haplotypes.<sup>19,32,33</sup> Studies of mtDNA HVS-1 sequence diversity demonstrate that most human populations show evidence of remote (Pleistocene) expansion, and that such demographic signals in patterns of DNA variation may be lost due to recent bottlenecks.<sup>13</sup> For example, nonsignificant  $F_S$  values observed in some hunter-gatherer populations have been explained by recent bottlenecks that may have occurred in response to the Neolithic expansions of food-producing populations. Recently, Cordeaux *et al.*<sup>25</sup> demonstrated that several southern Indian tribal popula-

tions exhibit higher frequencies of the zero and one classes of mismatch distributions and more positive  $F_S$  values than do northern populations. These authors inferred that southern tribal groups experienced enhanced genetic drift because of small population sizes and/or bottlenecks that erased the signature of Pleistocene population expansion. Here, we demonstrate that the Ashkenazi population also exhibits the signature of reduced population size as seen in the higher  $f_0$  and  $f_1$  mismatch class frequencies and  $F_u$ 's  $F_S$  values. However, the magnitude of the bottleneck in these populations was not so severe as to entirely erase the earlier signal of Pleistocene population expansion<sup>34</sup> and the shared history of population growth is clearly demonstrated by comparing the Ashkenazi Jewish mismatch distribution with those of Near Eastern populations (Figures 4a and c). The observed mutational frequency peak for the Ashkenazi and Near Eastern non-Jewish populations is similar and consistent with the age of the Pleistocene expansion, which is older than that inferred from the mutational frequency peak for European non-Jews. This is consistent with a Near East origin for a major portion of the Ashkenazi Jewish mtDNA pool.

If the Jewish population bottleneck did begin in the Near East, other Jewish populations from around the world are predicted to harbor similar values of  $f_0$  and  $f_1$  in their mismatch distributions. To test this prediction, we examined the mismatch distributions resulting from the data of Thomas *et al.*,<sup>12</sup> which includes samples of the Bukharan, Georgian, Indian, Iranian, Iraqi, Moroccan, and Yemenite Jewish communities. All HVS-1 sequence datasets showed a significantly elevated  $f_0$  (only Sephardic Jews showed an increase in  $f_1$ ) relative to Near Eastern non-Jewish populations, with  $\tau \approx 6$  (data not shown). This result implies that global Jewish communities suffered a common bottleneck in the Near East, or independent founder events during the Jewish Diaspora.

Our computer simulations confirm that the frequencies of the zero and one class of the Ashkenazi mismatch distribution are significantly elevated over that observed for the sequences sampled from Near Eastern populations. This is a strong indication of a recent population bottleneck<sup>24</sup> and further simulations suggest that the data best fit a 200-fold reduction in size ~150 generations ago. It should be noted that this estimate depends largely upon two major assumptions: the Ashkenazim have not admixed with European host populations and that the mutation rate is  $1.2 \times 10^{-3}$  per sequence per generation. It is therefore necessary to consider the effects of local admixture on the mismatch distributions of various Jewish populations. Admixture is known to decrease the frequencies of the zero and one classes, creating an ‘intermatch distribution’.<sup>35</sup> By assuming no admixture, our inferences of bottleneck events and their respective times may be biased towards more recent dates. However, a faster mutation rate will result in a more recent estimate of  $t_0$  and *vice versa* for a



slower rate. The chosen rate places the time of the Pleistocene expansion at 50 000 years before the present, a time that is consistent with other estimates from both autosomal and mitochondrial data.<sup>23,36–38</sup>

We may also infer the effects of bottlenecks on some specific Near Eastern non-Jewish populations who are known to have small effective sizes such as the Druze (Figure 4c).<sup>39</sup> This suggests the possibility that contemporary Ashkenazi mtDNA diversity may derive, in part, from a small and subdivided ancestral mtDNA gene pool, and is consistent with the hypothesis that some high frequency disease alleles originated before the separation of Jewish communities in the Near East.<sup>40,41</sup> Indeed, estimates of the age of mutations causing Ashkenazi genetic diseases range from recent times (ie, during demographic upheavals within Europe in the past 500 years),<sup>6,26,40,42</sup> to times when ancestral Ashkenazi populations were first migrating to and within Europe,<sup>43</sup> to times before Jewish populations migrated out of the Near East.<sup>40,44,45</sup>

## Conclusions

The combined mtDNA and disease mutation data suggest that Ashkenazi Jewish populations experienced a long period of accentuated genetic drift marked by an early bottleneck, perhaps beginning in the Near East. Prolonged periods of low effective population size can lead to the accumulation of slightly deleterious mutations throughout the genome.<sup>46</sup> Small founder populations derived from large ancestral populations are not always capable of purging these deleterious mutations. This may be the ultimate cause of the segregation of disease mutations in Ashkenazi Jews. However, this explanation does not preclude more proximal causes for the increase in frequency of disease mutations, such as those hypothesized by Risch *et al.*,<sup>7</sup> unequal contribution of a particular segment of the Ashkenazi Jewish community to the explosive population growth occurring in the Pale of Settlement approximately 25 generations ago. Low effective size may have enabled deleterious mutations to become established in the Jewish population, while the recent growth of affected segments of the community amplified these mutations to frequencies sufficiently high to form homozygotes.

## Acknowledgements

We thank Laurent Excoffier for valuable comments, Dr Marc-Alain Levy for help in collecting samples from the French Rhine Valley, the National Laboratory for the Genetics of Israeli Populations for contributing samples to this study and David Goldstein for providing non-Ashkenazi Jewish mtDNA data. This work was supported by a grant from the National Institute of General Medical Sciences (GM53566-06) to MH and an award grant from the Israeli Science Foundation to KS, as well as a kind donation from the Milin Charitable Foundation.

## References

- 1 Ben-Sasson HH: *History of the Jewish People*. Cambridge: Harvard University Press; 1976.
- 2 Goodman RM: *Genetic Disorders Among the Jewish People*. Baltimore: The Johns Hopkins University Press; 1979.
- 3 Weinryb BD: *The Jews of Poland: a Social and Economic History of the Jewish Community of Poland From 1100 to 1800*. Philadelphia: The Jewish Publication Society of America; 1972.
- 4 Dellapergola S: Major demographic trends of world Jewry: the last hundred years. in Adam A (ed): *Genetic Diversity Among Jews*. New York: Oxford University Press; 1992; **Chapter 1**: pp 3–30.
- 5 Jorde LB: Population diseases in the Ashkenazi population: evolutionary considerations. in Bonne-Tamir B, Adam A (eds): *Genetic Diversity Among Jews*. New York, Oxford University Press; 1992, pp 305–312.
- 6 Risch N, de Leon D, Ozelius L *et al*: Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 1995; **9**: 152–159.
- 7 Goldstein DB, Reich DE, Bradman N, Usher S, Seligsohn U, Peretz H: Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. *Am J Hum Genet* 1999; **64**: 1071–1075.
- 8 Fay JC, Wu CI: A human population bottleneck can account for the discordance between patterns of mitochondrial versus nuclear DNA variation. *Mol Biol Evol* 1999; **16**: 1003–1005.
- 9 Santachiara Benerecetti AS, Semino O, Passarino G *et al*: The common, Near-Eastern origin of Ashkenazi and Sephardi Jews supported by Y-chromosome similarity. *Ann Hum Genet* 1993; **57**: 55–64.
- 10 Hammer MF, Redd AJ, Wood ET *et al*: Jewish and middle eastern non-jewish populations share a common pool of Y-chromosome biallelic haplotypes (in process citation). *Proc Natl Acad Sci USA* 2000; **97**: 6769–6774.
- 11 Nebel A, Filon D, Brinkmann B, Majumder PP, Faerman M, Oppenheim A: The Y chromosome pool of Jews as part of the genetic landscape of the Middle East. *Am J Hum Genet* 2001; **69**: 1095–1112.
- 12 Thomas MG, Weale ME, Jones AL *et al*: Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* 2002; **70**: 1411–1420.
- 13 Excoffier L, Schneider S: Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc. Natl. Acad. Sci. USA* 1999; **96**: 10597–10602.
- 14 Richards M, Macaulay V, Hickey E *et al*: Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 2000; **67**: 1251–1276.
- 15 Macaulay V, Richards M, Hickey E *et al*: The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* 1999; **64**: 232–249.
- 16 Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS: Genetic evidence of an early exit of Homo sapiens sapiens from Africa through eastern Africa. *Nat Genet* 1999; **23**: 437–441.
- 17 Schneider S, Roessli D, Excoffier L: *Arlequin Software for Population Genetic Analysis*. Geneva: Genetics and Biometry Laboratory, University of Geneva; 2000.
- 18 Fu Y: Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 1997; **147**: pp 915–925.
- 19 Rogers A, Harpending HC: Population growth makes waves in the distribution of pairwise differences. *Mol Biol Evol* 1992; **9**: 552–569.
- 20 Bandelt HJ, Forster P, Rohlf A: Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 1999; **16**: 37–48.
- 21 Kruskal JB: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 1964; **29**: 1–27.
- 22 Rohlf FJ: *NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System*. Setauket, NY: Exeter Software; 1998.

- 23 Rogers AR: Genetic evidence for a Pleistocene population explosion. *Evolution* 1995; **49**: 608–615.
- 24 Excoffier L, Schneider S: Why hunter-gatherer populations do not show signs of pleistocene demographic expansions. *Proc Natl Acad Sci USA* 1999; **96**: 10597–10602.
- 25 Cordaux R, Saha N, Bentley GR, Aunger R, Sirajuddin SM, Stoneking M: Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet* 2003; **11**: 253–264.
- 26 Durst R, Colombo R, Shpitzen S *et al*: Recent origin and spread of a common Lithuanian mutation, G197del LDLR, causing familial hypercholesterolemia: positive selection is not always necessary to account for disease incidence among Ashkenazi Jews. *Am J Hum Genet* 2001; **68**: 1172–1188.
- 27 Weinryb BD: *The Jews of Poland: a Social and Economic History of the Jewish Community in Poland From 1100 to 1800*. Philadelphia: The Jewish Publication Society of America; 1972.
- 28 Chakravarti A, Chakraborty R: Elevated frequency of Tay-Sachs disease among Ashkenazic Jews unlikely by genetic drift alone. *Am J Hum Genet* 1978; **30**: 256–261.
- 29 Motulsky AG: Possible selective effects of urbanization on Ashkenazi Jews. in Goodman RM, Motulsky AG (eds): *Genetic Diseases Among Ashkenazi Jews*. New York: Raven Press; 1979, pp 301–314.
- 30 Diamond JM: Human Genetics. Jewish Lysosomes. *Nature* 1994; **368**: 291–292.
- 31 Rogers AR, Harpending HC: Population growth makes waves in the distribution of pairwise differences. *Mol Biol Evol* 1992; **9**: 552–569.
- 32 DiRienzo A, Wilson AC: Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc Natl Acad Sci USA* 1991; **88**: 1597–1601.
- 33 Slatkin M, Hudson RR: Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 1991; **129**: 555–562.
- 34 Rogers AR, Jorde LB: Genetic evidence on modern human origins. *Hum Biol* 1995; **67**: 1–36.
- 35 Harpending HC, Sherry ST, Rogers AR, Stoneking M: The genetic-structure of ancient human-populations. *Curr Anthropol* 1993; **34**: 483–496.
- 36 Ward RH, Frazier BL, Dew-Jager K, Paabo S: Extensive mitochondrial diversity within a single Amerindian tribe. *Proc Natl Acad Sci USA* 1991; **88**: 8720–8724.
- 37 Wall JD, Przeworski M: When did the human population size start increasing? *Genetics* 2000; **155**: 1865–1874.
- 38 Pluzhnikov A, Di Rienzo A, Hudson RR: Inferences about human demography based on multilocus analyses of noncoding sequences. *Genetics* 2002; **161**: 1209–1218.
- 39 Baldwin CT, Weiss S, Farrer LA *et al*: Linkage of congenital, recessive deafness (DFNB4) to chromosome 7q31 and evidence for genetic heterogeneity in the Middle Eastern Druze population. *Hum Mol Genet* 1995; **4**: 1637–1642.
- 40 Goldstein DB, Reich DE, Bradman N, Usher S, Seligsohn U, Peretz H: Age estimates of two common mutations causing factor XI deficiency: recent genetic drift is not necessary for elevated disease incidence among Ashkenazi Jews. *Am J Hum Genet* 1999; **64**: 1071–1075.
- 41 Ostrer H: A genetic profile of contemporary Jewish populations. *Nat Rev Genet* 2001; **2**: 891–898.
- 42 Ellis NA, Ciocci S, Proytcheva M, Lennon D, Groden J, German J: The Ashkenazic Jewish Bloom syndrome mutation blmAsh is present in non-Jewish Americans of Spanish ancestry. *Am J Hum Genet* 1998; **63**: 1685–1693.
- 43 Diaz GA, Gelb BD, Risch N *et al*: Gaucher disease: the origins of the Ashkenazi Jewish N370S and 84GG acid beta-glucosidase mutations. *Am J Hum Genet* 2000; **66**: 1821–1832.
- 44 Bar-Sade RB, Kruglikova A, Modan B *et al*: The 185delAG BRCA1 mutation originated before the dispersion of Jews in the diaspora and is not limited to Ashkenazim. *Hum Mol Genet* 1998; **7**: 801–805.
- 45 Patael Y, Figer A, Gershoni-Baruch R *et al*: Common origin of the I1307K APC polymorphism in Ashkenazi and non-Ashkenazi Jews. *Eur J Hum Genet* 1999; **7**: 555–559.
- 46 Ohta T: Slightly deleterious mutant substitutions in evolution. *Nature* 1973; **246**: 96–98.

(Supplementary information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>))