Am. J. Hum. Genet. 49:131–146, 1991

Genetic Affinities of Jewish Populations

Gregory Livshits,*,† Robert R. Sokal,* and Eugene Kobyliansky†

*Department of Ecology and Evolution, State University of New York at Stony Brook; and †Research Unit—Human Population Biology, Division of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv

Summary

Genetic relations between various Jewish (J) and non-Jewish (NJ) populations were assessed using two sets of data. The first set contained 12 pairs of matched J and NJ populations from Europe, the Middle East, and North Africa, for which 10 common polymorphic genetic systems (13 loci) were available. The second set included 22 polymorphic genetic systems (26 loci) with various numbers of populations (ranging from 21 to 51) for each system. Therefore, each system was studied separately. Nei's standard genetic distance (*D*) matrices obtained for these two sets of data were tested against design matrices specifying hypotheses concerning the affiliations of the tested populations. The tests against single designs were carried out by means of Mantel tests. Our results consistently show lower distances among J populations than with their NJ neighbors, most simply explained by the common origin of the former. Yet, there is evidence also of genetic similarity between J and corresponding NJ populations, suggesting reciprocal gene flow between these populations or convergent selection in a common environment. The results of our study also indicate that stochastic factors are likely to have played a role in masking the descent relationships of the J populations.

Introduction

Jewish populations dispersed from the Eastern Mediterranean in the main between 600 B.C. and A.D. 100. The history of their subsequent residence in and movements between various countries in Europe, North Africa, and the Middle East has resulted in a nearly unique pattern of genetic relationships among Jewish (J) populations and between them and the non-Jewish (NJ) peoples among whom they live. Whereas the factors molding the genetic structure of the typical European population are spatial differentiation, migration, and amalgamation of native gene pools (Sokal et al. 1989), J populations have the potential for exhibiting a more dendritic or hierarchic structure due to their migration and branching history. Spatial differentiation based on isolation by distance is possible for them only in situations where they were numerous and widespread enough and resident for a sufficient period to permit the effects to become manifest.

Computer simulation studies have shown that the percentage of shared polymorphic alleles declines very rapidly in groups of populations after their splitting, even when daughter populations are of as large an effective size as 10,000 individuals (Li and Nei 1977; Fuerst 1985). Because the sizes of J populations frequently were small, there is evidence (Carmelli and Cavalli-Sforza 1979; Morton et al. 1982) of the additional significant effects of chance sampling of the gene pool (genetic drift, bottleneck effects). Finally, although religious law and custom proscribed intermarriage with the surrounding NJ populations, there is little doubt that some genetic admixture took place.

In recent years various investigators have examined the genetic affinities of J populations to each other and to their NJ neighbors in an attempt to unravel what must be a complex system of interrelations. The published results and conclusions concerning this topic are quite contradictory. For example, Mourant et al. (1978, p. 57) maintained that "each major [Jewish] community as a whole bears some resemblance to indigenous peoples of the region where it first developed." Morton et al. (1982), using the modified kin-

Received June 18, 1990; final revision received February 19, 1991.

Address for correspondence and reprints: Dr. Robert R. Sokal, Department of Ecology and Evolution, State University of New York, Stony Brook, NY 11794–5245.

0002-9297/91/4901-0016$02.00

ship function of Malécot, concluded that there was substantial intermixture between J and neighboring NJ populations. Other investigators reached very different conclusions (e.g., Karlin et al. 1979; Kobyliansky et al. 1982; Bonne-Tamir 1985). Their studies of various J populations demonstrated a considerably greater genetic similarity for most pairs of J populations than between J and NJ communities. It has also been shown for ABO and MN that inbreeding within J populations as measured by F_{is} is higher than in NJ populations, whereas Wright's standardized genetic variance (F_{st}) among J groups is lower than among NJ communities (Kobyliansky and Livshits 1983). Rao and Boudreau (1984) concluded that (1) European Jews and non-Jews form two distinct and closely knit clusters, and (2) North African and Iraqi Jews are closer to the European Jews than to the corresponding non-Jews.

Some other authors obtained intermediate results. Wijsman (1984) believes that Ashkenazi Jews have a low to moderate NJ genetic component, consistent with both the low estimates suggested by Karlin et al. (1979) and higher estimates suggested by others (Carmelli and Cavalli-Sforza 1979; Morton et al. 1982). Computing the level of admixture separately for the different loci, Cavalli-Sforza and Carmelli (1979) have found no admixture for HLA-A and HLA-B loci, but various amounts of admixture for other loci. Similar variability across loci was reported by Motulsky (1980). The relative genetic contributions to the J gene pools of common ethnic origins and of admixture with neighboring populations continue, therefore, to interest investigators and to engender controversy. Unfortunately the various published studies are not entirely comparable. They differ in the number and provenance of both J and NJ populations investigated and in the number and identity of the genetic loci analyzed.

We readdress this issue in the present study for two reasons. We have assembled the largest combination of gene frequencies based on independent loci for matched J and NJ populations reported to date, and also the largest number of loci for individual J populations. Second, we introduce a new analytical technique to examine the problem—testing genetic distances by means of Mantel tests against design matrices. We believe that this technique is better suited to answering the relevant questions than previously applied methods, since it permits direct tests of various contrasts in genetic distance matrices.

Material and Methods

The Data Base

The gene frequencies on which the present study is based were extracted from numerous published sources and include also new data obtained by G.L. and E.K. Some of these data, mainly on blood group loci, were published elsewhere (Kobyliansky et al. 1982; Kobyliansky and Livshits 1989). Table 1 provides a list of populations used in the present study and the number of polymorphic loci available for each of these populations, as well as the average sample size across the loci per population. Altogether 22 genetic systems were employed, comprising the following 26 loci: (1) blood groups ABH, ABO, Duffy, Kell, Kidd, Lutheran, MNSs, P, and Rhesus (three loci); (2) enzymes and proteins ACP1, ADA, AK1, G6PD, GLO1, GPT1, PGW, PGM1, HPA, and TF; and (3) histocompatibility loci HLA-A, HLA-B, KM. The maximal number of populations available for any one locus was 51 for ABO, comprising 23 J and 28 NJ populations. However, these numbers (furnished in table 6 below) fluctuated considerably among loci. Although we did not always possess data for matching J and NJ populations from the same country, we had at least information from neighboring countries at our disposal.

Because matching J and NJ samples were not always available for any one locus, we assembled a subset of 12 pairs of matched population samples for which gene-frequency data were available for the following 10 common genetic systems (comprising 13 loci): ABO, Duffy, Kell, Kidd, MNSs, P, Rhesus (CCDEe). ACP1, AK1, and PGM1. These samples contained J and NJ populations from the following geographic regions and countries: Middle East: Yemen, Iran, and Iraq; North Africa: Morocco and Libya; eastern Europe: Poland, Russia, and Georgia; central Europe: Germany and Czechoslovakia; southern Europe: Bulgaria, Turkey (J only), and Spain (NJ only). This last pair was associated not only because we lacked sufficient systems for a matching Turkish NJ sample, but also because we believe that a Spanish NJ sample is the more appropriate match for these Judeo-Spanish-speaking Turkish Jews, who settled in Turkey some time after their expulsion in 1492 from Spain, where they had lived since Roman times.

The gene frequencies for most of the 26 loci noted above were treated as independent. The MNSs and Rhesus gene complexes were treated as single loci us-

Table 1

Number of Genetic Systems Available for Various J and NJ Populations and Average Sample Size across Systems

	J		NJ	
Country or Region	No. of Genetic Systems Available	Average Sample Size across Systems	No. of Genetic Systems Available	Average Sample Size across Systems
Algeria	5	654	13	989
Austria	13	112	18	1,820
Bukhara, USSR	17	125	2	167
Bulgaria	12	217	14	1,867
Cochin, India	12	136	13	230
Czechoslovakia	12	239	16	514
Egypt	15	274	20	333
Georgia, USSR	18	123	19	264
Germany	19	309	19	1,759
Greece	. . .	. . .	19	861
Hungary	11	166	18	592
Italy	. . .	. . .	21	584
Iran	19	235	20	202
Iraq	20	572	16	418
Israel/Jordan	. . .	. . .	20	183
Kurdistan	16	245	16	178
Lebanon/Syria	7	186	9	313
Libya	20	207	15	360
Morocco	20	340	11	201
Poland	19	548	22	529
Rumania	15	450	9	345
Russia, USSR	19	291	19	906
Saudi Arabia	. . .	. . .	19	283
Spain	. . .	. . .	19	370
Tunisia	13	104	8	291
Turkey	14	126	16	243
Yemen	22	371	15	207
Yugoslavia	6	105	9	1,467

ing haplotype frequencies. The alleles at the HLA-A and HLA-B loci are known to exhibit linkage disequilibrium (Hedrick et al. 1986), but the allelic association is not as strong as in the case of the MNSs and Rhesus loci, nor are haplotype frequencies available. We, therefore, treated them as independent loci, but for the analyses of separate systems the results for both HLA loci were averaged.

Statistical Analysis

Nei's standard genetic distance (*D;* Nei 1987) for each pair of populations was calculated for each locus over the available populations, and also for the 10 systems combined in the set of 12 population pairs mentioned above. To survey the structure inherent in these distance matrices, we used both hierarchical cluster analysis and ordination procedures. The clustering method employed was unweighted pair-group analysis using arithmetic averages (UPGMA; Sneath and Sokal 1973, p. 230) applied to the genetic distance matrices. We represented the results as phenograms—dendrograms of phenetic relationships (Sneath and Sokal 1973, p. 260). (This is the established meaning of phenogram in population and evolutionary biology and in systematics [Camin and Sokal 1965; Mayr 1965] and differs from the meaning of the same term coined by Cotterman [1953] for m-allele phenotype systems in human genetics.) To display the relationships among the populations by means of ordinations, we carried out nonmetric multidimensional scaling (Sneath and Sokal 1973, p. 249) on these distance matrices. Both of these computations were carried out

Table 2

Illustrative Example for Mantel Test of Genetic Distance Matrix against Design Matrix

	Distance Matrix					Design Matrix				
	J1	J2	J3	NJ1	NJ2	J1	J2	J3	NJ1	NJ2
J2.....	.190					−1				
J3.....	.190	.067				−1	−1			
NJ1.....	.106	.362	.326			1	1	1		
NJ2.....	.251	.150	.137	.452		1	1	1	1	
NJ3.....	.339	.201	.232	.590	.267	1	1	1	1	1

by means of the NTSYS computer program (Rohlf 1989).

The various genetic distance matrices were tested against specific design matrices specifying hypotheses concerning the affiliations of the tested populations. Because these methods may not be sufficiently familiar, they are explained in somewhat greater detail below. For illustrative purposes we show in the left portion of table 2 a small genetic distance matrix for three pairs of matched J and NJ populations (actually these are the distances for populations from Yemen, Georgia, and Bulgaria, extracted from table 3 below). The distances can be divided into four classes: J and J (specifically, J1 and J2, J1 and J3, and J2 and J3), J and matched NJ (J1 and NJ1, J2 and NJ2, and J3 and NJ3), J and unmatched NJ (J1 and NJ2, J1 and NJ3, J2 and NJ1, J2 and NJ3, J3 and NJ1, and J3 and NJ2), and NJ and NJ (NJ1 and NJ2, NJ1 and NJ3, and NJ2 and NJ3).

Suppose we wish to test whether distances among J populations are smaller than the other three kinds of distances. To carry out such a test we write down a design matrix (shown in the right portion of table 2). In this matrix the distances being contrasted are designated by arabic ones with opposite signs. It does not matter which distance classes are given positive and negative signs. One then applies a test developed by Mantel (1967; see also Sokal 1979). Corresponding elements of the two half-matrices are multiplied and summed. In our case this would yield $(.190 \times -1) + (.190 \times -1) + (.106 \times 1) + \ldots + (.267 \times 1) = 2.966$. This quantity, called Z, is compared to a distribution of Z values obtained as follows. If there is no difference among any of the four classes of distance (the null hypothesis), we can randomly relabel the rows and columns of the distance matrix and recompute Z. We do this repeatedly (in the results presented below we carried out 999 permutations of the rows and columns of the distance matrices) and if the null hypothesis holds, the observed Z value (for the actual distance matrix) should fall somewhere within the distribution of Z's from the randomized outcomes. If, on the other hand, the null hypothesis is wrong and the alternative hypothesis that the J × J distances are less than the other types of distances is true, then observed Z should be at the high end of the distribution of randomized outcomes. This is easy to understand since the negative small J × J distances do not diminish the large positive distances of the other three types appreciably.

For the example in table 2, only $6! = 720$ permutations are possible for the six rows and columns, so we could not carry out 999 randomizations as was done for the larger matrices reported in the Results section. Ten percent of the randomized Z values are as large as or larger than the observed Z of 2.966. We therefore conclude (on the basis of this small subset of our data) that the J × J distances are not significantly smaller than all other distances in the example matrix.

In actual practice, rather than compute Z as we have shown above, we rescale the distances suitably, so that their Mantel product Z yields an ordinary product-moment correlation r (Smouse et al. 1986). Such correlations are also called matrix correlations (Sneath and Sokal 1973, p. 280) and simply represent the pairwise correlations of corresponding elements of the two matrices. However, their significance is not tested in the conventional way but by the permutational test described above. Program MANTEL3R of the R-package, written by P. Legendre, was employed. Note that the design coefficients need not be ones, but can be weighted to represent specific hypotheses. Also, when a particular class of distances is not involved in a given null hypothesis, these distances are assigned zeros in the design matrix.

For certain hypotheses it is inappropriate to randomly interchange all rows and columns. In such cases, we practice restricted randomization in which

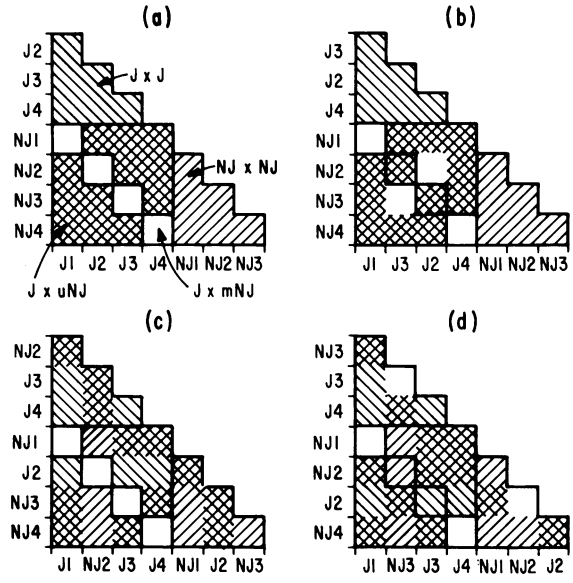

Figure 1 *a*, Schematic diagram of half-matrix of genetic distances for four paired J and NJ populations. Four classes of distances, J × J, NJ × NJ, J × matched NJ, and J × unmatched NJ, are labeled and differentiated by shading. mNJ = matched NJ; uNJ = unmatched NJ. Subsequent panels show the four possible types of interchanges of rows and columns, and changes in the types of distances that appear in the four nominal areas. *b*, J × J interchange, illustrated by J2 × J3. Some nominal J × matched NJ are now J × unmatched NJ, and vice versa. Interchange NJ × NJ is analogous and is not illustrated. *c*, J × matched NJ interchange, illustrated by J2 × NJ2. Now J × unmatched NJ distances appear in the nominal J × J and NJ × NJ matrices and, conversely, J × J and NJ × NJ distances appear among the nominal J × unmatched NJ distances. The J × matched NJ distances remain unchanged. *d*, J × unmatched NJ interchange, illustrated by J2 × NJ3. This results in J × NJ and J × unmatched NJ distances appearing in both nominal J × J and NJ × NJ distances. The nominal J × NJ distances now contain both J × J and NJ × NJ distances. Different randomization procedures permit different combinations of the interchanges illustrated in panels *b–d*.

only specified subsets of rows and columns are permuted. Sokal et al. (1987) developed such an approach in a related context. For didactic reasons the specific designs are described in the Results section together with the test results.

To visualize the effects of randomization more clearly, we show in figure 1*a* a schematic half-matrix of distances, this time for four pairs of matched J and NJ populations. The four classes of distances are labeled and also indicated by different shading. There are four types of interchanges possible during random permutations of the rows and columns of the matrix corresponding to the four classes of distances: J × J, NJ × NJ, J × matched NJ, J × unmatched NJ. Note in figure 1*b–d* that the consequences of these four types of interchanges differ. Permuting all rows and columns permits all four types of interchanges, restricted randomization only some. The differences in permutations make for differences in hypotheses tested even when the design matrix is the same (e.g., design 4 testing hypotheses 4 and 7).

We also quantified the degree of genetic variation within and between J and NJ populations by computing the coefficient of gene differentiation (Nei 1987). For each locus this coefficient is defined as $g_{st} = (h_t - h_s)h_t$, where h_t is the mean heterozygosity in the total set of populations, and h_s is the mean heterozygosity within these populations. Mean heterozygosity, h, per locus is defined as

$$h = 1 - \sum_{i=1}^{m} q_i^2,$$

where q_i is the population frequency of the ith allele at this locus, and m is the number of alleles. The quantity g_{st} obtained for two alleles at a locus is identical to Wright's (1969) F_{st}. However, if a locus contains more than two alleles, g_{st} is the weighted mean of F_{st} for each allele. The average gene diversity for a number of loci is $G_{ST} = (H_T - H_S) H_T$, where H_T and H_S are the arithmetic means of h_t and h_s for all loci under consideration. No bias correction was needed since average sample sizes are much larger than 50 (Nei 1987).

Results

Genetic Affinities among the Populations

Estimates of D between all pairs of J and NJ samples from 12 matched populations based on 10 common systems, as well as of their heterozygosity values, are shown in table 3. Figure 2 shows the UPGMA phenogram for these distances. There are three major clusters. One cluster consists of the J and NJ populations from Yemen. The second comprises most of the populations in the study. It includes all other J populations, as well as all NJ populations from Europe and those from Iraq and Iran. The third cluster comprises the two North African NJ populations. The internal structure of the second cluster is rather difficult to resolve, because the distance values upon which it is based are quite close. Some relationships reflect geographic proximity within either the J or NJ group (e.g., German and Czech NJ, Russian and Polish J); others indi-

Table 3

Nei's Genetic Distance *D* and Average Heterozygosity for 12 Matching J and NJ Populations for 10 Common Systems

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1. Yemen J	.398																							
2. Georgia J	190	.438																						
3. Bulgaria J	190	67	.434																					
4. Turkey J	265	104	60	.466																				
5. Czechoslovakia J	177	86	42	109	.434																			
6. Germany J	148	86	58	62	56	.452																		
7. Russia J	197	84	56	65	80	31	.444																	
8. Poland J	218	58	54	51	84	56	16	.445																
9. Morocco J	119	71	104	137	130	131	141	106	.442															
10. Libya J	258	130	108	135	192	135	93	95	155	.423														
11. Iraq J	278	51	51	54	124	100	60	31	100	95	.455													
12. Iran J	254	86	75	67	111	73	62	60	135	185	58	.456												
13. Yemen NJ	106	362	326	441	304	280	366	408	260	452	447	403	.359											
14. Georgia NJ	251	150	137	95	163	130	172	159	177	190	181	181	452	.433										
15. Bulgaria NJ	339	201	232	146	291	207	130	89	218	161	135	218	590	267	.460									
16. Spain NJ	232	106	117	115	170	108	86	84	148	60	80	120	412	152	146	.451								
17. Czechoslovakia NJ	232	131	115	69	194	98	67	58	157	80	75	111	408	137	78	65	.437							
18. Germany NJ	227	128	128	104	199	111	56	51	144	62	73	135	421	183	60	62	19	.442						
19. Russia NJ	251	64	86	53	124	80	38	21	117	91	45	86	478	109	93	84	58	47	.449					
20. Poland NJ	284	122	168	177	230	163	102	93	157	67	111	212	527	216	126	87	109	53	65	.446				
21. Morocco NJ	483	593	463	520	445	452	386	394	460	467	502	584	502	595	423	461	439	372	447	412	.385			
22. Libya NJ	256	322	269	295	249	267	238	243	240	306	300	377	322	329	291	284	300	247	241	258	161	.431		
23. Iraq NJ	161	62	43	67	100	82	67	45	49	126	32	51	313	170	163	104	91	97	75	155	456	258	.447	
24. Iran NJ	205	100	168	166	229	137	109	106	100	93	111	170	397	210	181	102	122	78	78	43	469	256	115	.444

NOTE:—Average heterozygosity is shown on the diagonal. All genetic distances shown here have been multiplied by 10^4.

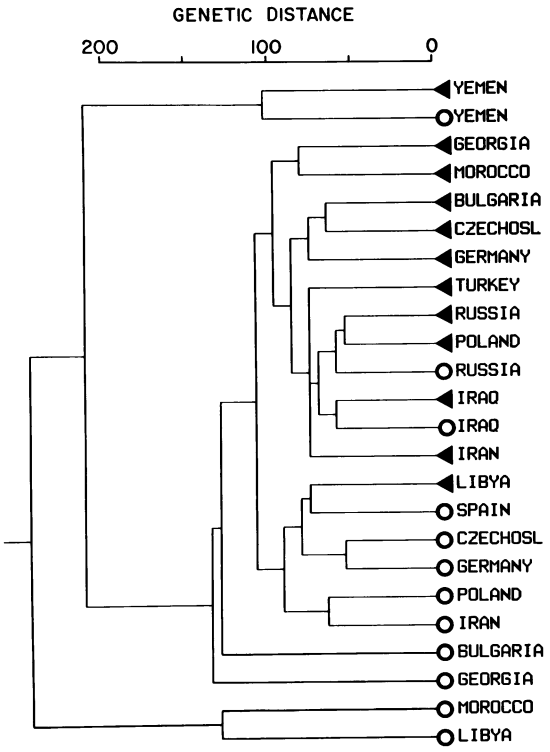

Figure 2 UPGMA phenogram for genetic distances between 12 J and 12 matched NJ populations. Triangles mark J populations, circles mark NJ populations. The abscissa represents D values multiplied by 10^4.

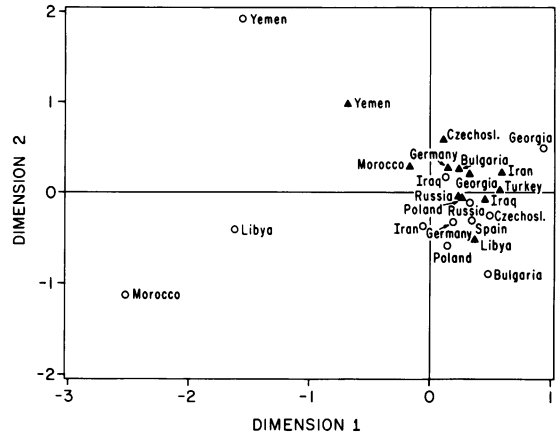

Figure 3 Two-dimensional ordination by nonmetric multidimensional scaling of 12 J and 12 matched NJ populations based on matrix of standard genetic distances (table 3). Triangles mark J populations, circles mark NJ populations.

cate genetic affinity that unites geographically close J with NJ populations (e.g., Iraqi J with NJ). Yet other relationships seem more puzzling, such as the closeness of Moroccan and Georgian J populations and that of the second North African J population (Libya) to the Spanish NJ populations. As a control we assembled a second set of 19 localities with 12 common genetic systems, including the highly informative HLA. The phenogram resulting from the distances based on these data was virtually identical.

However, the obtained phenograms may not necessarily reflect the genetic relationships with respect to the origins of these populations. The genetic distances vary continuously over geographic space, and a hierarchic representation may therefore be inappropriate. The cophenetic correlation coefficients (Sneath and Sokal 1973, p. 278) were only moderately high (r = .86 both times) for the two matrices.

For a better representation of the dissimilarity structure among the 12 matched pairs of J and 12 NJ populations, we proceeded to an ordination approach. The nonmetric multidimensional scaling used both two- and three-dimensional principal axis coordinates. Stress for both types of ordinations was moderate, .183 and .114, respectively. The results of two-dimensional scaling are illustrated in figure 3.

The pattern is reminiscent of that discovered by the UPGMA phenogram (fig. 2), but the similarity among J populations is now better established. There is an obvious compact cluster for the majority of J populations, surrounded by and interspersed with European and Middle Eastern NJ populations, with a more diffuse distribution. All these populations are centrally located within the ordination, with the exception of the NJ community from Bulgaria, whose position is relatively low on axis 2. There are also two pairs of outliers in the graph. One includes the two North African NJ populations located in the left lower quadrant, and the second the two J and NJ Yemenite populations, found in the left upper quadrant.

The position of the Yemenite J sample is of great interest. In contrast to their controversial conclusions with respect to other Jewish populations, earlier studies had agreed that Yemenite Jewry most likely originated from the native population converted to Judaism (Morton et al. 1982; Bonne-Tamir 1985). Kobyliansky et al. (1982) suggest population admixture with a possible effect of convergent selection in the J and NJ populations. The present figure 3 shows the Yemenite NJ population located in the left upper quadrant, well isolated from the rest of the populations, whereas the J sample lies intermediate between it and the cluster of the other J populations.

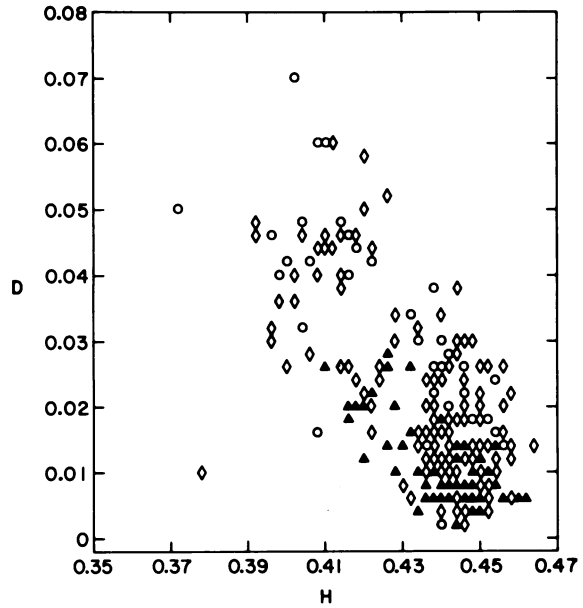

Figure 4 Correlation between average heterozygosity (H) and *D* for 24 J and NJ populations. Triangles indicate J populations; circles indicate NJ populations; diamonds indicate incidences of both types of populations.

Bottleneck Effects on the Genetic Distances in J and NJ Populations

Some of the J founding populations may have been quite small and their gene frequencies may have experienced random changes as a result. Computer simulation studies have shown that when genetic distance estimates are subject to large stochastic errors, UPGMA is superior to other distance matrix methods in the construction of phylogenetic trees (Tateno et al. 1982; Sourdis and Krimbas 1986). Yet, since the use of UPGMA as a technique for phylogenetic estimation assumes a constant rate of evolutionary change for all populations under consideration, if genetic distance estimates are affected by bottlenecks, then the dendrogram constructed from them may significantly deviate from the true phylogenetic tree, even in the absence of population intermixture.

Although the assumption of a constant rate of evolution in all populations is not relevant to an interpretation of multidimensional scaling, the matrix of genetic distances upon which it based might be distorted by the bottlenecks. In an attempt to estimate the extent of the bottleneck effect on the genetic distances in the 12 matched pairs of J and NJ populations, we examined the correlation between genetic distance and mean heterozygosity over all pairs of populations. Genetic distances and heterozygosities in this set of populations are given in table 3. The relationships between genetic distances and heterozygosity values among the 12 J and 12 NJ populations are illustrated in figure 4. The observed correlations are, as expected, negative, separately for J and NJ populations, as well as for the two combined. The degree of correlation is substantial: $-.677$ for J, $-.805$ for NJ, and $-.716$ for the combined data. These correlations are significant at $P < .01$ by Mantel tests. These results support the belief that bottleneck effects were involved in genetic differentiation not only of J populations, but possibly also of the parallel set of NJ populations. Nevertheless, a considerable portion of the variation of the genetic distances cannot be accounted for by variation in the heterozygosity, and factors other than bottleneck effects must have been involved in the evolution of these populations. We may suppose that the higher negative correlation between distance and heterozygosity among NJ populations indicates more profound and prolonged isolation than in the set of J populations. This suggestion is further supported by the analysis of genetic diversities within and between these two sets of populations as discussed in the next section.

Genetic Diversity in J and NJ Populations

The mean genetic distance over the 10 common systems was found to be .0108 (SD = .0064) for 66 possible pairwise combinations between 12 J populations; it is .0191 (SD = .0156) for 144 combinations between 12 J and 12 NJ populations. These quantities indicate a tendency for J populations to be more similar genetically to each other than to NJ populations. The 12 NJ populations showed the highest mean genetic distance over the 66 pairs between them, .0236 (SD = .0127). Jews are also more similar to their NJ neighbors than to other NJ populations. When we divide the distances between J and NJ populations into those between J and geographically matched NJ populations, and those between J and unmatched NJ populations, we obtain respective mean values of .0167 (SD = .0120, n = 12) and .0193 (SD = .0179, n = 132). Significance tests for these differences are reported below as part of the tests of different designs.

Table 4 presents values of the total and within-population genetic diversity, as well as coefficients of gene differentiation between populations by system, for 22 individual systems, and averaged for the 10 common systems for 12 matched J and NJ populations. It can be seen that the h_t values at each genetic system and for the average are very similar in the two

Table 4

Gene Diversity among J and NJ Populations

System	Combined J + NJ			Only J			Only NJ		
	h_s	h_t	g_{st}	h_s	h_t	g_{st}	h_s	h_t	g_{st}
ABH	.4851	.4991	.0280	.4819	.4991	.0345	.4889	.4992	.0205
ABO	.5651	.5734	.0145	.5776	.5859	.0140	.5554	.5634	.0142
Duffy	.4647	.4812	.0341	.4797	.4865	.0141	.4517	.4757	.0505
Kell	.0947	.0958	.0117	.1040	.1049	.0093	.0873	.0884	.0130
Kidd	.4882	.4967	.0171	.4828	.4921	.0188	.4949	.4997	.0097
Luther	.0463	.0476	.0266	.0456	.0463	.0163	.0467	.0481	.0308
MNSs	.7054	.7189	.0187	.7192	.7250	.0080	.6929	.7123	.0272
P	.4806	.4945	.0281	.4816	.4853	.0075	.4798	.4984	.0373
Rhesus	.6540	.6720	.0268	.6419	.6493	.0114	.6639	.6877	.0346
RH (D)	.4336	.4413	.0172	.4155	.4199	.0105	.4563	.4634	.0153
ACP1	.4363	.4484	.0271	.4374	.4505	.0292	.4352	.4464	.0250
ADA	.1797	.1819	.0122	.2002	.2018	.0079	.1664	.1687	.0135
AK1	.0793	.0858	.0076	.0868	.0869	.0104	.0756	.0760	.0046
G6PD	.0897	.1124	.2014	.0865	.1099	.2132	.0916	.1138	.1949
GLO1	.4501	.4572	.0156	.4301	.4365	.0145	.4701	.4780	.0164
GPT1	.4857	.4934	.0155	.4714	.4772	.0122	.4929	.4978	.0100
PGD	.0608	.0625	.0272	.0605	.0622	.0269	.0609	.0627	.0274
PGM1	.3923	.3977	.0137	.4005	.4064	.0156	.3845	.3890	.0119
HPA	.4322	.4411	.0200	.4089	.4132	.0105	.4479	.4567	.0193
TF	.0094	.0095	.0136	.0010	.0010	.0029	.0132	.0134	.0122
HLA-A	.8512	.8679	.0193	.8491	.8691	.0155	.8548	.8707	.0182
HBA-B	.8997	.9282	.0307	.8968	.9163	.0213	.9013	.9304	.0313
KM	.1375	.1469	.0642	.0921	.0945	.0253	.2017	.2149	.0614
	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}	H_S	H_T	G_{ST}
10 joint systems	.4096	.4228	.0312	.4175	.4223	.0130	.4103	.4206	.0245

sets of populations. This occurs despite the considerable variability between loci, which ranges from .0095 (TF) to .9282 (HLA-B) in the combined group of populations. Diversity at all loci for the most part is due to the within-population component. The proportion attributed to the between-population component of diversity varies from 0.76% for AK1 to 20.14% for G6PD, in the combined group. A similar pattern is observed in both the J and NJ populations (table 4). Nevertheless, the G_{ST} value over the 10 jointly studied systems is nearly twice as great among NJ populations as among J populations (2.45% vs. 1.30%). The quantity g_{st} is lower for J populations for 15 of the 22 systems. This difference approaches, but does not quite reach, statistical significance.

Tests of Specific Hypotheses for the Genetic Distances

Nine different hypotheses concerning the relations among the genetic distances are tested in the present section. These hypotheses, enumerated below, are tested by various combinations of six design matrices and three permutational schemes listed in table 5. For hypotheses 1–7 the same analyses were applied to genetic distances calculated independently for the 22 genetic systems and jointly for the 10 systems in the 12 matched population pairs. For hypotheses 8 and 9 only the joint data were analyzed because of computational and design complications. The Rhesus system was analyzed twice, first for the haplotype frequencies of the CcDEe gene complex, and second for the D locus only, because different numbers of population samples were available for these. For HLA, the genetic distances were averaged for the HLA-A and HLA-B loci. The probabilities for the separate systems were combined by Fisher's method (Sokal and Rohlf 1981) to obtain an overall probability of rejection of the null hypothesis in favor of the specified alternative hypothesis. Bonferroni probabilities (Sokal and Rohlf 1987) were first computed for the two Rhesus probabilities.

The results for hypotheses 1–6 are shown in table 6 as correlation coefficients between the distance ma-

Table 5

Six Design Matrices Representing Nine Hypotheses concerning Genetic Affinities of J and NJ Populations

Design Matrix	Genetic Distance between				For Hypothesis Number Below Permute		
	J and J	J and Matched NJ	J and Unmatched NJ	NJ and NJ	All Rows and Columns	J Rows and Columns Only	Matched J vs. NJ
1	−1	1	1	1	1		8
2	1	2	3	3	2		
3	2	1	3	3	3		
4	1	−1	1	1	4	7	
5	0	0	1	0	5		
6	0	0	0	1	6		9

Note.—Hypotheses: 1, only common origin; 2, predominantly common origin; 3, predominantly resemble neighboring NJ; 4, only resemble neighboring NJ; 5, J differ most from nonmatching NJ; 6, NJ differ most from each other; 7, matching J and NJ pairs are closer than unmatched pairs; 8, J pairs closer to each other than NJ or unmatched pairs; 9, NJ pairs farther from each other than J or unmatched pairs.

trices and the appropriate design matrix, together with their significance based on 999 random permutations. These permutations involved all rows and corresponding columns of the matrix, as is conventional in Mantel tests (these are the interchanges shown in fig. 1*b–d*). The signs of the design coefficients were chosen so as to result in positive correlations for the alternative hypothesis of interest. The separate hypotheses are not independent of each other, but each asks a specific question and is therefore of specific interest. In reading through the hypotheses, it may be helpful to refer back to figure 1, which illustrates the randomizations.

Hypothesis 1: Jews Show Only Common Origin.—The ancestral Jewish gene pool is the only factor determining genetic affinities of J and NJ populations. In consequence, the genetic distances between pairs of J populations will always be lower than those between J and NJ populations, regardless of the geographic location of the populations in question. We tested the null hypothesis that distances between pairs of J populations are equal to those between J and NJ populations, or to those between pairs of the latter. The alternative hypothesis is that J populations are less distant from each other than they are from any NJ population or than NJ populations are from each other.

The joint genetic distances based on the 10 systems (in the 12 matched population pairs) are consistent with the hypothesis that Jews show only common origin ($P < .05$), although the degree of (linear) correlation is only .311. Of the 22 tests performed on the separate genetic systems, seven indicate statistically significant departures from the null hypothesis, in support of the hypothesis that Jews show only common origin. Seven loci yield negative correlations, contradicting the hypothesis, but none of them is statistically significant. The overall probability of accepting the null hypothesis based on the outcomes for the separate systems is $P < .005$ ($X^2 = 72.62$). Clearly, the data support this extreme hypothesis of common origin.

Hypothesis 2: Jews Show Predominantly Common Origin.—The ancestral Jewish gene pool makes affinities closest among J populations. Some genetic admixture with neighboring NJ populations yields intermediate affinities with matched NJ populations. Affinities are least between J and unmatched NJ populations (and equally so among NJ populations), which share neither ancestry nor geography. We tested the null hypothesis that there is no correlation between the genetic distances and the design distances implied by the hypothesis that there is predominantly common origin. The alternative hypothesis is positive correlation between the two matrices.

The joint genetic distances based on 10 systems show only trivially higher significant correlation with design matrix 2 than with design matrix 1 ($r = .317$; $P < .05$). The same seven systems as in design matrix 1 exhibit statistically significant departures from the null hypothesis in favor of alternative hypothesis 2. This is not surprising in view of the strong correlation between the two designs. Overall, the probability of accepting the null hypothesis based on the outcomes for the separate systems is $P < .001$ ($X^2 = 79.93$). The hypothesis that Jews show predominantly common origin is also well supported.

Hypothesis 3: Jews Predominantly Resemble Neighboring Non-Jews.—This is similar to hypothesis 2, except that admixture with neighboring populations now leads to

Table 6

Mantel Tests of Seven Hypotheses concerning the Genetic Affinities of J and NJ Populations

Genetic Systems	Chromosome Location	No. of Populations J/Total	Hypothesis 1		2		3		4			5		6	
			r	*P*	*r*	*P*	*r*	*P*	*r*	*P*	P_R	*r*	*P*	*r*	*P*
Blood groups:															
ABH	19p	16/29	−.136	NS	−.145	NS	−.154	NS	−.061	NS	NS	−.054	*	−.111	NS
ABO	9q	23/51	−.007	NS	.003	NS	.031	NS	.065	**	***	.027	NS	−.018	NS
Duffy	1q	22/43	.168	+	.177	+	.178	*	.055	*	***	−.008	NS	.178	*
Kell	?	21/45	.028	NS	.029	NS	.028	NS	.008	NS	NS	.004	NS	.022	NS
Kidd	2p	21/38	−.048	NS	.053	+	−.063	NS	−.038	NS	NS	.060	NS	−.143	NS
Lutheran	19p	7/20	.066	NS	.053	NS	.015	NS	−.037	NS	NS	−.106	NS	.129	NS
MNSs	4q	20/42	.196	*	.200	*	.183	*	.025	NS	NS	−.000	NS	.192	+
P	6p	19/45	.216	**	.214	**	.182	**	−.004	NS	NS	.016	NS	.156	+
Rhesus	1p	23/48	.215	**	.210	**	.173	*	−.025	NS	NS	.009	NS	.171	*
Rh(D)	1p	20/36	.226	*	.221	*	.176	*	−.039	NS	NS	.200	NS	−.005	NS
Enzymes and proteins:															
ACP1	?	18/40	−.108	NS	−.104	NS	−.079	NS	.023	NS	NS	.003	NS	−.091	NS
ADA	20q	19/44	.060	NS	.070	NS	.083	NS	.054	NS	NS	.039	NS	.019	NS
AK1	9q	19/40	−.161	NS	−.159	+	−.132	NS	.010	NS	NS	.102	NS	−.262	NS
G6PD	Xq	13/36	−.172	NS	−.179	NS	−.172	NS	−.046	NS	NS	.077	NS	−.205	NS
GLO1	6p	8/22	.017	NS	.029	NS	.050	NS	.054	NS	+	.277	NS	−.252	NS
GPT1	16?	8/21	.051	NS	.058	NS	.067	NS	.041	NS	NS	.211	NS	−.171	NS
PGD	1p	11/33	.000	NS	.006	NS	.017	NS	.025	NS	NS	−.021	NS	.027	NS
PGM1	1p	21/41	−.068	NS	−.064	NS	−.044	NS	.025	NS	+	.018	NS	−.078	NS
HPA	16q	14/35	.176	*	.177	*	.149	*	.012	NS	NS	.110	NS	.021	NS
TF	3q	10/32	.135	+	.131	+	.098	NS	−.013	NS	NS	−.038	NS	.112	NS
Histocompatibility systems:															
HLA	6p	12/33	.175	*	.184	*	.175	*	.053	+	*	.206	NS	−.078	NS
KM	2p	17/29	.269	*	.271	*	.250	*	.007	NS	NS	.160	NS	.131	NS
10 joint systems ...		12/24	.311	*	.317	*	.256	*	.022	NS	+	.079	NS	.229	+

NOTE.—NS, not significant; +, $.05 < P \leq .10$; *, $.01 < P \leq .05$; **, $.001 < P \leq .01$; ***, $P \leq .001$. For identification of hypotheses, see notes to table 5; for explanation of hypotheses, see text. The correlations for hypothesis 7 are the same as for hypothesis 4. The significance levels for the tests of hypothesis 7 are found in the column marked P_R under hypothesis 4.

closest affinities for J with matched NJ populations. The intermediate affinities are between pairs of J populations and are due to common ancestry, whereas the least affinity is again shown between J and unmatched NJ populations and among NJ populations. We tested the null hypothesis that there is no correlation between the genetic distances and the design distances implied by the hypothesis that Jews predominantly resemble neighboring non-Jews. The alternative hypothesis is positive correlation between the two matrices.

The joint genetic distances based on 10 systems show weaker but significant correlation ($r = .256$; $P < .05$) with design matrix 3. Eight separate systems show significant departures from the null hypothesis in favor of the alternative hypothesis. However, all but one of these are the same systems that tested significant for hypotheses 1 and 2. Presumably they reflect the effects of common origins. The linear correlations between the design matrix and the genetic distance matrices for the separate systems tend to be lower in general than for the two previous hypotheses (table 6). The overall probability of accepting the null hypothesis based on the separate systems is $P < .005$ ($X^2 = 74.55$). Thus there is some support for the hypothesis that Jews predominantly resemble non-Jews, but this may be due to its common-origin component.

Hypothesis 4: Jews Resemble Only Neighboring Non-Jews.—Gene pools of Jewish populations from various geographic regions share no ancestral genes. They are in essence the same populations as the NJ peoples surrounding them, differing from them only in religious and cultural aspects. According to this hypothesis, the genetic distances for Jews and non-Jews from the same country (or geographic region) would always be lower than that between Jews (or non-Jews) of different geographic extractions, or between J and unmatched NJ populations. We tested the null hypothesis that distances between geographically matched J and NJ populations are equal to those between all other combinations of J and NJ populations. The alternative hypothesis is that the matched J and NJ pairs are closer to each other than J and unmatched NJ populations or than distances between pairs of J or of NJ populations. Significances for the results of our tests of hypothesis 4 should be looked up in the column headed P in table 6 (the column headed P_R furnishes significances for hypothesis 7).

The joint genetic distances based on 10 systems lack a significant correlation with design matrix 4. Only two of the separate systems deviate significantly from the null hypothesis. The overall probability of accepting the null hypothesis based on the separate systems is moderately significant at $P < .0405$ ($X^2 = 59.27$), but in view of the tiny correlations and the substantial number of systems (eight) deviating in the direction away from the alternative hypothesis, this cannot be interpreted as support for the hypothesis that Jews resemble only neighboring non-Jews.

Hypothesis 5: Jews Differ Most from Nonmatching Non-Jews.—Because Jewish gene pools reflect both common origin as well as admixture with neighboring populations, the genetic distances between J and NJ populations from different areas should be greater than those between the other possible combinations. This implies that the distance of the ancestral J population from the average NJ population is greater than the diversity among the NJ samples. The null hypothesis tested is that the distances between J and geographically unmatched NJ populations are equal to those between all other combinations of J and NJ populations. The alternative hypothesis is that the unmatched J and NJ pairs are farther from each other than matched J and NJ populations or than distances between pairs of J or of NJ populations.

The joint genetic distances based on 10 systems lack a significant correlation with design matrix 5. Only one of the separate systems is significant, and that is in a direction opposite to that expected under the specified alternative hypothesis. The overall probability of accepting the null hypothesis based on the separate systems is not significant ($X^2 = 20.83$). There is no support for this hypothesis.

Hypothesis 6: Non-Jews Differ Most from Each Other.—If the common origin of the NJ populations employed is more remote than that of the Jews, and if the latter share a common ancestor with one or more of the NJ populations, one would expect the distance matrix of pairs of NJ populations to have higher values than those of any other combination of populations. The null hypothesis tested is that the distances between pairs of NJ populations are equal to those between all other combinations of J and NJ populations. The alternative hypothesis is that the distances of NJ pairs are greater than those of J pairs or of combinations of J and NJ populations.

Although we have seen in an earlier section that the average distance of the NJ pairs is indeed higher than other average distances, neither the joint genetic distances based on 10 systems nor the overall probability for separate systems ($X^2 = 41.4$) is significant. We cannot confirm that non-Jews differ most from each other.

The remaining three hypotheses were tested for sig-

nificance by restricted randomizations of the genetic distance matrix against designs 4, 1, and 6, respectively, as shown in table 5. For hypothesis 7 we permuted only the rows and columns of J populations, leaving the NJ populations undisturbed. This has the effect of rearranging the order of the J populations, so that the J × matched NJ distances are confounded with the J × unmatched NJ distances. See figure 1*b* for an illustration. Applying this procedure to the data results in the same correlations as for hypothesis 4, but in different probability values, which are shown in table 6 in the column labeled P_R.

Hypothesis 7: Matching J and NJ Pairs Are Closer than Unmatched Pairs.—If geography matters, either because of genetic admixture between J and NJ populations inhabiting the same region, or because of adaptations to a common environment, the matched populations should be genetically closer than unmatched ones. The null hypothesis assumes equality of matched and unmatched distances; the alternative hypothesis is that matched distances are closer.

The probabilities in the column labeled P_R under hypothesis 4 in table 6 indicate that when the comparison is only between matched and unmatched populations the results are more significant than when the comparison is matched J and NJ against all others (column *P* under hypothesis 4). The joint genetic distances based on 10 systems are close to significance, three separate systems show significant correlation with the expected sign, and the overall probability of accepting the null hypothesis based on the separate systems is $P < .001$ ($X^2 = 78.20$). The common geography of the matched J and NJ pairs results in lower genetic distances than those of unmatched pairs. Still, however, the linear correlation are remarkably low, the maximum (at locus ABO) being .065.

For hypotheses 8 and 9 we permuted only rows and columns of geographically matched J and NJ populations. This caused matching J and NJ populations to exchange places in the distance matrix, so that at the end of each permutation one is almost always left with a pseudo-J and a pseudo-NJ population, each composed of a mixture of J and NJ samples. Note, however, that by this procedure no matching J and NJ pair can ever end up in the same group. While the J × matched NJ elements in the distance matrix are not disturbed by such a procedure, the other three types (J × J, J × unmatched NJ, and NJ × NJ) are jumbled so that each purported type can contain the other two as well. Figure 1*c* illustrates these points. Since there are only 4,096 possible permutations for the 12 matched populations, we carried out an exhaustive enumeration. Only the joint distances based on 10 systems were treated in this manner.

Hypothesis 8: J Pairs Are Closer to Each Other than NJ Pairs or Unmatched Pairs.—Note the subtle distinction from hypothesis 1, which includes matched pairs in the contrast. The observed correlation of .311 now has a probability $P = .00098$. Pairs of J populations are not only closer than all other combinations (hypothesis 1), but also closer than NJ pairs and unmatched J × NJ pairs (hypothesis 8).

Hypothesis 9: NJ Pairs Are Farther from Each Other than J Pairs or Unmatched Pairs.—Again note the distinction from hypothesis 6. The observed correlation of .229 now has a probability = .03955. Thus while we could not show that NJ pairs are more distant than all other combinations (hypothesis 6), when we exclude matched pairs from the comparison, they are farther apart at a moderate level of statistical significance.

Discussion

Evidence for Common Jewish Origins

During the process of their formation and settlement in a given territory, human populations are subjected to the influences of various evolutionary factors, such as migration, admixture with people from other populations, random differentiation, and specific selective processes. Despite the complicated interactions of these factors in the ethnic groups populating modern Europe, evidence of the effects of single processes, such as directional patterns caused by migration, as well as of stochastic differentiation, may be demonstrated (Sokal et al. 1989).

In the J populations examined in this study, these same factors must be operating as well. In the analyses of the preceding sections we have focused on what the affinities among J populations and those between Jews and their NJ neighbors reveal about the relative magnitudes of two opposing forces: the presumed common origin of modern J populations, and the reciprocal gene flow between J and NJ populations in the same area. Let us consider the evidence from the design matrix tests. Because the separate designs are not independent of each other, a Bonferroni approach must be adopted when evaluating the significance of either the tests of the joint genetic distances based on 10 systems or the overall probabilities obtained from the separate systems. To be significant with a 5% experimentwise type I error rate, each of the 16 tests (2 tests each of the first seven hypotheses, and one test each of hypotheses 8 and 9) would need to be significant at $P \leq .0031$.

By this criterion only one of the correlations for the joint genetic distances (hypothesis 8) is significant, but the overall probabilities based on separate systems are significant for hypotheses 1, 2, 3, and 7.

The significant tests accepting specified alternative hypotheses 1 and 8 (table 6 and text) clearly evidence the greater similarity of J populations to each other than to non-Jews. We also note that pairs of J populations are more similar than pairs of non-Jews. The findings are easiest to explain by deriving the modern J populations from a common original gene pool which underwent relatively few changes during the dispersion of the J people. Also, unless differential evolutionary rates are postulated, the common origin of the J populations must be more recent than that of the non-Jews.

There is evidence also for gene flow between Jews and their neighbors. The significance of hypothesis 7 (and possibly of hypothesis 3) supports geographic affinities between matched J and NJ populations. These affinities suggest gene flow, but, formally at least, convergent adaptation to a common environment cannot be ruled out. The significant outcome of the test for hypothesis 2 is in line with the presumed joint contribution of common origin and admixture. The lack of significance for hypothesis 4, that Jews resemble only neighboring non-Jews, suggests that an exclusively geographic interpretation of the relationships is untenable.

The reality of the gene-flow effect can be demonstrated in another way. When we compute the partial correlation of the genetic distance matrix among the 12 J populations with the matrix of the corresponding NJ populations while keeping a geographic distance matrix for these populations constant, we obtain a value of .440, which is significant at $P < .005$ according to the Smouse-Long-Sokal test (Smouse et al. 1986). This indicates that J population pairs reflect the similarities of their paired NJ neighbors even after allowing for similarity due to common geography. Unless such parallel similarity is induced by convergent selection in a common environment (an unlikely scenario for more than a very few loci), the explanation must be reciprocal gene flow between these populations.

Complications

Our findings that the distances among J populations and their neighbors can most strongly be accounted for by the common origin of the former might lead one to expect the modern genetic structure of J populations to reflect the history of dispersal. Yet although the phenogram of the populations (fig. 2) places all but two of the J populations in a common cluster, the fine structure of this cluster does not mirror the known history of these populations. Nor is this history especially reflected in the ordination results (fig. 3). A further test of this point is a correlation of the genetic distance matrix among J populations with a matrix of separation times (expressed as years ago) of these populations. The correlation is low (.157) and not significant according to the Mantel test. What factors may lie behind this lack of structure within the set of J populations?

Starting out as a single entity (comparable to other ethnic groups of that time), the J people at the beginning of the diaspora were subdivided into numerous populations which became dispersed in the course of dozens of generations to various parts of the world. Within each country in which they came to reside the Jews usually became nonintermarrying subpopulations. Consequently, a significant contribution of stochastic differentiation between populations and inbreeding within populations could be expected. The reduction of effective population size in such a set of populations may lead to a rapid increase of genetic distances between them, primarily due to the decrease of heterozygosity in each of these populations (Chakraborty and Nei 1977). The effect of a bottleneck on average heterozygosity may last for hundreds or even thousands of generations after the recovery of population size (Nei 1987). Thus estimates of the descent and branching relationships of populations may be seriously distorted by bottlenecks.

One way to study this problem is to examine the relationship between genetic distance and the average heterozygosity over pairs of populations (Livshits and Nei 1990). The negative correlation between genetic distance and heterozygosity of pairs of populations would be expected on condition that the populations were derived from the same ancestral stock and at about the same historical time. We also have to assume absence of migration, as well as of selection after separation.

Despite the fact that human populations can hardly satisfy the above conditions, especially regarding the total absence of migration, genetic distance values are strongly negatively correlated with heterozygosity estimates. Statistically significant negative correlations between genetic distance and heterozygosity were found in separate studies of caucasoid, Amerindian, and Far Eastern mongoloid populations by Livshits

and Nei (1990). Our results in this study are in agreement with the above findings. Therefore stochastic factors are likely to have played a role in masking the descent relationships of the J populations. Note, however, that when populations representative of different major races were combined, Livshits and Nei (1990) found the correlation to be almost negligible, implying that the time after splitting of the human species into the major races was long enough to restore equal heterozygosity in each stock.

The complex structure of the cluster of J populations must also be due to differences in the amount of admixture with neighboring NJ populations, because of both different lengths of contact and different rates of admixture. The first of these is obvious. The second is due to marked social and cultural differences among J communities living in different countries (see Deshen and Zenner 1982). These populations would be subject to different genetic consequences as a result of different interactions with their neighbors. By way of an example, whereas the Kurdish Jews were loosely knit, with a poorly developed sociocultural life and consequently considerable interactions with their NJ neighbors, the Jews of Yemen formed a socially homogeneous isolate in more recent times (although as noted above, they may have had a substantial component due to conversion in their early history).

These complicated relationships indicate that application of regular population genetic models to what is an aggregate of J populations may be inappropriate. This is in part the reason why we chose the nonparametric approaches detailed above for the analysis of the main trends in the genetic relationships between these populations and their corresponding NJ neighbors.

We have studied the problem of J affinities using genetic distances based on gene-frequency data for as many loci as were available, since genetic distances and average heterozygosities for single loci are subject to large stochastic errors (Nei 1987). However, useful information can also be obtained from a consideration of each locus singly. Inspection of table 6 shows that hypothesis 1, postulating a common origin of J populations, is supported by MNSs, P, Rh, HPA, HLA, and KM systems. Although stochastic events will affect all loci, the magnitude of the effect on each locus is random and depends on the gene frequency and effective population size (Wright 1969). We must assume that those systems not differentiating J from NJ populations drifted apart sufficiently to match the J–NJ distances. By contrast, consistent support for a geographic pattern, and hence admixture, is provided by only Duffy and HLA. Although admixture will also affect all loci, its effects will be noticeable only in those cases where initial large gene-frequency differences separate J from neighboring NJ populations. Since by chance it is unlikely that many loci will exhibit such differences and since admixture seems to have been a lesser factor in the shaping of the modern J gene pool, it is not surprising that we find only few loci supporting gene flow.

Our findings that J populations reflect their common origins are in agreement with the results supported by the majority of previous workers on the problems (e.g., Carmelli and Cavalli-Sforza 1979; Karlin et al. 1979; Kobyliansky et al. 1982; Kobyliansky and Livshits 1983; Rao and Boudreau 1984). Our results disagree with those of Morton et al. (1982), which, however, were based on far fewer genetic systems. Genetic admixture with neighboring populations seems to be a secondary effect (except possibly for the Yemenites). In this respect we again differ with the results reported by Morton et al. (1982) and agree more with findings obtained by different techniques by Carmelli and Cavalli-Sforza (1979), Wijsman (1984), and Motulsky (1980). However, we should point out that we cannot show detailed correspondence with Motulsky's study, the only one which reports differential rates for different loci. Motulsky's most "admixed" loci are PGM, Kell, ABO, G6PD, and Gm. By contrast, we find the highest correlations against design matrix 4, the most direct test for admixture, for loci ABO, Duffy, GLO1, and HLA. Only one locus, ABO, is common in these lists. Since Motulsky estimated admixture for Ashkenazi populations only and did not specify the population samples employed explicitly we are unable to resolve this contradiction.

Acknowledgments

This is contribution 780 in Ecology and Evolution from the State University of New York at Stony Brook. We are indebted to Barbara Thomson for computational assistance. Donna DiGiovanni and Cynthia Kurtz wordprocessed the manuscript and Joyce Schirmer prepared the illustrations. Drs. Neal Oden and Lev Ginzburg provided useful suggestions for data analysis. This research was carried out with the support of grant GM 28262 from the National Institutes of Health to R.R.S., and sabbatical leave funds from Tel Aviv University to G.L.

References

Bonne-Tamir B (1985) Oriental Jewish communities and their genetic relationships with S.W. Asian populations. In: Ahuja YR, Neel JV (eds) Genetic microdifferentiation in human and other animal populations. Indian Anthropologist, Occasional Papers in Anthropology no 1. Paragon, New Delhi, pp 153–170

Camin JH, Sokal RR (1965) A method for deducing branching sequences in phylogeny. Evolution 19:311–326

Carmelli D, Cavalli-Sforza LL (1979) The genetic origin of the Jews: a multivariate approach. Hum Biol 51:41–61

Cavalli-Sforza LL, Carmelli D (1979) The Ashkenazi gene pool interpretation. In: Goodman RM, Motulsky AG (eds) Genetic diseases among Ashkenazi Jews. Raven, New York, pp 93–102

Chakraborty R, Nei M (1977) Bottleneck effects on average heterozygosity and genetic distance with the stepwise mutation model. Evolution 31:347–356

Cotterman CW (1953) Regular two-allele and three-allele phenotype systems. I. Am J Hum Genet 5:193–235

Deshen S, Zenner WP (eds) (1982) Jewish societies in the Middle East: community, culture and authority. University Press of America, Washington, DC

Fuerst PA (1985) Evolutionary differentiation and the sharing of alleles between populations. In: Ahuja YR, Neel JV (eds) Genetic microdifferentiation in human and other animal populations. Indian Anthropologist, Occasional Papers in Anthropology no 1. Paragon, New Delhi, pp 15–30

Hedrick PW, Thomson G, Klitz W (1986) Evolutionary genetics: HLA as an exemplary system. In: Karlin S, Nevo E (eds) Evolutionary process and theory. Academic Press, New York, pp 583–606

Karlin S, Kenett R, Bonne-Tamir B (1979) Analysis of biochemical data on Jewish populations. II. Results and interpretations of heterogeneity indices and distance measures with respect to standards. Am J Hum Genet 31:341–365

Kobyliansky E, Livshits G (1983) Genetic composition of Jewish populations: diversity and inbreeding. Ann Hum Biol 10:453–464

——— (1989) Age-dependent changes in morphometric and biochemical traits. Ann Hum Biol 16:237–247

Kobyliansky E, Micle S, Goldschmidt-Nathan M, Arensburg B, Nathan H (1982) Jewish populations of the world: genetic likeness and differences. Ann Hum Biol 9:1–34

Li W-H, Nei M (1977) Persistence of common alleles in two related populations or species. Genetics 86:901–914

Livshits G, Nei M (1990) Relationship between intrapopulational and interpopulational genetic diversity in man. Ann Hum Biol 17:501–513

Mantel M (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Mayr E (1965) Numerical phenetics and taxonomic theory. Syst Zool 14:73–97

Morton NE, Kenett R, Lee S, Lew R (1982) Bioassay of kinship in populations of Middle Eastern origin and controls. Curr Anthropol 23:157–166

Motulsky AG (1980) Ashkenazi Jewish gene pools: admixture, drift and selection. In: Eriksson AW, Forsius HR, Nezanlinna HR, Workman PL, Norio RK (eds) Population structure and disorders. Academic Press, New York, pp 363–365

Mourant AE, Kopéc AC, Domaniewska-Sobczak K (1978) The genetics of the Jews. Oxford University Press, Oxford

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Rao CD, Boudreau R (1984) Diversity and cluster analyses of blood group data on some human populations. In: Chakravarti A (ed) Human population genetics: the Pittsburgh symposium. Van Nostrand Reinhold, New York, pp 331–362

Rohlf FJ (1989) Numerical taxonomy system of multivariate statistical programs. Technical report, Department of Ecology and Evolution, State University of New York at Stony Brook, Stony Brook

Smouse PE, Long JC, Sokal RR (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Syst Zool 35:627–632

Sneath PHE, Sokal RR (1973) Numerical taxonomy. W. H. Freeman, San Francisco

Sokal RR (1979) Testing statistical significance of geographic variation patterns. Syst Zool 28:227–232

Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthrop 80:267–294

Sokal RR, Lengyel IA, Derish PA, Wooten MC, Oden NL (1987) Spatial autocorrelation of ABO serotypes in mediaeval cemeteries as an indicator of ethnic and familial structure. J Archeol Sci 14:615–633

Sokal RR, Rohlf FJ (1981) Biometry, 2d ed. W. H. Freeman, New York

——— (1987) Introduction to biostatistics, 2d ed. W. H. Freeman, New York

Sourdis J, Krimbas C (1986) Accuracy of phylogenetic trees estimated from DNA sequence data. Mol Biol Evol 4: 159–66

Tateno Y, Nei M, Tajima F (1982) Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. J Mol Evol 18:397–404

Wijsman EM (1984) Techniques for estimating genetic admixture and applications to the problem of the origin of the Icelanders and the Ashkenazi Jews. Hum Genet 67: 441–448

Wright S (1969) Evolution and the genetics of populations, vol 2. University of Chicago Press, Chicago