Am J Hum Genet 31:341–365, 1979

Analysis of Biochemical Genetic Data on Jewish Populations: II. Results and Interpretations of Heterogeneity Indices and Distance Measures with Respect to Standards

SAMUEL KARLIN,[1] RON KENETT,[2] AND BATSHEVA BONNÉ-TAMIR[3]

SUMMARY

A nonparametric statistical methodology is used for the analysis of biochemical frequency data observed on a series of nine Jewish and six non-Jewish populations. Two categories of statistics are used: heterogeneity indices and various distance measures with respect to a standard. The latter are more discriminating in exploiting historical, geographical and culturally relevant information. A number of partial orderings and distance relationships among the populations are determined. Our concern in this study is to analyze similarities and differences among the Jewish populations, in terms of the gene frequency distributions for a number of genetic markers. Typical questions discussed are as follows: These Jewish populations differ in certain morphological and anthropometric traits. Are there corresponding differences in biochemical genetic constitution? How can we assess the extent of heterogeneity between and within groupings? Which class of markers (blood typings or protein loci) discriminates better among the separate populations? The results are quite surprising. For example, we found the Ashkenazi, Sephardi and Iraqi Jewish populations to be consistently close in genetic constitution and distant from all the other populations, namely the Yemenite and Cochin Jews, the Arabs, and the non-Jewish German and Russian populations. We found the Polish Jewish community the most heterogeneous among all Jewish populations. The blood loci discriminate better than the protein loci. A number of possible interpretations and hypotheses for these and other results are offered. The method devised for this analysis should prove useful in studying similarities and differences for other groups of populations for which substantial biochemical polymorphic data are available.

Received July 24, 1978; revised November 28, 1978.

This study was supported in part by grant GM 10452-15 from the National Institutes of Health, and grant MCS76-80624-A01 from the National Science Foundation.

[1] Department of Mathematics, Stanford University, Stanford, California 94305.

[2] Department of Mathematics, The Weizmann Institute of Science, Rehovot, Israel.

[3] Department of Human Genetics, Sackler School of Medicine, Tel-Aviv University, Ramat-Aviv, Israel.

INTRODUCTION

In this paper we present a nonparametric methodology for the analysis of gene frequency data observed in a series of populations. Various groupings of the populations by historical, geographical, cultural, and environmental criteria are considered in order to assess relationships within and between them. A hierarchy of statistics, including heterogeneity and distance measures relative to a standard, are proposed in order to infer orderings for appropriate population groupings.

This study concentrates on the analysis of biochemical genetic frequency data collected on nine Jewish and six non-Jewish populations. The specific populations and corresponding sample sizes are listed in table 1.

The data consist of allele frequencies at the 14 loci in table 2. The eight blood markers and six protein loci are all polymorphic to at least 1%; actually, in all cases, the second most frequent allele exceeded 5%. Observations on six other enzyme loci were discarded, since each was monomorphic for the same allele throughout the population range. Table 2 shows the numbers of alleles of the loci with respect to the aggregate population.

The loci are mostly located on distinct chromosomes which, in conjunction with other population studies pertaining to blood groups and protein markers, portends that the allelic variants of these loci likely segregate independently. The only exception is the two loci in the HLA complex, but because of their manifold allelism, again the assumption of loci independence is quite adequate.

The listings of table 1 conform to the "historical"-geographical identifications of Ashkenazi, Sephardi, and Oriental Jewish groupings. The three groups have many

TABLE 1

POPULATIONS STUDIED AND SAMPLE SIZES

Population	No. in sample
Jewish populations:	
Ashkenazi (A)—European Jews	
Polish	126
Russian	123
German	73
Rumanian	130
Sephardi (S)—North African Jews	
Moroccan	141
Libyan	189
Oriental (O)—Asian Jews	
Iraqi	123
Yemenite	197
Cochin	184
Non-Jewish populations:	
Middle Eastern	
Arabs	100
Armenians	90
Samaritans	114
European	
German	462
Polish	400
Russian	126

TABLE 2

GENETIC MARKERS

Markers	Numbers of alleles of loci, with respect to the aggregate population
Blood Groups:	
HLA-A	10
HLA-B	16
ABO	5
MNSs	4
Rh	7
P	2
Duffy	3
Kell	2
Proteins:	
Acid phosphatase	3
Adenylate kinase	2
Adenosine deaminase	2
Phosphoglucomutase	2
6-Phosphogluconate dehydrogenase	2
Haptoglobin	2

different characteristics. The Ashkenazi populations of relatively large sizes have enjoyed many contacts with one another and, to some extent, with neighboring non-Jews. The Sephardi populations are also large but have been more separated. The Moroccan and Libyan Jews constitute only two representatives of the Sephardi groups. No complete data are available as yet on other populations in this grouping (e.g., Turkish, Greek, and Egyptian Jews), but we surmise the results achieved may apply broadly. The Oriental group is the most heterogeneous; the Yemenite and Cochin Jews, in particular, were small, isolated communities with significant inbreeding. Some pertinent history on the specific Jewish populations of this study with respect to migrations, persecutions, conversions, demographic variations, degree of consanguinity, intermarriage, life-styles, behavior norms, health practices, social institutions, etc., is outlined in a companion paper in this issue [1]; see also references cited therein. Relevant information on the nature of table 1 is also provided [1].

This study analyzes similarities and differences in the Jewish populations and between the Jewish and the non-Jewish populations in terms of the frequency distributions for a number of genetic markers. Some questions discussed are as follows: (1) These Jewish populations differ in certain morphological and anthropometric traits. Are there corresponding differences in biochemical genetic constitution? (2) How can we assess the extent of heterogeneity between and within the populations? (3) How valid and meaningful with respect to biochemical traits are the historical-geographical groupings? (4) Which class of markers (blood typings or protein loci) discriminates better between the populations? (5) What can we learn about the degree of admixture between them? (6) How can we assess the extent, form, and relative influence of selection, drift, population structure, and migration? (7) There is a contention that the HLA system, because of its large multiple allelism, is a good basis for classifying population relationships. Is this claim valid in the present context?

On a more local level we may note other questions of interest: (8) Are Iraqi Jews "closer" to Libyan Jews than to Russian Jews? (9) Are German Jews "closer" to Russian Jews than to German non-Jews? This last question is related to the familiar debate over the extent to which a Jewish minority living in an area for several generations resembles the non-Jewish majority of that area. Further problems along these lines will be taken up.

Until recently, such questions have typically been pursued in the framework of physical anthropological characteristics (e.g., bone structure, facial features, skin texture, dermatoglyphics) or in terms of a few Mendelian genetic traits [2]. Comparisons are commonly made one locus at a time and often yield contradictory results. There are problems in combining measurements from separate polymorphic loci which may involve dependencies among loci and correlations among alleles at a locus. How to represent and compare vectors of different dimensions in a valid manner is not obvious. Some authors [3–7] lay out separate loci frequency vectors as one extended array and proceed with deductions based on multivariate normal distribution theory, using variants of linear discriminant functions. Others [8–12] calculate some distance measure and then employ a clustering procedure.

It is generally thought that genetic distances used to compare populations serve equally well to discriminate between populations [5, 7, 9, 12]. Most techniques simply add distances over loci to obtain the total distance between two populations. Ours does not, since it is unclear both how to weight different loci with different numbers of alleles, and how to weight intralocus alleles.

We describe our methodology first in qualitative terms. Suppose we wish to relate two populations $P^{(1)}$ and $P^{(2)}$ with respect to their allelic frequency arrays for a set of polymorphic genetic markers. The idea is to compute a "distance" (or set of distances) to a standard for $P^{(1)}$ and $P^{(2)}$ at each marker and compare the distances. We say that one population, $P^{(1)}$ or $P^{(2)}$, is *closer to the standard* than the other population at the 5% significance level, if its distance from the standard is smaller than the other population's for at least 11 of the 14 markers measured. If neither population is closer to the standard 11 times or more, we say that the two populations cannot be ordered (are "not separable") relative to the standard. (For α and n different from 5% of 14, an appropriate n^* is calculated.)

We employ four classes of standards in order to assess different forms and levels of relatedness among the populations of table 1. These are: (a) absolute standards; (b) the "world Jewish average" standard; (c) standards based on the three main geographical-historical groupings of Jews; and (d) each individual population as a standard.

When the standard to which the separate populations are compared is an absolute set of frequencies (a), we refer to the associated distances as heterogeneity indices. For the Rh locus with seven alleles, the heterogeneity indices effectively measure distance to the central frequency state, one-seventh for each allele. Thus a population with a larger heterogeneity index has a relatively more central allelic frequency distribution. The distances of the separate Jewish population allele frequency vectors from the pooled world Jewish standard (b) enable us to assess the extent of homogeneity among the various Jewish populations. Distances of separate Jewish populations from the three group standards (c) offer insights into the meaningfulness of these groupings.

The allele frequency vectors of each individual population (d) provide standards in order to determine the relative closeness of the other populations to the specified population. These latter comparisons are pertinent to questions (8) and (9) highlighted earlier.

METHODS

At each locus, we have available for each population its allele frequency array, which we wish to compare or contrast across the separate populations.

We will deal with two categories of statistics: heterogeneity indices, and distance measures with respect to a standard. The latter are more discriminating as they allow greater flexibility in exploiting relevant historical, geographical, and cultural information.

Heterogeneity indices

Let $x = (x_1, \ldots, x_N)$ = frequency vector;

$$\sum_{i=1}^{N} x_i = 1 \ .$$

At each locus N refers to the number of all possible alleles observed for the aggregate sample of Table 1.

Consider

$$f(x,\alpha) = \frac{1 - \sum_{i=1}^{N} x_i^{\alpha+1}}{\alpha} \ ; \ \alpha \geq 0 \text{ for which we highlight the special choices,} \qquad (1)$$

$$f(x,0) = -\sum_{i=1}^{N} x_i \log x_i \text{ the } \textit{information number } (\alpha = 0), \text{ and} \qquad (2a)$$

$$f(x,1) = 1 - \sum_{i=1}^{N} x_i^2 \text{ the } \textit{heterozygosity measure } (\alpha = 1). \qquad (2b)$$

The functions $f(x,\alpha)$ are maximal in a population where all alleles occur at equal frequencies and zero when the population is fixed on one allele. Of course, with more alleles (N increasing) the values in equations (2a) and (2b) tend to increase. These indices reflect estimates of heterogeneity for each allele frequency vector. Concomitantly, such functionals tend to be less sensitive for discrimination purposes as will be evident in the results expounded later.

Distance measures referring to an appropriate standard

The concept is as follows. For each locus (say i) we prescribe an allele frequency vector as a standard

$$(s_1^{(i)}, s_2^{(i)}, \ldots, s_{m_i}^{(i)}), \ i = 1,2, \ldots, 14 \ . \qquad (3)$$

m_i denotes at locus i the number of *all* possible alleles observed in the aggregate sample of table 1. (We do not exclude the contingency that the standard has some zero components.) An associated distance statistic, relative to the standard, is an expression of the form,

$$u_{(i,j)}(s) = \sum_{k=1}^{m_i} w(s_k^{(i)}) f(|x_{k,j}^{(i)} - s_k^{(i)}|) \ , \tag{4}$$

where $w(s_k^{(i)})$ are weights assigned to the allele frequencies in the standard, and the deviation function $f(|x_{k,j}^{(i)} - s_k^{(i)}|)$ calculates an extent of departure from the standard for the kth allele frequency of the jth population at locus i. Different choices of weights will place more or less emphasis on the rare or abundant alleles relative to the standard. We mostly take

$$f(u-v) = |u-v|^p \ , \tag{5}$$

and then the statistics of equation (4) become

$$\sum_{k=1}^{m_i} w(s_k^{(i)}) |x_{k,j}^{(i)} - s_k^{(i)}|^p \ . \tag{6}$$

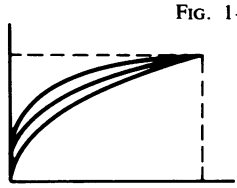
In order to assure robustness with respect to the conclusions emanating from our analysis, we will execute the computations for a wide spectrum of choices of weights coupled with the deviation functions of the form (5) for $p = 2,1,½$ (i.e., the squared, absolute, and square root deviation calculations which emphasize differently the small, against moderate to large deviations). By robust results we mean qualitative results little affected by the use of different weights and deviation functions.

A hierarchy of weight functions and some motivations for their interest are indicated in figure 1.

The versatility in the figure 1 weight functions place different emphasis on large,

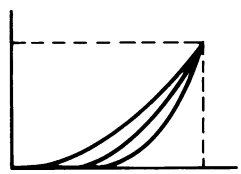
FIG. 1—Weight Functions.

A. $s^{½}$ $s^{¼}$ $s^{⅛}$



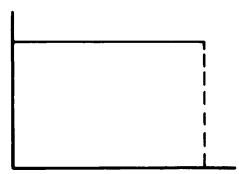
A.—These weight functions share the property that they increase to various degrees the relative contribution of the alleles at the standard with small frequency as against those with high frequency.

B. s^2 s^4 s^8



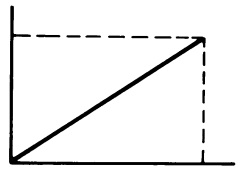
B.—The small values of s are much reduced in importance relative to the large values. For practical purposes, with s^4 and s^8, we are ignoring the alleles occurring with small frequency.

C. $s^0 \equiv 1$



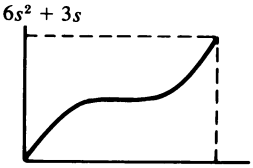
C.—This weight prescription assigns equal value to all admissible values of s.

D. s



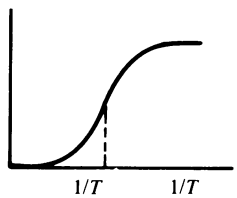
D.—The weightings are directly proportional to the allele frequency of the standard.

E. $Y = 4s^3 - 6s^2 + 3s$



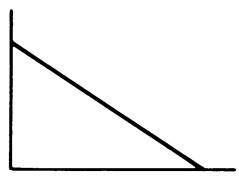
E.—This cubic increases the relative importance of small values of s and reduces the relative contribution of large values of s.

F. $V(s) = \dfrac{sQ'(s)}{nQ(s)}$; $Q(s) = \sum_{i=1}^{n} (Ts)^i$



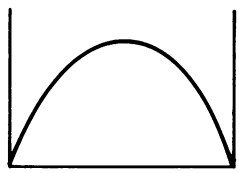
F.—For large $n(> 50)$ this function produces a threshold at $s = 1/T$ changing from 0 to 1 quite quickly. We use a less abrupt threshold effect as induced by $n = 10$; $1/T = .1,.2,.3$; One could also use a logistic function $1/(1 + ke^{-\lambda s})$.

G. $1-s$



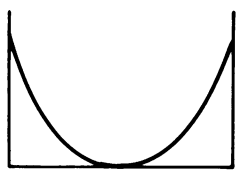
G.—Strongly weights small frequencies of the standard and virtually ignores high frequencies.

H. $s(1-s)$



H.—Emphasizes intermediate frequency values of the standard.

J. $s^2 - s + 1$



J.—An opposite oriented weighting to that of the previous example. This is similar to the weight function $1/s(1 - s)$, but does not suffer from infinities of zero frequencies.

medium and/or small allele frequencies of the standards. The aim is to extract results robust with respect to the different weight and deviation functions.

The determination of significant orderings among the populations

All the preceding heterogeneity indices and/or distance measures provide us with different facets of the variability inherent to each locus. We now combine the information endowed to the various measures for the 14 loci.

The analysis consists of five steps;

1. Choose a statistic; either one of the heterogeneity indices or a distance measure from a standard of the type in equation (6).

2. Compute for every population at each locus the statistic values.

3. Compare the statistic values at the 14 loci for each pair of populations. The assumption of loci independence permits the use of the *sign test* to uncover an ordering, where one exists, for the two populations under consideration. (Since the loci differ in the number of alleles, applying a rank order test, Wilcoxon or otherwise is inappropriate.)

The frequency vectors for the different loci are of varying dimensions, and undoubtedly the underlying distributions for the individual loci measurements differ. On this basis, it is natural to invoke the sign test with $n = 14$ number of paired comparisons in ordering the populations.

4. If for the pair of populations P_1 and P_2, the statistics $u_{i1}(s)$ of P_1 exceeds $u_{i2}(s)$ of P_2 (or vice versa), 11 times or more, then the sign test guarantees at a better than 5% significance level that one of these populations is "larger" or "smaller" as measured by the particular statistic used over the second population.

5. With each statistic, the above calculations and comparisons are made for every pair of populations, and a *partial ordering* among the populations is thereby ascertained.

RESULTS

We obtain a partial ordering of the populations by invoking steps 1–5 of the procedure for each specified statistic. Significant comparisons over the 14 loci permit us to decide which of two populations is larger, or whether they are not separable. When there is no significant difference, the populations are displayed in the same column, as in tables 3–5.

We start by describing in table 3 the realizations of the comparisons for the heterogeneity indices of equations (2a) and (2b).

The place in the ordering for Rumanian Jews is based on blood group data only (since other data were not available); inferences in this respect should be regarded as tentative. We have complete data on the Samaritan, Arab, and Armenian populations, so that their relative position on the heterogeneity scale is discernible. The populations at the extreme right are those endowed with more uniform allele frequency distributions; those at the extreme left bear relatively few segregating alleles per locus.

Table 3 shows a relatively increased heterogeneity value for the Ashkenazi group, while showing no significant differences within this group. The Oriental group was at the other extreme; the Iraqi Jews fell in with the Ashkenazi group at the top of the

TABLE 3

MEASURES OF HETEROGENEITY

Information number (2a) →			
Cochin J. Samaritans	Yemenite J. Libyan J.	Arabs Armenians	Iraqi J. Moroccan J. German J. Polish J. Russian J. (Rumanian J.)*
Heterozygosity statistic (2b) →			
Samaritans	Yemenite J. Cochin J.	Moroccan J. Libyan J. Arabs Armenians	Iraqi J. German J. Polish J. Russian J. (Rumanian J.)*

NOTE.—Increasing values of the statistic from left to right. Populations in same column are not separable. Populations in any column are mostly "larger," with respect to the statistic, than populations in a column to the left.

* Based on blood group data only.

scale, whereas the Yemenite and Cochin populations showed a strongly noncentral genetic composition, yielding the smallest information and heterozygosity value across the 14 loci. The Sephardic populations lie between the Ashkenazi and Oriental populations, but are strongly tilted toward the Ashkenazi, as will be evident from tables 4 and 5.

We now present results obtained with the various distances defined in equation (6). In table 4 we take as our standard the pooled world-Jewish allele frequency arrays, with different weightings and deviation functions. In table 5 we consider the standards as the allele frequency vectors of the historical-geographic groups, with the distance for each population referred to its own group standard. Since the sample sizes of the Jewish populations were quite similar, the world-Jewish and historical-geographic group standards were not dominated by any particular population.

Discussion and comments on tables 4 and 5

Juxtaposing tables 4 and 5 reveals that the use of historical-geographic group averages as standards establishes a much finer ordering among the populations than we obtained in table 3.

The Iraqi Jews are again "closer" in genetic composition to the Ashkenazi and Moroccan populations than to the other Oriental populations. The closeness between the Iraqi Jews and the Ashkenazi and Moroccan populations remains an enigma and needs more study.

The Yemenite and Cochin Jews exhibit non-uniform allele distributions and much divergence with respect to each other, and since they also differ substantially from the world-Jewish standard, our Oriental group seems not to be properly a group at all. An important factor is presumably that the Yemenite and Cochin populations were composed of small, isolated, inbred groups.

TABLE 4

ORDERINGS INDUCED BY DISTANCES WITH "WORLD"-JEWISH FREQUENCY VECTOR USED AS STANDARD

$$\text{Distance} = \sum_k w(S_k^{(i)})(x_{k,j} - S_k^{(i)})^2$$

Weight function	Population groups		
$w(s) = 4s^3 - 6s^2 + 3s$	Polish J.	Iraqi J. German J. Russian J. (Rumanian J.) Moroccan J. Libyan J.	Yemenite J. Cochin J.
$w(s) = s^{\frac{1}{2}}$	Polish J.	Iraqi J. German J. Russian J. (Rumanian J.) Moroccan J.	Libyan J.	Yemenite J. Cochin J.
$w(s) = s^0 \equiv 1$	Polish J.	Iraqi J. German J. Russian J. (Rumanian J.) Moroccan J.	Libyan J.	Yemenite J. Cochin J.
$w(s) = s^2$	Polish J. Russian J. (Rumanian J.)	Iraqi J. German J. Moroccan J. Libyan J.	Yemenite J.	Cochin J.

NOTE.—Results for four typical statistics. Results were exceptionally robust and consistent with minor adjustment in the orderings when considering a wide spectrum of 27 statistics constructed using the p values of 2, 1 or 1/2 in equation (6), and independently, the 9 variations of the weightings of figure 1.

The Ashkenazi populations proved both relatively heterogeneous in allele frequency distributions and rather homogeneous in terms of distance from their own historical group standard. A striking finding, however, is the distinctive position of the Polish Jews. They were significantly more heterogeneous than the Russian and German Jews, basically exhibiting the most "uniform" (i.e., most central) allele distributions over most of the loci. This outcome conforms with the fact that the Polish Jews registered the smallest distances from the world-Jewish and Ashkenazi standards. In another vernacular, we can characterize the Polish Jews as the Jewish population "closest" to the average "world"-Jewish population, and to the average Ashkenazi population.

The Moroccan Jews are closely akin to the Ashkenazi Jews in genetic composition. Some possible factors bearing on this matter are discussed in the Discussion and Conclusions section.

The Libyan Jews are somewhat less homogeneous than the Moroccan Jews, and show somewhat greater differences from the Sephardi group and world-Jewish standards. As with the Yemenite and Cochin populations, we may be partly seeing the consequences of small effective population size.

Two reassuring features of tables 4 and 5 are the robustness of the comparisons for all the different weight functions chosen, and the overwhelming prevalence of significant orderings. It seems clear that the orderings of these tables are real and not spurious.

TABLE 5

ORDERINGS INDUCED BY DISTANCES WITH HISTORICAL-GEOGRAPHICAL GROUP AVERAGES AS STANDARDS

$$\text{Distance} = \sum_k w(g_{k,\,j}^{(i)})\,(x_{k,\,j} - g_{k,\,j}^{(i)})^{2*}$$

Weight function	Population groups					
$w(s) = 4s^3 - 6s^2 + 3s$	(A) Polish J.	(A) German J. Russian J. (Rumanian J.)	(S) Moroccan J.	(S) Libyan J.	(O) Yemenite J.	(O) Cochin J. Iraqi J.
$w(s) = s^{\frac{1}{2}}$	(A) Polish J.	(A) German J. Russian J. (Rumanian J.)	(S) Moroccan J.	(S) Libyan J.	(O) Yemenite J.	(O) Cochin J. Iraqi J.
$w(s) = s^0 = 1$	(A) Polish J.	Moroccan J. (S) German J. Russian J. (Rumanian J.) (A)	(S) Libyan J.	(O) Yemenite J.	(O) Cochin J. Iraqi J.	
$w(s) = s^2$	(A) Polish J.	(A) German J. Russian J. (Rumanian J.)	(S) Moroccan J.	(S) Libyan J.	(O) Yemenite J. Cochin J.	(O) Iraqi J.

NOTE. — The computation of each statistic for a population referred to the standard of its own historical-geographical group as prescribed in table 1. Almost all comparisons between columns were significant.

* Where the groupings J = 1, 2, 3 stand for 1 = Ashkenazi (A), 2 = Sephardi (S), and 3 = Oriental (O).

Another finding is that we can partly contrast the information derived from blood group loci with those derived from protein loci. In all our computations, protein loci proved to be less affected by changes in the weight functions, probably because of their generally smaller number of segregating alleles per locus. The largest differences occurred when taking weightings with high power (viz., s^4 and s^8), where the corresponding distance measures computed for blood group loci were reduced to almost zero, but those computed for protein loci were little affected. Within the Ashkenazi and Sephardi groups both blood group loci and protein loci exhibit similar trends with respect to the comparisons between populations. Within the Oriental group, both blood group loci and protein loci indicate that the Yemenite Jews are nearer than the Cochin Jews to both the Oriental standard and the world-Jewish standard. But when we compare the Iraqi Jews to the Yemenite and Cochin Jews, some conflicting trends appear. The Iraqi Jews show larger distances from the standards for blood group loci and smaller distances for protein loci. This conforms with the widely recognized tenet that protein loci tend to be less variable than blood group loci.

Comparison of populations with different arrangements of the group standards

Can we assess the relative distance between the historical groups? For example, can we say that the Sephardi populations are "halfway" between the Ashkenazi and Oriental Jews? To deal with this question, we executed the analogous calculations and comparisons using the Sephardic group standard as the common reference. (We also considered all permutations of the group standards in making the comparisons.) Table 6 records the relative ordering of all the Jewish populations with respect to the Sephardic standard.

Population orderings with respect to the separate populations as standards

We implement next the methodology using each separate population as a standard. The results are strikingly robust and consistent founded on a wide spectrum of statistics of the form (6). Apart from discerning finer relationships among the separate Jewish populations, we also compare and contrast them with six non-Jewish populations.

The interpretation of the orderings of table 7 paraphrases that of tables 4 and 5. Thus, populations appearing in the same column are not separable and cannot be ordered, while a population in a more left column entails *significantly* (at the 5% level of

TABLE 6

ORDERING INDUCED BY DISTANCES RELATIVE TO SEPHARDI GROUP AVERAGE AS STANDARD

$\Sigma w(s_i)\,(x_i - s_i)^2$

Standard	Population groups			
Sephardi	Moroccan J.	Libyan J.	Iraqi J. German J. Polish J. Russian J. (Rumanian J.)	Yemenite J. Cochin J.

significance) smaller distance from the standard than the populations listed in a column to the right. *The populations that do not appear signify that no orderings can be established between these populations and any of the others.* For definiteness, we record in table 7 the partial orderings for the specific statistic.

$$\sum_{k=1}^{m_i} \sqrt{s_k^{(i)}}\,(x_{k,j}^{(i)} - s_k^{(i)})^2 \ ,$$

where $s_k^{(i)}$ refers to the relevant allele frequency component of the standard at locus i and $x_{k,j}^{(i)}$, the corresponding allele frequency vector of population j.

A cautionary word on the data: The Rumanian Jews and the Polish and Russian non-Jewish populations have some missing data. Their positions are, therefore, based on less than 14 comparisons. We place them in parentheses suggesting their probable position.

Some discussion of table 7

In table 7 we indicated significant orderings at the 5% level. There is a slight degree of subjectivity in completing parts of the table, but this can affect only small details. The overall results are extremely robust with respect to all kinds of transformations and rescalings of the data, and the use of different weights and deviation functions.

As in table 5, the Russian and Polish Jews show up closer to each other than to the German Jews. The Ashkenazi are quite homogeneous as a group and simultaneously relatively heterogeneous (i.e., more central in all their allele frequency distributions). With the German Jews as standard, these conclusions are reinforced. Again we see from table 7 that the Ashkenazi, Sephardi and Iraqi Jews can be well identified with respect to the 15 populations at hand, whereas the Yemenite and Cochin Jews are no more closely related to the other Jews than to the German non-Jews, Arabs, and Samaritans.

A curious fact is that the population closest in genetic composition to the Iraqi Jews are the Polish Jews.

Table 7 also shows that the genetic constitution of the Yemenite Jewish population is so discordant from all the other populations considered that, *with the Yemenite standard, no separations emerged* among all other populations.

When the Arabs and Armenians are taken as standard, the Ashkenazi, Sephardic, and Iraqi populations are indistinguishable, and relatively far from the standard, while the Yemenite and Cochin Jews exhibit a higher order of divergence. The Samaritans appear only remotely comparable to other populations, as might be expected of this small isolate. The complete absence of certain allelic variants of some genes occurs in both the Cochin and Samaritans, especially pronounced in the HLA loci.

The last section of table 7 indicates "closer" connections between German non-Jews and Polish and Russian Jews than between German non-Jews and North African Sephardic, Iraqi and German Jews. This is probably because Polish and Russian Jews are more central in their allele frequency distributions, so they show overall reduced distance when compared with a quite far population.

TABLE 7

ORDERING INDUCED BY DISTANCES RELATIVE TO SEPARATE POPULATIONS AS STANDARDS

$\Sigma \sqrt{s_i}(x_i - s_i)^2 \longrightarrow$ (Increasing distance)

Standards	Population groups			
Ashkenazi Jewish:				
1. Polish	Russian J. (Rumanian J.)	Iraqi J. Moroccan J. Libyan J. German J. Arabs Armenians		Yemenite J. Cochin J. Samaritans Germans (Poles) (Russians)
2. Russian	Polish J.	Iraqi J. Moroccan J. Libyan J. German J. (Rumanian J.)	Arabs Armenians Germans (Poles) (Russians)	Yemenite J. Cochin J.	Samaritans
3. German		Polish J. Russian J. (Rumanian J.)	Iraqi J. Moroccan J. Libyan J. Arabs Armenians Germans (Poles) (Russians)		Yemenite J. Cochin J. Samaritans
Sephardi Jewish:				
4. Moroccan	Libyan J. Iraqi J. German J. Polish J. Russian J. (Rumanian J.)		Arabs Armenians Germans (Poles) (Russians)	Yemenite J.	Cochin J. Samaritans
5. Libyan	Only 3 significant separations obtained: Moroccan J. Russian J.			Yemenite J. Germans	Samaritans

TABLE 7 (*continued*)

Oriental Jewish:				
6. Iraqi	Polish J.	German J. Russian J. (Rumanian J.) Moroccan J. Libyan J.	Yemenite J. Arabs Armenians	Cochin J. Samaritans Germans (Poles) (Russians)
7. Yemenite	No significant separations obtained.			
8. Cochin	Only 2 significant separations obtained:		Polish J.	Iraqi J. Germans
Middle Eastern:				
9. Arabs		Polish J. German J. Russian J. (Rumanian J.) Iraqi J. Moroccan J. Libyan J. Armenians	Samaritans Yemenite J. Cochin J.	
10. Armenians		Moroccan J.	Iraqi J. Arabs German J. Polish J. Russian J. (Rumanian J.)	Samaritans Yemenite J. Cochin J. Germans
11. Samaritans		Armenians German J. Polish J. Russian J.		Iraqi J. Moroccan J.
German non-Jewish:				
12. Germans	Polish J. Russian J. (Rumanian J.)	Libyan J. Iraqi J. Moroccan J. German J.	Armenians Arabs	Yemenite J. Samaritans Cochin J.

NOTE. — Because of missing data, Polish and Russian non-Jewish populations could not be used as standards.

Central range estimates

In order to better *interpret the ordering relations* established in table 7, we proffer estimates of degrees of closeness for the positions of the populations in terms of the distances relative to a standard determined by each particular population. For a given statistic with respect to a standard evaluated at the 14 loci for a specific population, we delete the three largest and smallest values and plot the remaining ("*central range*") values. This is done on a linear scale and also via a log transformation for various choices of weight and deviation measures.

Some comments on figure 2A and B

The central range plots provide an ancillary perspective on the ordering relations of table 7. Observe that with the standard of the Polish Jews, a few small values occur in the central range of the German Jews, while Russian Jews are much closer since none of the corresponding points actually appear. The Yemenites and German non-Jews are in the same order relatively far from the Polish Jew standard.

The German Jews are significantly closer to the Russian and Polish Jews than to the German non-Jews, as is manifest from the shift of the range to the left. There is no discernible relationship of the German Jews in comparing them to the Yemenites against the German non-Jewish populations. Note also the left translation of the range of the Iraqi Jews relative to the Polish and Russian Jewish standards when compared with the German non-Jewish and Yemenites.

The Cochin population manifest an outlier relative to the five standards, and in fact, exhibit this attribute with respect to all the other population group standards. The central range for the Samaritan community relative to the five standards are also very spread, as would be expected from a small isolate.

It is interesting that the Armenians and Arabs show quite similar plots. The range for the Yemenite population are considerably spread with respect to all standards. The German non-Jews are a little further from the Sephardic populations compared to the Ashkenazi, but actually relatively far from both.

The complications inherent to orderings obtained from individual markers

Many "genetic distances" have been proposed to evaluate differences in the genetic composition of populations. Most of them, however, do not adopt a multivariate analysis in assessing multilocus genetic data, but merely rely on a weighted averaging of distances obtained for each locus separately. The investigation of separate markers one at a time does not provide a coherent picture; nor does an averaging over loci, since locus effects can show partially different trends that may cancel out in the averaging process. To illustrate the problems with single-locus analysis, let us examine the HLA complex, which we choose because of its multiple alleles and the widespread contention that these loci are particularly informative. Table 8 shows the ordering of the populations for HLA-A, HLA-B, and the Rh blood group for some of the statistics of the form (6) used in our analysis.

The conclusions from our previous analysis could *not* have been validated by considering the HLA system alone, or the Rh locus. Notice, for example, the disparity in the ordering for Cochin in the information number, and the almost random position

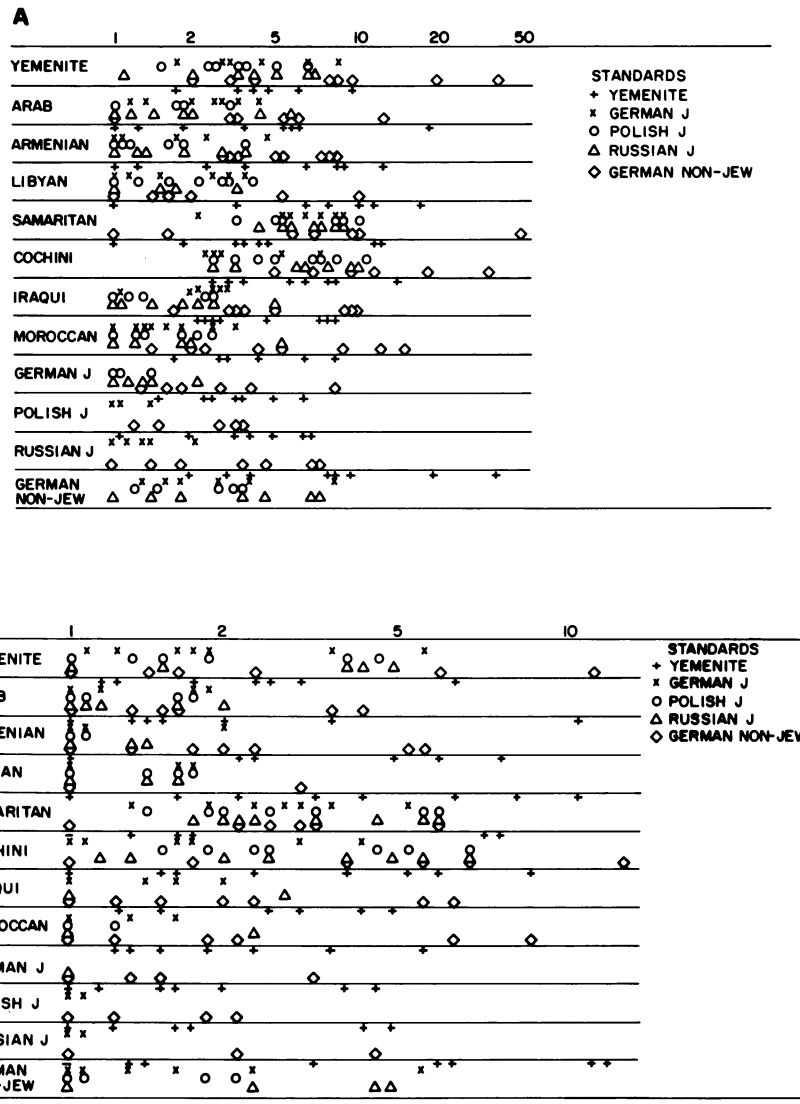


FIG. 2.—These present the central range where the largest and smallest three values, $i = 1,2, \ldots, 14$, for the statistic, (a) $1000 \left(\sum_{\kappa}[x_{k,j}^{(i)} - s_k^{(i)}]^2\right)$, are discarded for population j.

The plots are done for all populations with respect to five specific standards in A.

The central range statistic is described in B for the statistic,

(b) $1000 \sum_{k} \sqrt{s_k}\,(x_{k,j}^{(i)} - s_k^{(i)})^2$, $i = 1,2, \ldots, 14$ for population j.

The plots of statistics (a) and (b) contrast two different weightings of the allele frequencies of the standard. Points of the central range when of magnitude less than 1 are not shown; that is, missing points imply some very close points of the central range.

TABLE 8

ORDERING OF POPULATIONS WITH THE HLA AND RH BLOOD GROUPS

	HLA-A	HLA-B	Rh
Information number ↑	Cochin J.	Libyan J.	Arabs
	Polish J.	Moroccan J.	Russian J.
	German J.	Russian J.	Iraqi J.
	Rumanian J.	Rumanian J.	German J.
	Armenians	Polish J.	Rumanian J.
	Iraqi J.	German J.	Polish J.
	Libyan J.	Cochin J.	Moroccan J.
	Russian J.	Iraqi J.	Yemenite J.
	Arabs	Yemenite J.	Armenians
	Moroccan J.	Armenians	Libyan J.
	Yemenite J.	Arabs	Samaritans
	Samaritans	Samaritans	Cochin J.
Variability distances: (world-Jewish average as standard, squared root weight) ↑	Yemenite J.	Yemenite J.	Libyan J.
	Moroccan J.	Cochin J.	Iraqi J.
	Iraqi J.	Iraqi J.	Cochin J.
	Cochin J.	German J.	German J.
	German J.	Rumanian J.	Moroccan J.
	Libyan J.	Libyan J.	Yemenite J.
	Rumanian J.	Moroccan J.	Russian J.
	Polish J.	Russian J.	Polish J.
	Russian J.	Polish J.	Rumanian J.
Variability distances: (group averages as standard, squared root weight) ↑	Iraqi J.	Cochin J.	Libyan J.
	Yemenite J.	Iraqi J.	Iraqi J.
	Cochin J.	Yemenite J.	German J.
	Libyan J.	Libyan J.	Cochin J.
	German J.	Moroccan J.	Moroccan J.
	Moroccan J.	German J.	Yemenite J.
	Polish J.	Rumanian J.	Russian J.
	Rumanian J.	Polish J.	Polish J.
	Russian J.	Russian J.	Rumanian J.

of the Yemenite population with respect to the three loci above. The Libyan position is also quite variable. However, the Russian and Polish Jewish populations tend to go together.

DISCUSSION AND CONCLUSIONS

The method devised for this analysis should prove useful in studying similarities and differences for other groups of populations for which substantial biochemical polymorphic data are available.

In this last section we first summarize several salient conclusions of the analysis and then suggest possible interpretations of the results. More information has to be collected in order to confirm, modify, and/or extend the interpretations. Two kinds of information are desirable: information on individuals permitting calculations of interactions between loci, and information on more loci and on more populations, both Jewish and non-Jewish. In a future publication we expect to present a statistical analysis of haplotype measurements on individuals from the nine Jewish populations studied here.

Main findings

Despite the classifications of table 1, we found the Ashkenazi, Sephardi, and Iraqi Jewish populations to be consistently close in genetic constitution and, significantly, equally distant from all the other populations, namely the Yemenite and Cochin Jews, the Arabs, and the non-Jewish German and Russian populations. This closeness appears a priori to be quite surprising, since morphological and anthropometric differences between Ashkenazi and Sephardi groups are well recognized.

We found the Polish Jewish community the most heterogeneous (exhibiting the most central allele frequency distribution) among all the Jewish populations, and the Jewish population closest to the world-Jewish average. We offer later a hypothesis that partially explains this finding. A less surprising result is that Polish and Russian Jews are closer genetically to each other than to German Jews. We found also that all Ashkenazi Jews are very distant from German non-Jews in genetic composition.

The individual Ashkenazi populations are closer to their group frequency distribution than the individual Sephardi populations are to theirs. The Oriental populations do not constitute a meaningful group. Indeed, for all practical purposes the Yemenite Jews and the Cochin Jews bear little biochemical genetic resemblance to each other, or to any other population considered in this study.

Three technical findings are of interest. One is that any analysis based on a single locus or even a few loci, notably the HLA complex or the Rh factor (table 8), usually does not do justice to the problem. Another, the bulk of our orderings, can be arrived at by examining the eight blood loci only; the protein loci add further significance, but tend to discriminate less among the populations. Finally, the heterogeneity indices (i.e., measures of distance from centrality), namely the information number and the heterozygosity value, discriminate less sensitively than the distance measures relative to the appropriate group standards.

Some comments on recent Jewish history

Before proceeding to interpretations of our results, it is pertinent to note some information on recent Jewish history. A number of works on Jewish population structure and historical events are relevant to our population studies [13–30]. Basically, they describe a decrease in the world Jewish population from about 1,250,000 in the 11th century to perhaps 250,000 by the end of the 15th century distributed over many small communities in Eastern Europe, the Balkans, North Africa, Iraq, and Persia. After 1500 the Jewish population started to grow again.

Significant migrations of Jewish groups from Western Europe, principally Germany, occurred during the 13th to 16th centuries settling in the Polish dominion, extending from the Baltic to the Black Sea. During the Golden Age of Poland, from 1500 to 1648, the Jewish population increased to over 500,000 [13, 24, 29]; but the Cossack massacres of 1648 and other causes reduced their numbers to below 100,000. The subsequent intermittent pogroms caused some Jews to move elsewhere within Poland, others to migrate to other lands, and still others to convert to Christianity.

Following the partitioning of Poland in 1792–95, Polish Jews were left more to themselves. However, in Eastern Poland, now part of Russia, the "Pale" was

established, and its area was periodically shrunk in order to cause more severe living constraints with a partial objective of stamping out Jewish separateness. Other measures were taken to this same end: for example, all young men were forced to serve in the Army, and various inducements to assimilation were provided. Jewish leaders reacted by insisting on religious orthodoxy and endogamy. At the same time, the division of Jews into religious factions (Hasidim, non-Hasidim, local "Rebbi" constellations) subdivided the Jewish population into smaller endogamous communities.

Following the cholera epidemic of 1868 and the major Russian pogroms of 1881–84 and 1903–05, some 70% of all Russian Jews migrated, mostly to the United States. A lesser percentage (on the order of 40%) migrated from Poland between 1880 and World War II [24, 29]. Baron [24] estimates the Jewish population of Poland at about 400,000 in 1820, 600,000 in 1850, over one million in 1880, and close to two million in 1920. In all Russian cities in 1898, Jews numbered only about 600,000.

The situation of the Jews in Germany during the last 500 years is briefly as follows. It is known that large portions of the German Jewish population (culminating a conglomeration of persecutions and epidemics by the 14th to 16th centuries) for all practical purposes, disappeared or moved away (mostly to Poland). There are even cases of family migrations from Germany to Iraq in the early 1500s [17, 19]. A renovation of Jewish settlements throughout Germany commenced about the middle to late 17th century from migrations, primarily of Polish origins, and secondarily of Sephardi Jews by way of Holland and Italy.

The relationship of North African and Iraqi Jews to European Jewish centers is poorly documented. A perhaps relevant event concerns the reverberations of the messianic movements, particularly of Sabbatai Zvi of the 17th century, which started in Turkey but reached segments of Polish Jewry and touched Baghdad and parts of Iran [30]. Most of their adherents were ultimately lost to the Jewish fold, but some remnant probably remained.

Connections of Iraqi Jews with the Balkan states to the west, and Persia and India to the east, have been established [15]. Other works [17, 19] tell of successful Jewish tradesmen centered in Baghdad, especially in the 19th century. A substantial number of Sephardi Jewish communities, including many merchants, were located along the trade routes from Italy (e.g., in Livorno) through Mesopotamia and Iraq, to Persia and India [17, 18, 23]. There were settlements of Ashkenazi Jews in Egypt and North Africa during the 18th and 19th centuries [20, 25]. We know also that Jews migrated from Persia to southern Russia in the Middle Ages [23] and the 17th century [16]. The spread of the Turkish empire during the 17th to 19th centuries, which reached regions of Hungary and included parts of Iraq, may also have benefitted Jewish movement in these areas, as Jews fared well in their dominion.

Seemingly, then, contacts between Jews in Europe, North Africa, and Iraq persisted as late as the 19th century. Parallels exist between the codification of the Halachah (Jewish ritual and custom) promulgated by Joseph Karo in Palestine, who had connections with Jews of Iraq and Egypt, and the codification established (completed by Moshe Isserlish) in German-Polish Jewish religious centers. We know also that there were contacts between Rabbinical cohorts in Israel, Egypt, and Iraq, probably

extending to other communities of North Africa. For example, Jewish settlements of Moroccan Jews in Egypt are recorded [20].

Possible interpretations, hypotheses, and speculations

The fact that Ashkenazi and Sephardi Jews, and to some extent Iraqi Jews, constitute a reasonably consistent grouping for the genetic composition of the 14 loci of table 2, poses a fascinating and challenging problem both historically and genetically. The morphological and anthropometric differences are well recognized. How can we account for the genetic similarity? Another tantalizing result concerns the centrality of the Polish Jews more so than any other Ashkenazi Jewish population. The exclusion of the Yemenite Jews is also striking. We shall venture some possible explanations. We suggest at this point rereading the discussion concerning tables 3–5 and table 7.

Consistent with our results is the thesis that the progenitors of the present-day Ashkenazi and Sephardi Jews were the remnants of several small Jewish populations of the 14th and 15th centuries. Actually, the current Ashkenazi and Sephardi Jews appear to represent a significant expansion of smaller populations that survived the tribulations of the 14th to late 15th century. Some later population declines in the 16th and 17th centuries in Eastern Europe may also be relevant. The biological relationship of these groups to the "original" (Biblical) Jews, if such a group can be defined, is tenuous at best. Since the Christian reconquest at the end of the Middle Ages, the contribution of non-Jews to the Jewish gene pool has been extremely small. The flow has been largely in the other direction, owing in good measure to conversions to promote individual success and avoid physical and mental suffering.

A meaningful hypothesis asserts that the blood and protein loci at hand, under conditions where the admixture rate with large outside sources are minimal, have maintained rather constant gene frequency distribution for the past 500–800 years in these Jewish populations. The contacts between Jewish centers in Europe, North Africa, and Iraq up to that period are established. Thereafter, these populations separated more, and the inflow of non-Jews into Jewish populations was minuscule, whereas outflow from the Jewish fold was substantial. The almost constancy of the gene frequency distribution is reasonable where no significant selection component or adaptability contrasts is endowed to these loci, and we could indeed expect the frequency distribution similarities between and within Ashkenazi, Sephardi, and Iraqi Jewish populations as reflected in our analysis.

Why are Russian and Polish Jews closer to each other genetically than either population is to German Jews? The shifting borders and changing political climate of Poland and Russia from the 17th century through the 19th probably led to a certain amount of mixing of their Jewish populations. Also a number of the Jews who left Russia following the widespread pogroms of 1881 and 1903 were absorbed by Polish Jewish communities, though most migrated to the United States. Further, Russian and Polish Jews had very similar social institutions and life styles: both groups strictly observed the Jewish traditions and tended to live in small enclaves, an arrangement conducive to inbreeding and high birthrates, especially in the families of Jewish scholars and merchants.

The significant heterogeneity of the Polish Jewish population gene frequency

distributions could be a result of averaging over numerous small communities with an assortment of alleles predominant in different village complexes; that is, it may be related to the Karlin-McGregor weak coupling principle applicable under conditions of small migration flow between the numerous subpopulations [31]. A further contributing factor may be historical and partly related to the life style of Polish Jews and the "steytl" social system. Polish Jews constituted the largest Jewish community in the world from the 19th century to World War II. As a result of intermarriage, assimilation, and persecutions, a nontrivial flow of Polish Jews into the neighboring non-Jewish populations occurred, along with considerable migration to distant places. The remaining Jews were largely the ardent orthodox who thrived in the confines of small villages through Eastern and Central Europe. Into this population there was a negligible inflow from non-Jewish sources. By contrast, the German Jews from the 18th century on were more open to Christian influences, yielding concomitantly more mixing among themselves. Of course, in Germany as in Poland, the progeny of Jews who married Christians overwhelmingly left the Jewish fold.

The interesting finding that the Polish Jews are closer to the average world Jewish population may involve several influences. A contributing factor pertains to the large population explosion of Polish Jews in the 18th and 19th centuries, coupled with the intermittent severe persecution under the Polish Empire and subsequent Russian, Polish, and Prussian overlords, which led to many dispersions and migrations. In this way, the influence of the Polish Jewish gene pool may have been partly distributed into the other Ashkenazi and Sephardi populations. The centrality of Polish Jews, rather than the Russian-Polish combination, may be an artifact of the data founded on sampling Israeli individuals. We know that the bulk of Russian Jewry immigrated to the United States admitting some inflow into Poland. The Polish Jewish contingent for the first half of the 20th century (the largest such Jewish community), apparently maintained ties throughout Europe, and possibly with some North African populations (e.g., as in the Ashkenazi centers in Egypt). Noting that the majority of World War II refugees absorbed in Israel are Europeans, North Africans, and Iraqis, the pronounced Polish gene expression is understandably more central.

The Yemenite community appears to be a residual of a large population of the early Middle Ages. The admixture with current Arabic groups appears to be minimal, as revealed by the genetic marker comparisons (table 7) and corroborated further by known historical developments.

The physical differences (mostly inherited as "polygenic" characters) between the Ashkenazi and Sephardi populations could possibly be adaptive (i.e., a result of natural selection forces operating on adaptive polygenic traits) rather than of gene flow from non-Jewish to Jewish populations in Europe, which, as we have seen, was almost surely negligible. The same considerations apply in comparing Oriental Jewish populations. In general, our findings suggest that whereas different patterns in biochemical group gene frequencies reproduce possible migration paths of a population, physical differences tend to reflect environmental conditions acting at the phenotype level.

Our findings and analyses argue against any but very slight gene flow from non-Jewish neighboring populations over the last five centuries. Moreover, our findings

bear out the view [13, 26, 30, 32] that the progenitors of the present Jewish people derive from a number of small Jewish remnant communities, mostly scattered through Europe and North Africa, that survived the persecutions, epidemics, and famines of the 14th and 15th centuries. These results and interpretations contrast with the discussion of [5, 6]; compare also with [28] and [11].

Relationships with "Jewish genetic diseases"

Adam [32] has provided an excellent review of the distribution of certain genetic diseases prominent and rare (or absent) among various Jewish communities and some associated population groupings. He highlights a number of neurological disorders, including Tay-Sachs, Gaucher, and familial Dysautonomia, among Ashkenazi Jews; familial Mediterranean fever among Sephardi Jews; glycogen storage disease type III, which is exclusive to Moroccan Jews; and favism, which has a high frequency among Iraqi Jews. In explaining the differences between Jews and non-Jews in many of these disease frequencies, Adam assigns a large weight to founder effects (related to the small Jewish population size of the 15th century) rather than selection-environmental correlates. Fraikor [13] argues in favor of a founder effect in Tay-Sachs disorder accompanied by continued population changes, special features associated with the steytl societies; the nature of Jewish migrations, and organizations of the immigrant American Jewish communities. Interestingly enough, Fraikor notes that Tay-Sachs disease, a big problem among Russian and Polish Jewish populations of near Balkan ancestry, is found among German Jews as rarely as it is among European non-Jews. This fact conforms nicely with the population orderings of tables 4, 5, and 7.

Three kinds of hypotheses are commonly advanced to explain an excessive frequency of a rare genetic disease trait. One attributes the excess to inbreeding on an isolate. Another invokes the founder principle, with subsequent genetic drift concomitant to population increases. The third postulates some form of mutation-selection balance or possibly a case of selection overdominance.

Efforts to develop population classifications by comparisons based on frequency distributions of rare genetic diseases are fraught with pitfalls. Properly speaking, the so-called "Jewish genetic diseases" are not always irrevokable. For example, Tay-Sachs disease has been primarily reported among American Ashkenazi Jews of Polish or Russian of primary Lithuanian origin; its estimated incidence among their Israeli counterparts appears to be lower. The usual frequencies quoted are averages of large Ashkenazi populations. Adam [32] stresses that the "Ashkenazi diseases" vary in frequency according to the geographic origin of the patients' recent ancestors [33]. It should be emphasized that systematic screening programs on these diseases for other populations are lacking. Bloom syndrome, once considered a "Jewish genetic disease," can no longer be placed in this category according to recent studies by German [34].

The existence of "Jewish genetic diseases" is less pronounced among Sephardi and Iraqi Jews. G6PD deficiency and increased cases of Thalassemia α and β presumably confer a selective advantage in malarial regions, but these occur also in other populations subject to these conditions.

In our opinion, "Jewish genetic diseases" can be largely an artifact of the intensive

data collection among Jewish populations. As more population groupings and/or isolates are researched, the concept of "Jewish genetic diseases" will probably disappear. McKusick [35] places the Amish isolate among the populations with the highest percent of rare recessive disorders ascribed to excessive inbreeding (see also [36]). Similar representations are expected for most isolates. In any case, the use of rare genetic disease frequencies is likely inadequate as a basis for relating or classifying populations.

REFERENCES

1. BONNÉ-TAMIR B, KARLIN S, KENETT R: Analysis of genetic data on Jewish populations, I. historical background, demographic features, and genetic markers. *Am J Hum Genet* 31: 324-340
2. PATAI R, PATAI-WING J: *The Myth of the Jewish Race*. New York, Scribner, 1975
3. SPIELMAN R, SMOUSE P: Multivariate classification of human populations, I. Allocation of Yanomama Indians to Villages. *Am J Hum Genet* 28:317–331, 1976
4. SMOUSE P, NEEL J: Multivariate analysis of gametic equilibrium in the Yanomama. *Genetics* 85:733–752, 1977
5. CARMELLI D, CAVALLI-SFORZA LL: The genetic origin of the Jews: A multivariate approach. *Hum Biol*. In press, 1979
6. CAVALLI-SFORZA LL, CARMELLI D: The Ashkenazi gene pool: Discussion and interpretation, in *Genetic Diseases Among Ashkenazi Jews*, edited by GOODMAN RM, MOTULSKY AG, New York, Raven Press, 1979
7. SMOUSE P, SPIELMAN R, PARK M: Multiple-locus classification of individuals as a function of the genetic variation within and among human populations. Submitted for publication
8. ROYALTEY HH, ASTRACHUN E, SOKAL RR: Tests for pattern in geographic pattern. *Geographical Analysis* VII. 369–395, 1975
9. FELSENSTEIN J: Maximum likelihood estimation of evolutionary trees from continuous characters. *Am J Hum Genet* 24:471–492, 1973
10. GOODMAN M: Distance analysis in biology. *Syst Zool* 21:174–186, 1972
11. BONNÉ-TAMIR B, BODMER J, BODMER W, PICKBOURNE P, BRAUTBAR C, GAZIT E, NEVO S, ZAMIR R: HLA polymorphism in Israel, an overall comparative analysis. *Tissue Antigens* 11:235–250, 1973
12. CAVALLI-SFORZA LL, EDWARDS A: Phylogenetic analysis: Models and estimation procedures. *Evolution* 21:550–570, 1967
13. FRAIKOR A: Tay-Sachs disease: genetic drift among the Ashkenazi Jews. *Hum Biol* 24:117–134, 1977
14. MARCUS JR: *The Jew in the Medieval World*. New York, Harper, 1965
15. GOITEIN SD: *Letters of Medieval Jewish Traders*. Princeton University Press, 1973
16. Mediterranean Society. *The Jewish Communities of the Arab World as Portrayed in the Documents of the Cairo Gneiza*, vol 3, Berkeley, University of California Press, 1973
17. BEN-JACOB A: *The Jews of Mesopotamia, 1938–1960*, Jerusalem, 1965
18. ROTH C: *The Sassoon Dynasty*. London, 1941
19. SASSOON DS: *A History of the Jews in Baghdad*. Letchworth, 1949
20. LANDAU A: *Jews in Nineteenth Century Egypt*. New York, Jewish Publishing Soc, 1971
21. JACOBS L: *Jewish Law*, New York, Behrman House, 1968
22. HIRSCHBERG HZ: *History of Jews of North Africa*. Leiden, Brill Publishing, 1973
23. LOEB L: *Jewish Life in Iran*. New York, Gordon Breach, 1978
24. BARON SW: *The Russian Jew under Tsars and Soviets. Russian Civilization Series*, edited by FLORINSKY MT, New York, MacMillan, 1964
25. CHOURAQUI AN: *A History of North African Jews*. New York, Jewish Publishing Soc, 1973
26. BARON SW: *A Social and Religious History of the Jews*. New York, Jewish Publishing Soc, 1960

27. COHEN HJ: *The Jews in the Middle Eastern Countries 1860–1971* (in Hebrew). Jerusalem, 1972
28. MOTULSKY AG: Ashkenazi Jewish gene pools: Admixture, drift and selection, in *Population Studies on Isolates*, edited by JUSELIUS S, London, Academic Press, 1979
29. WEINRYB BD: *The Jews of Poland*. Philadelphia, Jewish Publishing Soc, 1972
30. DUBNOV S: *Divrey Yemey Am Olam* (in Hebrew). vol, 7–9, Tel-Aviv, Dvir Publication, 1948
31. KARLIN S, MCGREGOR J: Application of method of small parameters to multi-niche population genetic models. *Theor Popul Biol* 3:186–208, 1972
32. ADAM AC: Genetic diseases among Jews. *Isr J Med Sci* 9:1383–1392, 1973
33. SHEBA C, SZEINBERG A, RAMOT B, ADAM A, ASHKENAZI Y: Epidemiological surveys of deleterious genes in different population groups in Israel. *Am J Public Health* 52:1101–1108, 1962
34. GERMAN JL: The current status of Bloom's Syndrome, in *Genetic Diseases Among Ashkenazi Jews*, edited by GOODMAN RM, MOTULSKY AG, New York, Raven Press, 1979
35. MCKUSICK VA: The ethnic distribution of disease in the United States. *J Chronic Dis* 20:115, 1967
36. BRUNT PW, MCKUSICK VA: Familial dysautonomia. A report of genetic and clinical studies with a review of the literature. *Medicine* 49:343–349, 1970