Synthetic Maps of Human Gene Frequencies in Europeans

These maps indicate that early farmers of the Near East spread to all of Europe in the Neolithic.

P. Menozzi, A. Piazza, L. Cavalli-Sforza

Summary. Multivariate techniques can be used to condense the information for a large number of loci and alleles into one or a few synthetic variables. The geographic distribution of synthetic variables can be plotted by the same technique used in mapping the gene frequency of a single allele. Synthetic maps were constructed for Europe and the Near East, with the use of principal components to condense the information of 38 independent alleles from ten loci. The first principal component summarizes close to 30 percent of the total information and shows gradients. Maps thus constructed show clines in remarkable agreement with those expected on the basis of the spread of early farming in Europe, thus supporting the hypothesis that this spread was a demic spread rather than a cultural diffusion of farming technology.

The study of the geographic distribution of genes has been useful in suggesting selective mechanisms that favor one or another of the alleles or loci. The correlation of the geographic distribution of thalassemia or sickle cell anemia with that of malaria indicated the possible advantage of malarial resistance among heterozygotes for these diseases (*1*). Such correlations can only suggest hypotheses, however, and these must be tested by other, more direct types of observations at the individual rather than at the population level.

Another approach for studying the genetic differentiation of populations has made use of coefficients of genetic distance between population pairs. Here the emphasis is on the overall genetic differentiation accumulated between two (or more) populations. It is very unlikely that the total difference is caused by a single selective factor; rather, the distance should express the outcome of many different selective processes, most of them unrelated to each other, and of random events, such as those determining genetic drift. As a result, the cumulative genetic difference between two populations can be expected to be a summary of their evolutionary history, being proportional to the time of separation and inversely related to the intermigration between them (*2*).

Distances refer to the comparison of population pairs. Some other approach, however, is needed if one is interested in representing the evolutionary history of the populations of a large area, or their overall similarities. A reasonable approach is that of using one (or few) variables that summarize most of the available information. Some of the information may be lost but one can minimize that amount by appropriate multivariate techniques. The synthetic variables thus obtained are usually, as in Eq. 1, linear combinations of the original variables, the gene frequencies. If the gene frequency of an allele at locus 1 in a given population is p_1, that of another allele at the same time or another locus is p_2, and so on, with a total of k alleles (usually belonging to many loci), then the population can be represented by at least one new synthetic variable S_α.

$$S_\alpha = \alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_k p_k \quad (1)$$

Using one of various related statistical models one can determine the coefficients $\alpha_1, \alpha_2, \cdots \alpha_k$ in such a way as to minimize the amount of information lost. One can then renew the process by computing another set of coefficients $\beta_1 \cdots \beta_k$ (to replace the α's) and thus generate a new synthetic variable S_β which retrieves as much as possible of the information lost by S_α; and so on. The amount of information that can be actually summarized with a small number of variables depends on the data set. Clearly, the procedure will be most useful if much of the information can be summarized with the use of only a few variables.

Methods of this kind have been widely used in biology, particularly in genetics (*3*), to generate plots useful, for instance, in clustering analysis. The application we suggest is, however, a different one; we attempt to use a synthetic variable to make a single geographic map summarizing the individual geographic maps supplied by the alleles whose frequencies have been used to generate the synthetic variable. We hope that geographic maps of synthetic variables will give us new insight into the evolutionary history of the populations represented in the map. Given that the synthetic variables are constructed so as to be statistically independent of one another, each of them may provide a different picture. Moreover, we may summarize more than one such variable and easily up to three, in one single geographic map, by using a different color for each one and making use of the capacity of the human eye to synthesize three colors.

In this article we describe the application of this technique to the problem of expansion of early farming in Europe. Previous work (*4*) demonstrated that the spread of farming from the area of origin of agriculture in the Near East was relatively slow and regular, starting about 9000 years ago and finishing about 5000 years ago. The radial rate of advance was of approximately 1 kilometer per year. It was argued that the process of spread might have occurred either (i) by diffusion of the farmers themselves (demic diffusion); or (ii) the technology might have spread to preexisting populations of hunters or gatherers living in these areas (cultural diffusion). The two hypotheses are not mutually exclusive. We are interested here in the fact that the genetic consequences of the two modes of diffusion are quite different. The migration of farmers from the Near East toward Europe should spread the farmers' genes to all of Europe. By contrast, with purely cultural diffusion no direct effect on genes is expected (except for selection due to the changed way of life). Finally, if farmers spread and also mixed with preexisting hunter-gatherers, one

Dr. Menozzi is lecturer of statistics at the University of Parma, Parma, Italy. Dr. Piazza is associate professor of medical statistics at the University of Turin, Turin, Italy. Dr. Cavalli-Sforza is professor of genetics at Stanford University School of Medicine, Stanford, California 94305.

0036-8075/78/0901-0786$02.00/0

should observe in Europe clines, that is, gradients of gene frequencies originating in the Near East, which is shown in the southeast corner of the map (see cover), and expanding radially from this corner. Data from single genes [for example, the Rh-negative alleles (*5–7*)] are in agreement with the idea of an earlier, largely Rh-negative population occupying mostly the west of Europe with which a slow wave coming from the east intermingled gradually. Other genes (especially some HLA-B alleles) show a similar northwest–southeast gradient across Europe (*8*). Is the northwest–southeast gradient shown by these genes confirmed when all the information from all other available markers is accumulated, or is it idiosyncratic to these genes? We have used European gene frequency data in order to test the demic and cultural diffusion hypotheses and the general usefulness of the method.

Methods for Map Construction and Sources of Data

The methods used are only briefly described here; for more details, see (*9*). The variable to be mapped (gene frequency, or principal component, or other) is interpolated at the nodes of a regular rectangular grid, generated by a Mercator projection. The interpolation followed a mixed strategy. The information available for each grid node within a circle of given radius around the node to be interpolated was first evaluated by counting the number of data points available in the circle (*9*). The radius of the circle was computed by a simple formula and varied from 7° to 17° great circle distance. If there were at least four data points available within the circle in each of at least three quadrants of the circle, then the strategy of interpolation was to fit a second degree surface to the data within the circle, weighting each point within the circle by its sample size and by the square function of the distance between the data point and the grid node being interpolated. If the information available was less than the above, then a weighted average of the data available in the circle was used, with weights as above. Details and reasons for the choice, and validation procedures will be found in (*9*).

Computer time is a direct function of the square of the mesh of the interpolation grid; the resolution of the map in terms of accuracy of representation of the original data is an inverse function of the mesh. In Europe we used an interpolation grid having constant spacing between nodes of 3° longitude and 1°30′ latitude. Thus the nearest representation of a data point on the grid may be removed from the actual location of the point itself, at most 1°30′ longitude and 45′ latitude. This was considered a reasonable compromise between the opposite requirements of keeping computer time within limits and improving resolution. For purposes of graphical presentation, however, a thinner grid is necessary. This was obtained by employing splines under tension to interpolate map values every half degree. Matrices with this thinner grid were printed directly on a terminal with type suitable for imitating various degrees of shading, or were projected on a color television screen. The Grinnell image processor was used which allows up to 16 grades of the color green or red and up to eight grades for blue (*10*).

Data were obtained for a total of ten loci and 38 "independent" alleles (counting for each locus the number of alleles for which adequate data were available, minus one). For 21 alleles from two loci for histocompatibility, HLA-A (*A1, A2, A3, A9, A10, A11, A28*) and HLA-B (*B5, B7, B8, B12, B13, B14, B18, B27, B15, B17, B21, B22, B35, B40*) the summary by Ryder *et al.* (*8*) was used, with the addition of other populations (*11*). Non-HLA markers used were obtained from the tables (indicated in parentheses) of Mourant *et al.* (*7*): the locus for blood group substances ABO (alleles A_1, *B, O*, tables 1.2, 1.21, and tables II, III); the Rh (rhesus) factor locus (CDe, Cde, cDE, cDe, cde, tables 4.13, 4.13.1, 4.17, and 4.17.1); the locus for the MN blood group factors (MS, Ms, Ns, tables 2.7 and 2.11); Lewis (Le^a, tables 7.3 and 7.7.1); Duffy (Fy^a, tables 8.1, 8.1.1, 8.3.1, and 8.3.2); haptoglobin (Hp_1, table 36.1); and acid phosphatase (P^a, P^b, table 50.1.1); and phosphoglucomutase (PGM_1^1, table 53). Data used belong to the region indicated in Figs. 1 to 6. Gypsies and Jews from non–Middle East regions were excluded, whereas Lapps were included. The counts of alleles refer to the total of "independent" alleles: for each locus, the total number of alleles on which sufficient data were available, minus one.

The Construction of Synthetic Variables and Their Meaning

Principal components supply excellent "synthetic variables" (at least under a specific evolutionary hypothesis) especially with a dispersion matrix obtained after subtracting from all gene frequencies the average gene frequency over all populations considered in the analysis (*12*). In a study in which five different transformations of gene frequencies were used, it was found that there was little to be gained by using various transformations (*13*). In that work, however, the standardization of variances led to unsatisfactory results. Accordingly, it was not applied here. The original gene frequencies were therefore employed.

In any type of multivariate analysis, missing data may cause considerable difficulty. In the case of HLA genes in Europe (21 independent alleles for two loci) data were available for all of the alleles in each of the 67 populations and the problem did not arise. For non-HLA data (17 independent alleles), however, it is extremely rare to find data for all markers on the same population. This makes it impossible to compute directly principal components for non-HLA data. The procedure we employed, therefore, was first to construct maps for each allele and then to obtain from each map the interpolated gene frequency at each of 400 preselected locations. The locations chosen were all the interpolated grid points in regions north of the 31° parallel. This eliminated most of North Africa, Arabia, and India, where observations are especially rare and are limited to too few markers. The interpolated gene frequencies at the 400 map locations for all alleles were subjected to principal components analysis. The first three principal components thus obtained were used to generate geographic maps, each by itself and also in superposition (with different colors). It was found useful to use the most visible color, green, for the first principal component, blue for the second principal component at a slightly lower average intensity, and, at a still slightly lower intensity, red for the third component. Decrease in average intensity from first to third was meant to parallel, even if roughly, the relative importance of the three components, as indicated by the relative fraction of variance each of the components explains. More sophisticated ways of representing the relative importance did not seem warranted at this stage.

We have already mentioned that unlike the non-HLA markers, HLA frequencies formed a compact set of data without missing values. They therefore lent themselves to direct application to principal components analysis of the observed population data, and were therefore so analyzed. But they were also analyzed by the "indirect" technique described above, of first building a map for each allele, and then computing principal

components from the systematic set of allele frequencies interpolated at 400 map locations. The two procedures gave final maps of principal components in good agreement, thus validating our indirect procedure of using map values to build principal components. We tested the agreement by computing the correlation between map values obtained by the technique of using population data directly and values obtained by the indirect technique. This correlation was found to be 0.912 ± 0.030. We also performed a discriminant analysis on the 67 populations using expected multinomial variances and covariances within loci for the matrix within populations. The first discriminant thus obtained showed a correlation of $r = 0.974$ with the first principal component, showing that in the present case the two methods give very close results.

Fig. 1. The first principal component of gene frequencies from 38 independent alleles at the human loci: ABO, Rh, MNS, Le, Fy, Hp, PGM_1, HLA-A, and HLA-B. Shades indicate different intensities of the first principal component, which accounts for 27 percent of the total variation and is represented with green shades in the photograph on the cover.

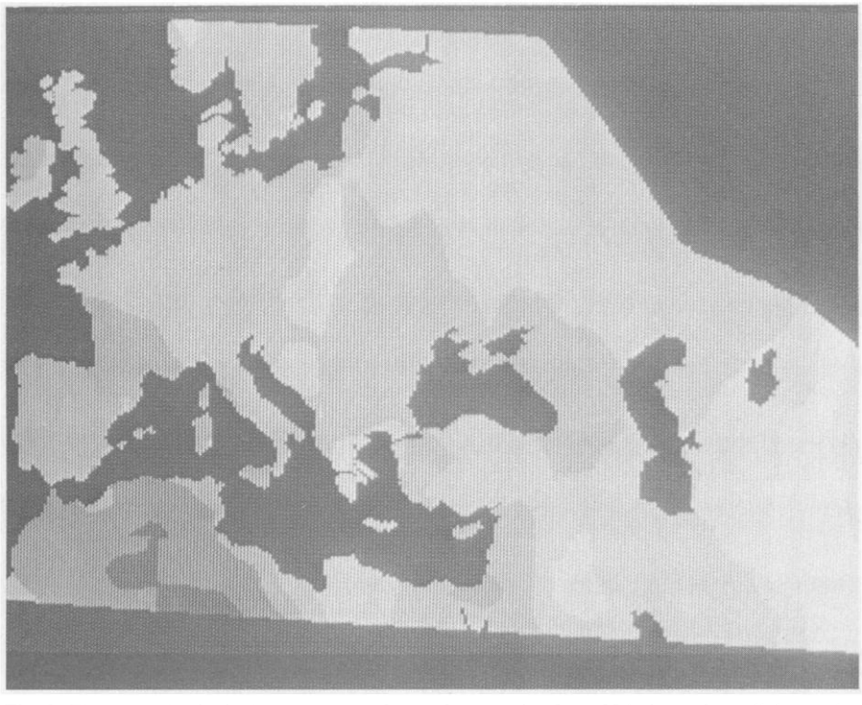

Fig. 2. The second principal component of gene frequencies from 38 independent alleles at the human loci: ABO, Rh, MNS, Le, Fy, Hp, PGM_1, HLA-A, and HLA-B. Shades indicate different intensities of the second principal component which accounts for 18 percent of the total variation and is represented with blue shades on the cover.

The Spread of Farmers in Europe and the Near East

We have mentioned that if there was a slow and gradual migration of farmers originating from the Near East into Europe, starting therefore from the southeast corner of the map shown on the cover and in Figs. 1 to 3, with progressive admixture with local European hunter-gatherers, a gradient of gene frequencies originating from the southeast should be observed. This postulated gradient will be referred to for convenience as the northwest-southeast cline or gradient. Such a gradient is of course expected (i) only for those genes for which there were, before the spread, clear-cut differences in allele frequencies between the European hunter-gatherers and the Near East farmers (other genes cannot show a gradient), and (ii) if subsequent evolutionary events, selective and migrational, were not of such magnitude as to modify or alter the gradient beyond recognition.

The areas to which expansion of Near East farmers may have spread includes in addition to Europe, North Africa (the Sahara not being a desert at the time), Arabia and East Africa, and Southwest Asia as far as the Indus Valley. The archeological information about these areas is, however, much more limited than that available for Europe. In addition, the genetic information for these areas is quite meager. While some non-European population data from neighboring regions have been used in map construction to improve the construction of map edges, we have basically confined our study to Europe and the Near East. In particular, the maps are especially poor for North Africa, where the paucity of data makes it necessary to employ rather remote populations, thus generating considerable uncertainty about the estimates for this area.

Principal component analysis was done for all 38 independent alleles, and also separately for all 17 non-HLA alleles as well as for the 21 HLA alleles. For the HLA alleles, the A and B loci were also analyzed separately. Table 1 shows the proportion of information (the variance between populations) explained

by the three major principal components. The first component explains about one-third of the total variation. Three major components together explain over one-half of the variation. As might be expected, with increasing number of alleles, there is a decrease in the fraction of variance explained. The decrease, however, is not at all a rapid one. It seems rather as if the fraction of variance explained tends to stabilize to an asymptotic value, becoming independent of the number of markers (on the order of 20 or more). This indicates that the picture should not change in an important way on addition of further markers.

On the cover of this issue we show a trichromic synthesis of the first three principal components from the 38 alleles. The individual components are also shown in Figs. 1 to 3. In Figs. 4 and 5 we show the first component of the two subsets of data, HLA and non-HLA. For a comparison, we also include a map (Fig. 6) showing the spread of early farming constructed entirely on the basis of archeological data.

There is a remarkable similarity between the map of the first principal component (Fig. 1 and the green shading in the map on the cover) and the archeological map of the advance of early farming (Fig. 6). In both cases there is a general gradient originating in the southeast corner of the map and spreading to the rest of the land. Both the genetic and the archeological map correspond to a series of concentric circles centered in the Middle East, tending to flatten into ellipses with the major diameter being horizontal. This is the northwest-southeast gradient or cline. It should be noted that it is totally arbitrary to set the values of a principal component in ascending or descending order; the presence of the lightest shades in the southeast corner and the darkest in the northeast corner has no special meaning and could be reversed. Also, the scale of principal components is to a large extent uninformative and is omitted.

The second principal component (Fig. 2 and the blue shading in the map on the cover) shows a marked east-west gradient, with few irregularities of uncertain meaning. The third principal component (Fig. 3 and the red shading in the map on the cover) shows a gradient from the northeast to the southwest, almost perpendicular to that of agriculture, with a central area in the Ukraine.

The trichromic map on the cover is especially suited for looking at overall similarities between populations. Among similarities of populations relatively distant geographically one from the other, one may note that of Scandinavia with Scotland and Northern Ireland. It is well known that these areas of the British Isles were those most densely settled by the Vikings, as can be judged for instance by place names (*14*). The two spots of color different from background in southern England and northern France correspond to the locations of London and Paris, respectively, being slightly misplaced because of low geographic resolution. Note that the colors indicate admixture in London with immigrants from the north of England and in Paris with immigrants from the south of France. The southern parts of France, Italy, Greece, and Turkey are all very similar, indicating a probable influence of Greek colonizations. In general, the central and eastern parts of the Mediterranean are very similar. Sardinia, and Algeria in the part nearest to Sardinia, show an even more striking similarity to the Middle East than the rest of the cen-

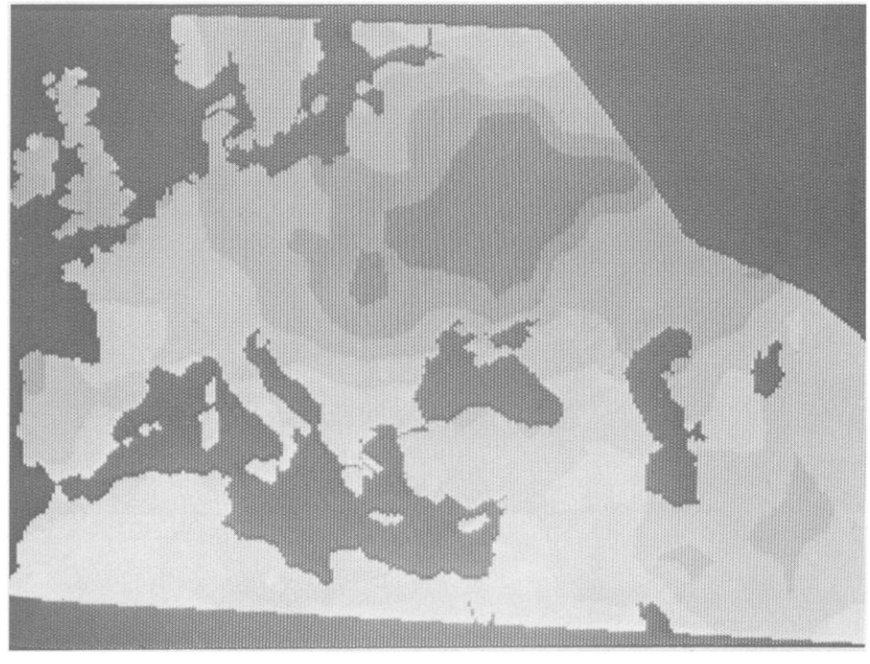

Fig. 3. The third principal component of gene frequencies from 38 independent alleles at the human loci: ABO, Rh, MNS, Le, Fy, Hp, PGM_1, HLA-A, and HLA-B. Shades indicate different intensities of the third principal component which accounts for 11 percent of the total variation and is represented with red shades on the cover.

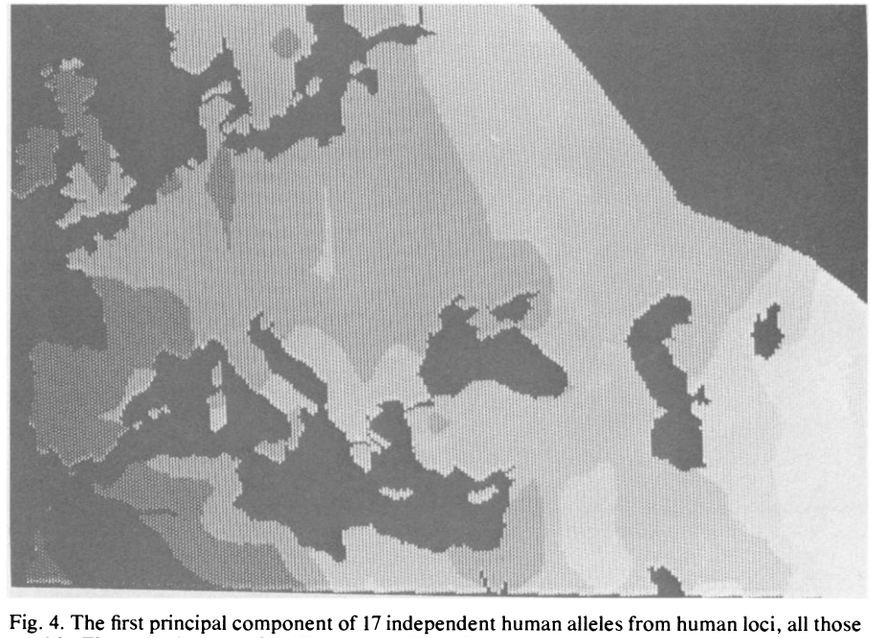

Fig. 4. The first principal component of 17 independent human alleles from human loci, all those used in Figs. 1 to 3 except for HLA-A and HLA-B.

tral Mediterranean. The information on Algeria is extremely limited, and the map values for Algeria are strongly influenced by information available for the nearby island of Sardinia. This island has been remarkably well investigated. Its low Rh-negative frequency, and the abnormal frequencies for all other genes have been an anthropological mystery. The archeology of Sardinia shows: (i) the earliest occupation was Neolithic; there were no Paleolithics; (ii) there was a substantial colonization both by Phoenicians and by the Phoenician colonies in Africa (Punics and Carthaginians) (*15*). The genetic evidence from the map agrees with the idea that part of the gene pool of Sardinia owes its origin to direct or indirect immigrations from the Middle East, with little admixture with other populations of European origin, thus supporting the theory of Phoenician and Punic immigration; it does not deny the importance of the contribution from southern Italy, from which the first farmers probably came.

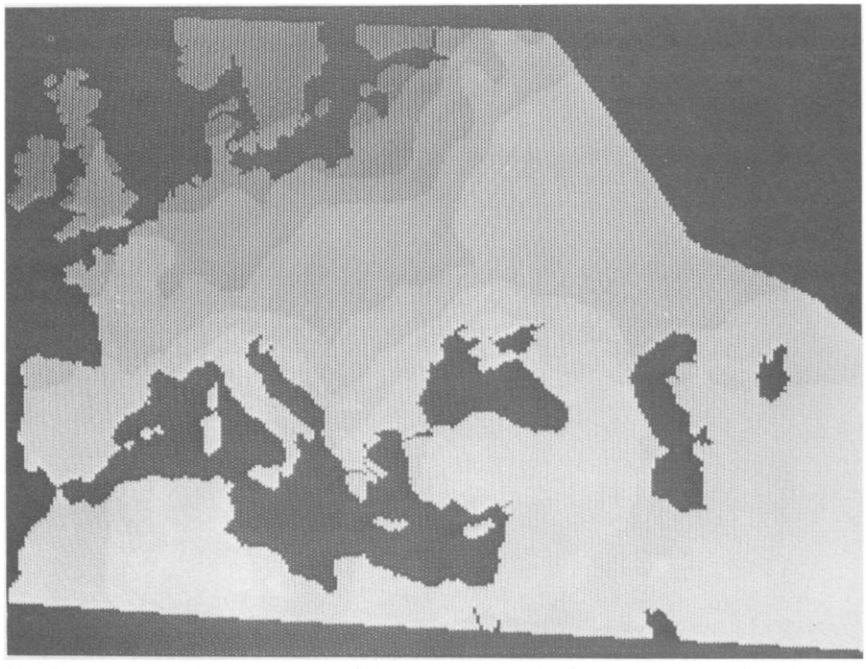

Fig. 5. The first principal component of 21 independent alleles from the human loci HLA-A and HLA-B.

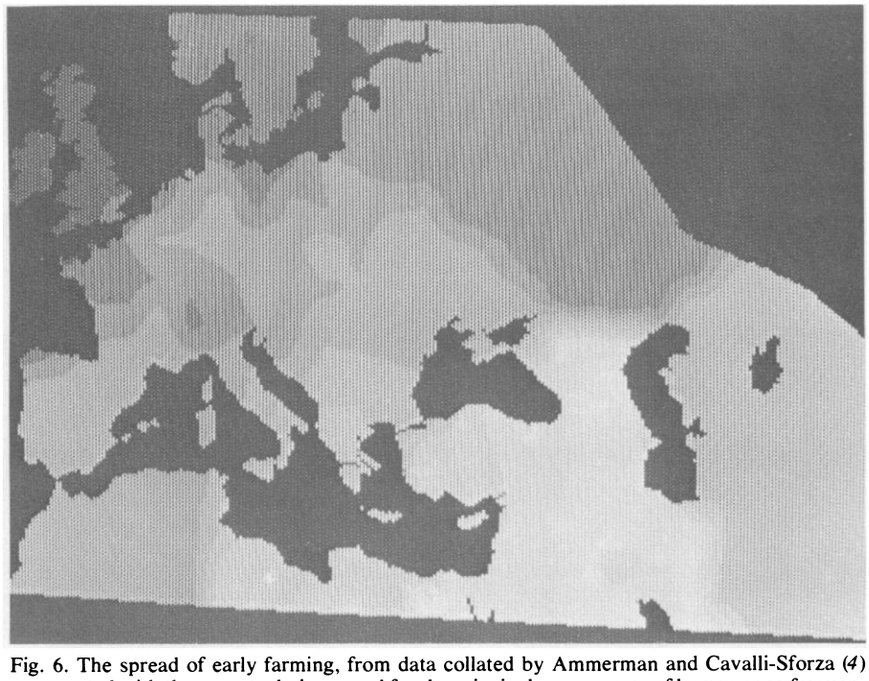

Fig. 6. The spread of early farming, from data collated by Ammerman and Cavalli-Sforza (*4*) represented with the same technique used for the principal components of human gene frequencies. The lightest shades correspond to areas in which early farming was present in year 8500 before present (uncorrected radiocarbon dates): the next shades correspond to 8000 to 8500, 7500 to 8000, 7000 to 7500, 6500 to 7000, 6000 to 6500, 5500 to 6000, ≤5500 periods of earliest occupation in years before present. Note the similarity with the first principal component of gene frequencies. Unlike maps published earlier (*4*, *6*) this map has included dates for Switzerland which was occupied somewhat later than the residual country around.

It is of interest to note which genes contribute most to the principal component values. Correlations of gene frequencies and the principal components are given in Table 2. Some genes are found to contribute to more than one principal component, as for instance genes B, Rh−, and Hp which contribute to both the first and second principal component. In the case of Rh−, for instance, this behavior is simply explained by the fact that the concentration of Rh− genes was probably very high—maybe close to 100 percent—in the hunter-gatherers of western Europe, including probably the British Isles, whereas positive genes arrived both from the east and the southeast, perhaps in different migrations.

The 38 alleles analyzed in Figs. 1 to 3 and on the cover are made of two groups of almost equal size: 17 non-HLA alleles and 21 HLA alleles. These groups were analyzed separately by the same techniques. The first principal component of each of the two groups is shown in Figs. 5 and 6. The non-HLA alleles show in the first principal components an admixture of east-west and the northwest-southeast cline; in the second (not shown) the pattern is very similar to the third principal component of the 38 alleles (Fig. 3). The third principal component of the non-HLA alleles (not shown) is characterized by a more complex pattern, with peaks in the Caucasus, Iran, and Algeria, of difficult interpretation.

The 21 HLA alleles show in their first principal component the same pattern as the first such component of the global analysis (Fig. 1), namely the same cline as the spread of farmers (Fig. 4). We do not give the figures of the second principal component which shows again an east-west gradient, and of the third component which repeats the northwest-southeast cline. On analysis of the two subsets of HLA genes, gene HLA-A shows the east-west pattern in its first principal component and the northwest-

southeast in the second, whereas the opposite is true of the HLA-B gene.

Thus there are some clearly recurring patterns, which are found more clearly in HLA but also in non-HLA genes: (i) the dominant one is a cline centered in the southeast of the map which parallels very closely the gradient of diffusion of early farming; (ii) an east-west cline is also clear cut; (iii) less clear but not negligible is another cline whose axis is opposite to that of the spread of agriculture, and seems to have its center in the Ukraine.

Discussion and Conclusions

It is intuitively clear that the synthetic variables give us information on the evolutionary history of the populations concerned. But it is also clear that they cannot tell us whether a cline is due to the existence of different selective forces at the extremes of the cline (migration being responsible for the gradual variation between the two extremes) or to an ancient difference between two formerly isolated populations (now disappearing under the influence of migration), or finally to the geographic spread of a population which has undergone an expansion within a not too distant past. The three explanations correspond respectively to (i) an equilibrium due to the balance between selection and migration, (ii) a transient cline disappearing under migration, and (iii) the consequences of an episode of demic diffusion. In a similar vein, similarities between populations might be due to environmental similarity or recent admixture; differences may be the outcome of drift or selection. Study of the synthetic variables alone does not permit a choice between these possibilities. This kind of uncertainty is characteristic, of course, of inference on historical events on the basis of present time data, which can sometimes be resolved by using other existing sources of information.

In our present case, however, we were led to look at a particular part of the world map with a specific hypothesis in mind. This hypothesis was generated from data of archeological studies, and was therefore of entirely different origin. The source of our curiosity was the history of expansion of early farming from the Near East to Europe, replacing earlier techniques of food production. Much of the archeological evidence (especially in well investigated areas like that occupied by the Bandkeramik in south central and north central Europe) speaks in favor of a demic, rather than a cultural diffusion. One could almost say that a gradient of genes like the observed one was expected, but it was not known if the conditions for its presence and its persistence were met.

The patterns observed in the synthetic map indicate various possible centers of radiation, that of the Near East being the dominant one. A diffusion from the East and one from a region approximately corresponding to the northern part of modern Ukraine are two other independent ones. Further work and especially more data from the eastern part of the area would be necessary to assure that these two last migration patterns are truly distinct. It is not surprising that the diffusion pattern connected with the spread of early farming is the dominant one, even if it is relatively ancient, and perhaps older than those connected with the other patterns. The chance of observing at the genetic level a mass migration depends among other things on how massive the migration is in terms of the ratio of the numbers of colonists to those of the people who lived earlier in the area. It also depends on patterns of differential reproduction of coexisting groups. Of all known migrations, that connected with early farming is likely to have been among the most important ones in terms of the ratio of colonizers to colonized, given that the transition to agriculture creates the potential for a major increase in population density (by more than one order of magnitude). By contrast, the invasions of "barbarians" of central and eastern Europe are poorly known and archeologically difficult to trace. Those that took place in historical times did not usually involve numbers large enough to influence gene frequencies strikingly. Thus, even though it is tempting to assume that the other clines were also associated with demic spreads it is difficult to link them unequivocally with any particular one.

An interesting negative conclusion is the absence, among the first three principal components, of north-south clines. These could have been expected under selective response to climate, provided a sufficient number of genes showed such a response.

It is worth extending the method to the other situations as a way of detecting sources of population spreads. Other directions of spread from the Near East, the Bantu expansion of Africa, and spreads from other centers of origin of agriculture in Asia and the Americas are natural candidates. The necessity of an adequate number of markers is, however, to be kept in mind. Europe is at present a unique area of the world in this respect, having been the subject of many archeological and genetic investigations. The study of gene distribution in this area had already been approached with another method (*16*) which had given similar results, though not as clear as

Table 1. Percentage of total variance explained by first three principal components of the gene frequency data for Europe and the Near East.

Principal component	All 38 alleles	HLA alleles	HLA-A alleles	HLA-B alleles	17 non-HLA alleles
First	27.0	31.8	35.3	37.9	32.7
Second	18.4	12.0	18.3	13.7	15.9
Third	11.0	11.1	15.5	10.6	9.0
Total	56.4	54.9	69.1	62.2	57.6

Table 2. Relative importance of various alleles in determining the first three principal components of Europe and the Middle East, as determined by absolute values of correlation coefficients with principal components.

Alleles having correlation values (absolute)		
0.8 to 1	0.6 to 0.8	0.4 to 0.6
	With first principal component	
HLA-B5, HLA-B7	HLA-A2, -A3, -A11; HLA-B8, -B12, -B27, -B15, -B35, -B40; Hp; cde; B	HLA-A1, -A9, -B21; P^b; Ms, Ns
	With second principal component	
Fy^a	HLA-A28, -B14; Hp; CDe, cde; B; 0	HLA-A2, -A11, -B18, -B21; cDE, cDe; Ms, Ns
	With third principal component	
	P^a, P^b	HLA-A1, -A3, -A9, -A10, -B13; PGM^1; A_1

those here reported. In this work one finds also a kinetic analysis of admixture, which provides ball-park estimates of rates of growth, migration, and "acculturation" (hunter-gatherers entering the farmers' pool) compatible with observations.

References and Notes

1. J. B. S. Haldane, *Ric. Sci* **19**, 3 (1949); A. C. Allison, *Trans. R. Soc. Trop. Med. Hyg.* **48**, 312 (1954).
2. L. L. Cavalli-Sforza and W. F. Bodmer, *The Genetics of Human Populations* (Freeman, San Francisco, 1971); L. L. Cavalli-Sforza, *Am. J. Hum. Genet.* **25**, 82 (1973).
3. C. R. Rao, *Advanced Statistical Methods in Biometric Research* (Wiley, New York, 1952); L. L. Cavalli-Sforza and A. W. F. Edwards, in *Genetics Today*, S. J. Geerts, Ed. (Pergamon, New York, 1965), vol. 3; P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy* (Freeman, San Francisco, 1973).
4. A. Ammerman and L. L. Cavalli-Sforza, *Man* **6**, 674 (1971); ———, in *The Explanation of Culture Change*, C. Renfrew, Ed. (Duckworth, London, 1973).
5. A. E. Mourant, *The Distribution of the Human Blood Groups* (Blackwell, Oxford, 1956).
6. L. L. Cavalli-Sforza, *Proc. R. Anthropol. Inst. G. B. Ireland* (1972), p. 15.
7. A. E. Mourant, A. C. Kopec, K. Domaniewska-Sobczak, *The Distribution of the Human Blood Groups* (Oxford Univ. Press, Oxford, 1976).
8. L. P. Ryder, E. Anderson, A. Svejgaard, *Hum. Hered.* **28**, 171 (1978).
9. A. Piazza, P. Menozzi, L. L. Cavalli-Sforza, in preparation.
10. We used the program Grimage developed by R. Lyon and his associates at Stanford University.
11. Altogether 67 populations were to be used and listed in (9).

→ L. Cavalli-Sforza and A. Piazza, *Theor. Pop. Biol.* **8**, 127 (1975).

13. A. Piazza, L. Sgaramella-Zonta, P. Gluckman, L. Cavalli-Sforza, *Tissue Antigens* **5**, 445 (1975).
14. D. F. Roberts, in *Genetic Variation in Britain*, D. F. Roberts and E. Sutherland, Eds. (Taylor & Francis, London, 1973).
15. F. Barreca, *La Sardegna Fenicia e Punica* (Chiarella, Sassari, 1974).
16. L. Sgaramella-Zonta and L. Cavalli-Sforza, *Proceedings of a Workshop on Population Structure* (Hawaii Univ. Press, Honolulu, 1973), pp. 128–135; see also other unpublished results by the same authors.

17. Work supported by grants NIH GM 20467, 20832, ERDA EY 76 S 03 0326, and by CNR, Centro di studio per l'Immunogenetica e l'Istocompatihilita, Turin, Italy. We thank R. Lyon for letting us use the TV color screen and program he has developed for use in analysis of satellite images for geology. The color TV system is a gift to the Earth Sciences Department of Stanford University by the Kerr McGee Corp. of Oklahoma.

Present address of P. Menozzi: Instituto di Ecologia, Universita di Parma, Italy; of A. Piazza, Instituto de Genetica Medica, Universita di Torino, Italy.

NEWS AND COMMENT

Report of Fusion Breakthrough Proves to Be a Media Event

On the weekend of 12 and 13 August, the unlikely subject of fusion suddenly became the leading news story in the country.

"Scientists at Princeton University have produced a controlled thermonuclear fusion reaction that experts are hailing as a major technical breakthrough," said the Knight-Ridder wire service in a story that was carried by 50 to 100 newspapers. "U.S. Makes Major Advance in Nuclear Fusion," was the banner headline of the Washington *Post*'s leading front-page story on Sunday morning. Radio and TV stations throughout the weekend reported the story with all the urgency of an international crisis, and by the end of the 2-day media blitz, many citizens apparently got the impression that after years of waiting for proof, fusion had finally been achieved. The message was so strong and so positive that it seemed—for 48 hours at least—that the energy crisis was over, solar energy and nuclear power were no longer needed, and that the future would be assured through fusion.

The heady optimism did not last long. By Monday afternoon, the Department of Energy, which had sponsored the Princeton research, was saying that no breakthrough had occurred, and that the results, while "gratifying," would make no change in the timetable or the funding for government fusion research, which is expected to require at least 50 more years to bear fruition. John Deutch, of the Department of Energy, said that the Princeton result came "sooner and in stronger form than we anticipated," but he characterized it only as "an item that bears on the first step" of a lengthy, costly, technically demanding development process.

One of the principal reasons for confusion was that the reports of the weekend had seemed to indicate that fusion had reached the long-awaited goal of energy breakeven—the point where a reacting fusion plasma produces more energy than it consumes. But the head of the Princeton laboratory, Melvin Gottlieb, said that the experiment in question had not made breakeven. However, he told a Washington press conference on Monday afternoon, 14 August, that "we're on schedule and I'm confident we will achieve breakeven" with a larger experiment due to begin operation in the 1980's.

What actually happened at Princeton that garnered so much attention? It was the dramatic conclusion of a sometimes discouraging experiment with a 3-meter diameter doughnut-shaped device called a tokamak, which can serve as a type of "magnetic bottle" for containing fusion reactions. The device, named the Princeton Large Torus, ran into severe engineering difficulties soon after it was built in 1976, but by spring of this year it was working well and by summer it was producing the highest temperature ever recorded for a tokamak. That temperature, according to Harold Eubank who conducted the experiment along with Walter Stodiek, was 50 to 55 million degrees Celsius, about six times higher than the temperature in the fusion experiment that had come closest to breakeven, one carried out with the Alcator tokamak at MIT in 1976. In order to get such a high temperature, however, Eubank and Stodiek had to lower the density of the plasma in their experiment. Temperature, density, and the length of time the plasma is contained are all important in magnetic fusion experiments and must simultaneously meet certain criteria for a self-sustained reaction to be achieved. Although the temperature was six times higher than MIT's, the combined measure of density and confinement time (which was 15 thousandths of a second) gave a value, Eubank told *Science*, which was 30 times worse than that attained in the MIT experiment. These results were obtained in July.

The significance of the Princeton result was not that it came close to breakeven, because MIT had improved somewhat on its 1976 result and still held the nearness-to-breakeven record. Rather, it was that in reaching such a high temperature the Princeton experiment had entered a plasma regime where wild fluctuations, called "trapped ion instabilities," were expected to degrade the confinement properties of a tokamak. These fluctuations had been earmarked by many in the fusion program as the biggest unresolved physics question that stood in the way of the development of tokamaks, which have been the leading candidates among various types of magnetic bottles since soon after they were invented in Russia in 1968. No evidence of the predicted fluctuations was found.

0036-8075/78/0901-0792$00.50/0