Am J Hum Genet 29:169–177, 1977

Population Genetics in the Province of Ferrara. I. Genetic Distances and Geographic Distances

P. Zanardi,[1] G. Dell'Acqua,[1] C. Menini,[2] and I. Barrai[1]

INTRODUCTION

The province of Ferrara (fig. 1), in particular the area between the town of Ferrara and the Adriatic, had been marshy and malarious for several centuries before World War II, after which time the marshes were drained. The population has a high frequency of alleles correlated with anomalies of hemoglobin production; variants of hemoglobin structure and red cell enzymes are common.

The most frequent trait determined by such genes is β^0 thalassemia. The high frequency of thalassemia in the province of Ferrara [1] is believed to be the result of selection in favor of the heterozygous genotype for the thalassemia gene. The fitness of the heterozygous genotype under malaria has been estimated to be 8% above the fitness of the genotype homozygous for the normal production of hemoglobin [2], using the distribution of gene frequencies in samples collected in the 26 residential units (comuni) of the province. The gene for β^0 thalassemia has an overall frequency of 6.5 ± 2.3%. The allele for β^0 thalassemia, which is almost invariably lethal in homozygotes, is thought to be the first to be selected by malaria. The other genes associated with malaria, notably the gene for G6PD deficiency, are in subpolymorphic frequencies [3].

The genes for the ABO and Rh blood group systems do not seem to be correlated with any of the genes associated under malarial conditions, and the selective advantage they confer in the heterozygous state is probably much lower than the advantage conferred by a single gene for thalassemia in a malarious environment [4].

The objective of the present work is the detection of the effect of genetic drift on the differentiation of gene frequencies for blood group alleles in 26 residential units in the province of Ferrara, assuming that for these systems the effect of selection is so small that it does not mask the effect of drift. Migration, another major force which opposes the effect of drift, reduces the variance of gene frequencies between subpopulations.

In the province of Ferrara from 1960 to 1971, the immigration rate was 2.6 ± 0.2% and the emigration rate was 3.7 ± 1.1%. These figures include the movement from one

---

Received August 4, 1976; revised October 27, 1976.

This work was supported in part by a grant from the World Health Organization to the Department of Zoology of the University of Ferrara.

[1] Department of Zoology, University of Ferrara, Ferrara, Italy

[2] Blood Bank, General Hospital, Ferrara, Italy.

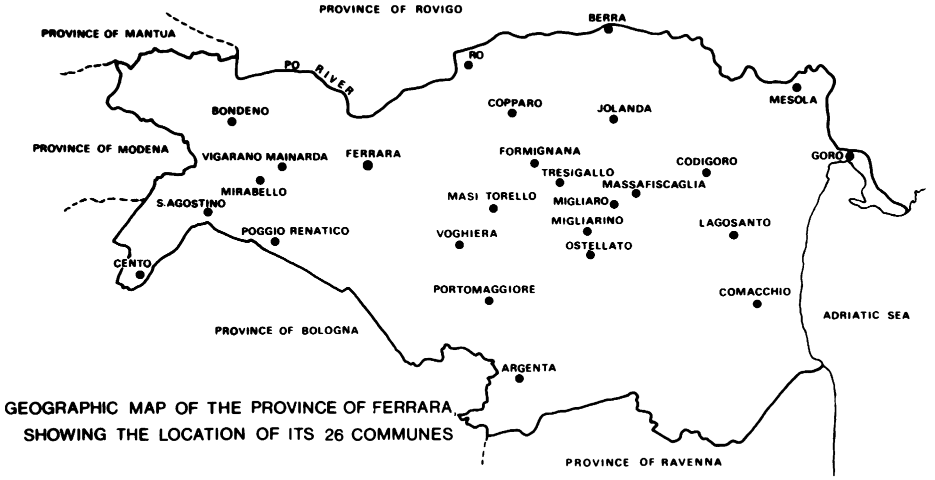

FIG. 1.—Geographic map of province of Ferrara showing location of 26 communes.

residential unit to another inside the province, which is about 80% of the present immigration rate.

Since the population has been growing from 1860 to 1950 (with a slow decrease after 1951) and since we have at present no data available to estimate equilibrium conditions of migration rates, in this paper we investigate the gene frequency variation as a function of genetic distance between residential units and the geographic distance between them. We hypothesize that, if some degree of isolation has persisted until recent times in the subunits of the population, genetic drift and other factors may have had an effect on the differentiation of gene frequencies. Under such an hypothesis, and assuming that inside the province there is an unimodal distribution of migration distances, the differentiation would be greater between units which are farther apart. We would, therefore, expect a significant correlation between indicators of genetic distances between places and measures of geographic distances.

MATERIALS AND METHODS

From the blood bank of the General Hospital, we obtained the blood groups of the ABO system of 8,255 donors from 26 communes. The Rh groups, using only anti-D, were obtained for 8,700 donors from the same communes; the phenotype frequencies are given in table 1. The chi-squares for association between phenotype frequencies and communes are not significant for ABO or Rh. One might, therefore, infer that migration has an important equalizing effect for the gene frequencies in these systems. However, our purpose is not to measure significant heterogeneity, but given a difference between frequencies among villages, we want to measure its extent and whether the genetic distance resulting from it is correlated with geographic distance. Such a correlation would indicate that migration has not canceled the effects of drift.

The gene frequencies for the alleles of the ABO system were computed with the inefficient method of Bernstein, and the square root method was used to estimate the frequencies of D and d in the Rh system. The estimates obtained for the 26 communes are given in table 2. Although the variation is not significant, there is a considerable variation of frequencies for the various alleles. For allele A, the range is .21–.33; for B, .04–.14; and for O, .58–.73; the minimum frequency of the D allele in the Rh system is .61 and the maximum, .74. The gene frequencies were used to compute indicators of genetic distance according to the methods of Cavalli-Sforza and Morton [5, 6].

TABLE 1

PHENOTYPE FREQUENCIES OF ABO AND RH SYSTEMS IN 26 COMMUNES OF THE PROVINCE OF FERRARA

	PHENOTYPES							
COMMUNE	A	B	O	AB	TOTAL	RH(+)	RH(−)	TOTAL
Argenta	215	44	212	14	485	433	52	485
Berra	131	39	145	13	328	298	30	328
Bondeno	180	50	201	23	454	414	40	454
Cento	105	22	111	9	247	210	37	247
Codigoro	164	25	193	14	396	356	40	396
Comacchio	67	10	81	13	171	150	21	171
Copparo	271	54	245	28	598	537	61	598
Ferrara	1,172	280	1,148	103	2,703	2,430	273	2,703
Formignana	74	21	71	4	170	171	29	200
Goro	30	9	27	2	68	84	11	95
Jolanda	91	22	74	10	197	174	23	197
Lagosanto	27	5	27	6	65	111	10	121
Masi Torello	6	3	8	1	18	24	3	27
Massa Fiscaglia	98	29	99	18	244	225	19	244
Mesola	71	16	106	9	202	186	16	202
Migliarino	119	21	87	12	239	253	33	286
Migliaro	8	4	7	2	21	21	2	23
Mirabello	62	8	45	6	121	119	21	140
Ostellato	145	38	143	8	334	364	33	397
Poggio Renatico	42	5	36	3	86	140	13	153
Portomaggiore	210	54	192	24	480	420	60	480
Ro Ferrarese	35	10	38	4	87	82	5	87
Sant'Agostino	46	11	50	6	113	137	23	160
Tresigallo	58	9	61	5	133	123	10	133
Vigarano Mainarda	79	14	89	10	192	179	27	206
Voghiera	49	7	44	3	103	145	22	167
Province of Ferrara	3,555	810	3,540	350	8,255	7,786	914	8,700

The genetic distance according to the method of Cavalli-Sforza is a function of the angle between population points in a multidimensional space where the coordinates are the gene frequencies. The distance d_{ij} is measured in terms of θ, the angle, and the frequencies of individual alleles. For any two population groups i and j, the distance is $d_{ij} = \sqrt{1 - \cos\theta}$ where $\cos\theta = \sqrt{p_{1i}\,p_{1j}} + \sqrt{p_{2i}\,p_{2j}} \ldots + \sqrt{p_{ki}\,p_{kJ}}$, where the products under the square root refer to the product of the frequencies of the same allele in the two different groups. This formulation of distance, called the euclidean distance, disregards a constant and can be combined over several loci [5]. It is an appealing indicator of divergence between groups because it can be referred to as a metric of allele substitutions or a function of allele substitutions [6].

We believe that, since euclidean distance and other indicators of divergence are strictly correlated [7], most of these indicators are useful descriptions of genetic distance or of genetic similarity between groups. However, we calculated the sample kinship according to the method of Morton [6], which is as widely used as the method of Cavalli-Sforza [5]. Following Workman [8] and using sample kinship, we avoid the interpretation of kinship as a probability of equality by descent, which is complicated by the fact that kinship often assumes negative values.

The sample kinship is the scaled gene frequency covariance, which is defined by

$$r_{ij} = \frac{(p_{ik} - \bar{p}_k)\,(p_{jk} - \bar{p}_k)}{\bar{p}_k\,(1 - \bar{p}_k)},$$

TABLE 2

GENE FREQUENCIES OF ABO AND RH SYSTEMS IN 26 COMMUNES IN PROVINCE OF FERRARA

	ABO			RH	
COMMUNE	p	q	r	D	d
Argenta	.2766	.0652	.6581	.6726	.3274
Berra	.2522	.0840	.6637	.6976	.3024
Bondeno	.2514	.0784	.6702	.6822	.3178
Cento	.2650	.0635	.6716	.6130	.3870
Codigoro	.2522	.0440	.7038	.6822	.3178
Comacchio	.2455	.0418	.7126	.6496	.3504
Copparo	.2894	.0672	.6434	.6806	.3194
Ferrara	.2745	.0751	.6504	.6822	.3178
Formignana	.2755	.0888	.6356	.6192	.3808
Goro	.2836	.0968	.6196	.6597	.3403
Jolanda	.3022	.0852	.6126	.6583	.3417
Lagosanto	.2712	.0580	.6708	.7125	.2875
Masi Torello	.2156	.1152	.6692	.6667	.3333
Massa Fiscaglia	.2634	.0879	.6486	.7209	.2790
Mesola	.2129	.0530	.7341	.7186	.2814
Migliarino	.3255	.0690	.6055	.6603	.3397
Migliaro	.2689	.1470	.5840	.7051	.2949
Mirabello	.3318	.0522	.6160	.6127	.3873
Ostellato	.2728	.0814	.6458	.7117	.2883
Poggio Renatico	.3060	.0436	.6504	.7085	.2915
Portomaggiore	.2829	.0835	.6336	.6464	.3535
Ro Ferrarese	.2554	.0820	.6626	.7603	.2397
Sant'Agostino	.2576	.0698	.6725	.6208	.3791
Tresigallo	.2694	.0484	.6822	.7258	.2742
Vigarano Mainarda	.2562	.0519	.6918	.6380	.3620
Voghiera	.2966	.0501	.6533	.6370	.3629
Province of Ferrara	.2725	.0711	.6564	.6759	.3241

where p_{jk} and p_{ik} are the frequencies of allele k in groups j and i, respectively. In the present case, the groups are the communes of the province and $\bar{p}_k$ is the regional average of the same allele; in the present analysis, the frequency over all the province. The values of r_{ij} are averaged over all different alleles and loci [9]. We note that sample kinship is homologous to the correlation between uniting gametes of S. Wright, which although related to inbreeding, is not a probability. When it is negative, it means that the two populations' F_1 would produce an excess of heterozygotes. A negative sample kinship means that there would be an excess hybridity between the two populations should these produce a common offspring population. The 26 × 26 distance and kinship matrices of the province may be obtained from the authors.

RESULTS AND DISCUSSION

The values of distance or kinship constitute a symmetrical matrix with a 26 × 26 dimension which poses obvious problems for study and interpretation. Methods which visually represent the distance between different communes, such as the bidimensional projection of the matrices using principal components analysis [10] and the dendrograms constructed according to the technique of the unweighted average pair-group [11], are the most useful. The results of both methods (figs. 2 and 3) are essentially identical.

It makes little difference on the bidimensional projection whether distance or kinship

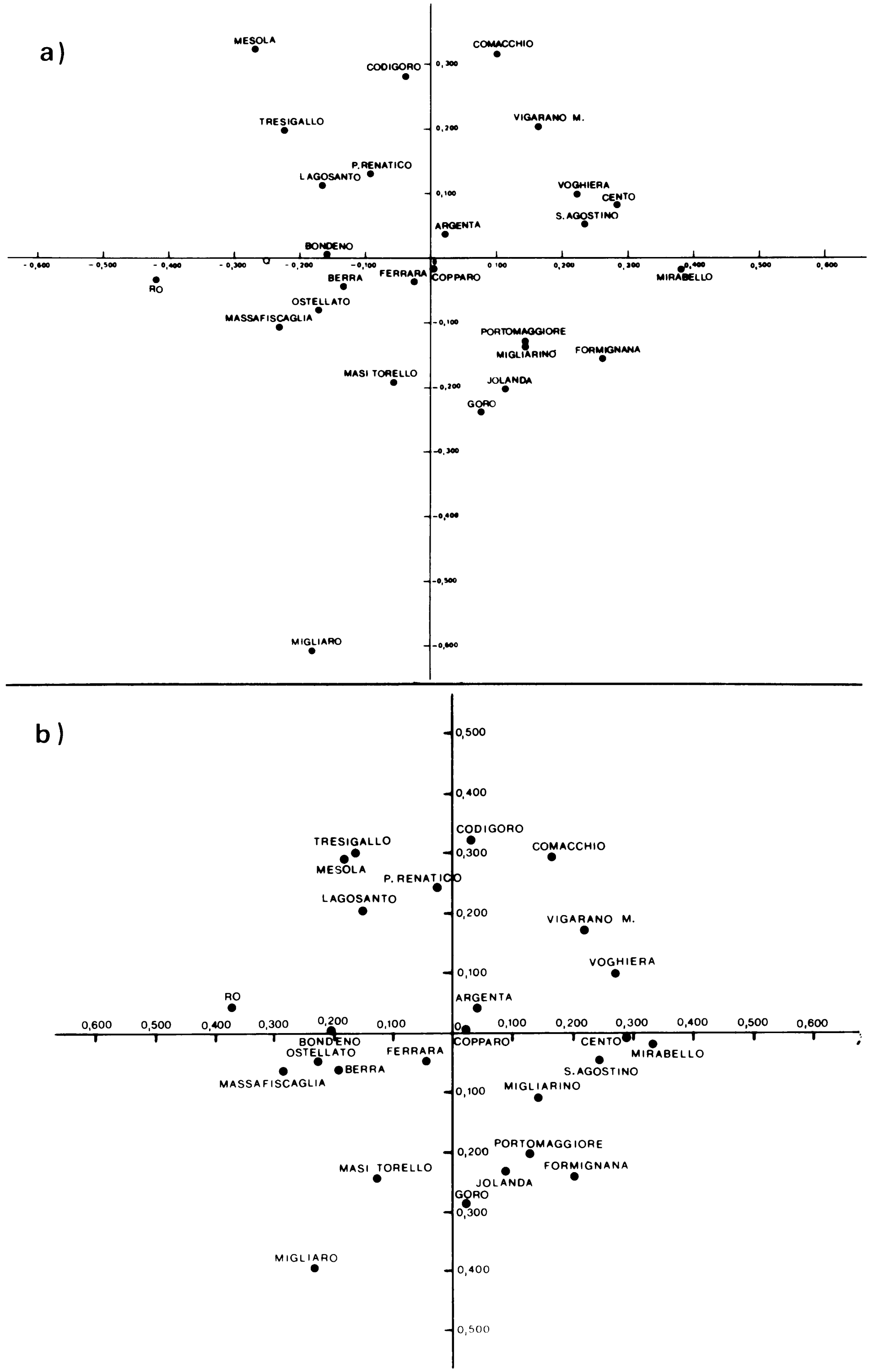

FIG. 2.—Genetic map of 26 communes obtained by bidimensional projection of matrix of sample kinship (*a*) and matrix of euclidean distances (*b*).

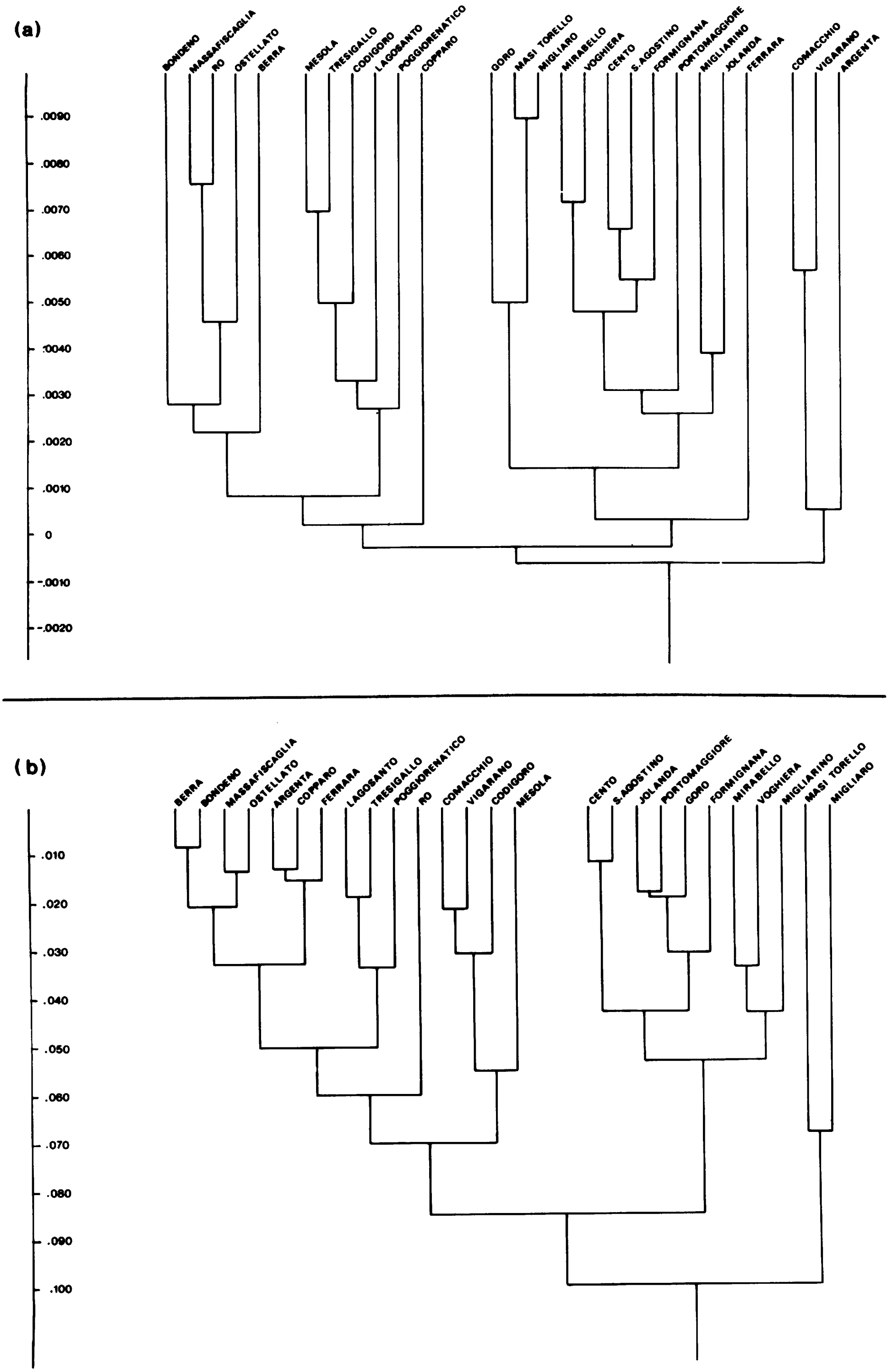

FIG. 3.—Dendrograms of the communes of the province of Ferrara: *a*, based on kinship; *b*, based on euclidean distance between communes.

are used to extract the coordinates of the points on the plane. The ratio of the first two eigenvalues over the sum of eigenvalues is .46 for the distance matrix and .83 for the kinship matrix. Both matrices were centroid adjusted according to a technique suggested by Morton and Lalouel [10]. After rotation, the projections of kinship and distance are almost superimposable (the projection of the distance matrix is rotated 180° on the main axis in fig. 3). The dendrograms, although similar, are not as similar as the projections. There are two main dichotomies in both representations, and the same groups of communes belong to each in the two dendrograms; however, the relative positions on the trees are not respected. For example, although the two communes of Masi Torello and Migliaro, belong to the same dichotomy in the two trees, they seem closer on the tree of kinship than on the tree of distance. The trees appear to be more sensitive to small variations of quantities in matrices than the projections. However, since distance and sample kinship have different metrics, the observed effect might be due to the nonlinear part of the correlation between kinship and distance. The coefficient of correlation between the two measures is negative as expected ($r = -.72$). This is a high value, which indicates that there is substantial similarity between the two indicators.

The most interesting results from testing the hypothesis are the coefficients of correlation between kinship and geographic distance (say r_{kd}) and between genetic distance and the same geographic distances (say r_{dd}). We have obtained $r_{kd} = -.23$; $t = -4.33$ and $r_{dd} = .26$; $t = 4.93$. Both indicators correlate equally with geographic distance, and the correlations are highly significant, with the sign expected for each case. Since sample kinship measures a similarity, it is negatively correlated with distance, and the euclidean distance is positively correlated with geography. This correlation indicates that even if drift is no longer active in differentiating gene frequencies in this area it was probably active in the past, and its effect can still be perceived with this method.

A clearer picture may emerge when indicators of genetic distance based on genes with strong selective advantage are correlated with geography in the same population. At this stage, however, the correlation between genetic and geographic distance is high. Since the correlation is high, we fit the model of isolation by distance to our data on sample kinship in the province of Ferrara. Although, as noted earlier, the sample kinship is not, in general, a probability, it may be used to estimate a priori kinship by fitting the equation: $r_{ij} = (1 - L)\ \varphi_{ij} + L$, where $\varphi_{ij} = ae^{-bd}$, according to Malécot's model of exponential decay of kinship with geographic distance. In this model, a is the a priori kinship at distance zero, for the same commune in our case, b is the rate of exponential decay, and d is the distance in kilometers between two communes. The value of L is a measure of the deviation of contemporary gene frequencies from the founder gene frequencies. To estimate a, b, and L, we minimized the function

$$Q = \sum_i \sum_j \left\{ r_{ij} - [(1 - L)ae^{-bd} + L] \right\}^2$$

using the Newton-Raphson method and numerical derivatives.

In table 3, the results of the simultaneous estimation of the parameters, which are compared with the values obtained in other European regions [12], are given.

TABLE 3

ISOLATION BY DISTANCE IN FERRARA PROVINCE, SWITZERLAND, AND BELGIUM

Ecology	Source	*a*	*b*	*c*
Ferrara Province	ABO AND RH	.0058	.0261	−.0029
Switzerland	ABO	.0025	.0185	−.0006
Belgium	ABO	.0009	.0247	−.0001

The a priori kinship for two random individuals from the same commune is estimated as $a = .0058$, with an exponential decline of .0261 per kilometer. The exponential decline is close to the one observed for Switzerland and Belgium. Random individuals from Ferrara have a kinship, relative to their founders, of $\varphi_R = .0029$, which is five times higher than random Swiss and about 30 times higher than random Belgians.

From the a priori kinship, it is possible to estimate the expected value of hybridity between two random persons born at distance d in an area, which is

$$\Theta_d = \frac{\varphi(0) - \varphi(d)}{2 - \varphi(0) - \varphi(d)} .$$

Estimates of hybridity would be optimal to test hypotheses on biological indicators in the progeny of any couple born at distance d. However, geographic distance between birthplaces of any two persons can be obtained routinely, whereas estimates of hybridity cannot. Since geographic distance is correlated with genetic distance, we may assume that distance between birthplaces of spouses might be used as an indicator, although not a very strong one, of the hybridity of their offspring; this would seem valid for the Ferrara area at least.

Should similar studies of correlation between genetic and geographic distance show significant association, geographic distance between birthplace of spouses might be used to test genetic hypotheses on biological indicators in the progeny.

SUMMARY

The ABO and Rh systems of the population in 26 residential units in the province of Ferrara were studied to detect the effect of genetic drift on the differentiation of gene frequencies. Results obtained from principal components analysis and dendrograms were similar.

A significant correlation between genetic and geographic distance was found. It was therefore hypothesized that in the area of Ferrara children born to parents originating from different communes were heterozygous at more loci than children born to parents originating from the same commune. It was also thought that it may be possible to use geographic distance between birthplaces of parents as an indicator of hybridity in their children.

REFERENCES

1. Silvestroni E, Bianco I: Screening for microcytemia in Italy: analysis of data collected in the past 30 years. *Am J Hum Genet* 27:198–212, 1975
2. Barrai I, Mandel SPH: Equilibrium under inbreeding in balanced polymorphisms. *Bull WHO* 45:201–208, 1971
3. Gandini E, Menini C, De Filippis A, Dell'Acqua G: Erytrocytary glucose-6-phosphate dehydrogenase deficiency. *Acta Genet Med Gemellol (Roma),* 3:271–284, 1969
4. Chung CS, Morton NE: Selection at the ABO locus. *Am J Hum Genet* 13:9–27, 1960
5. Cavalli-Sforza LL, Edwards AWF: Phylogenetic analysis models and estimation procedures. *Am J Hum Genet* 19:233–257, 1967
6. Morton NE: *Genetic Structure of Populations*. Honolulu, Univ. of Hawaii Press, 1973
7. Barrai I: Indicators of genetic distance, in *Proceedings 4th International Congress Human Genetics,* edited by Crow J, Denniston C, Amsterdam, Excerpta Medica, 1973
8. Workman P, Harpending H, Lalouel J, Lynch J, Niswander J, Singleton R: Population studies in southwestern Indian tribes. *Population Genetics Monograph* 3:166–194, 1973
9. Harpending H, Jenkins T: Genetic distance among southern African populations, in *Methods and Theories in Anthropological Genetics*. Albuquerque, Univ. of New Mexico Press, 1971
10. Morton NE, Lalouel JM: Topology of kinship in Micronesia. *Am J Hum Genet* 25:422–432, 1973
11. Sokal PR, Sneath PHA: *Principles of Numerical Taxonomy*. San Francisco, W. H. Freeman, 1968
12. Morton NE: Population structure and historical genetics of isolates. *Isr J Med Sci* 9:1299–1307, 1973