Amer J Hum Genet 25:347–361, 1973

Genetic Structure of Switzerland

N. E. MORTON,[1] D. KLEIN,[2] I. E. HUSSELS,[3] P. DODINVAL,[4] A. TODOROV,[5] R. LEW,[1] AND S. YEE[1]

Four years have passed since the first study of isolation by distance in human populations, using ABO blood groups in Switzerland [1]. Since then phenotype bioassay has been improved [2], new data have been collected [3, 4], and new methods have been developed to study population structure [5]. The time seems ripe for a synthesis.

Our data consist of four sources: (1) Rosin's report on the ABO blood groups of 275,664 Swiss army personnel and air raid wardens from 1940 through 1945, grouped according to 3,101 communes of origin [6]; (2) genealogies of Saas, an Alpine isolate [7]; (3) federal census data on each commune [8]; and (4) pedigrees of retinal degeneration, hemophilia, and myotonic dystrophy ascertained through the Institut de Génétique Médicale, Geneva [9–11].

Switzerland was selected for this study because of the high quality of its records and the belief that it is typical of developed populations with residual isolates. Rosin's study is remarkable for its size and the partition of his sample into communities. Saas seems representative of Alpine isolates, and its parish registers published by Zurbriggen [12] provide a genealogy since the sixteenth century. The three diseases were selected because they exemplify different modes of inheritance with a high level of ascertainment for the whole country. Retinal degeneration has an appreciable recessive component, unlike hemophilia and myotonic dystrophy, which serve as a control. Undoubtedly other regions could provide as extensive data on genetic structure of the population, but rarely in such accessible detail.

THE ABO BLOOD GROUPS

Each Swiss has one or more communes of origin, where the vital statistics of his family are kept. The commune of origin may coincide with his birthplace, but

Received August 2, 1972; revised February 26, 1973.

PGL paper no. 95. This work was supported by grant GM 17173 from the U.S. National Institutes of Health.

[1] Population Genetics Laboratory, University of Hawaii, Honolulu, Hawaii 96822.

[2] Institut Universitaire de Génétique Médicale, Hôpital Cantonal, 1206 Genève 8, chemin Thury, Switzerland.

[3] Division of Medical Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205.

[4] Service de Génétique Humaine, Université de Liège, Liège, Belgium.

[5] Université de Genève, Genève, Switzerland.

often it represents the birthplace of his father or paternal grandfather or more remote male ancestor in the paternal line. It is possible although not common for a foreigner, or even a Swiss, to acquire a commune of origin in which his family has not been resident. Men ordinarily take their father's commune of origin, unless they elect place of residence as a new commune of origin. In that case the previous origin may be retained or abandoned. Individuals with two or more communes of origin were assigned the principal or preferred affiliation in Rosin's study and the most historic or isolated in ours.

Variation of gene frequencies among the more than 3,000 communes may be described by a few parameters, which are provided by Malécot's theory of isolation by distance. The expected value of kinship between two random individuals at distance d is

$$\phi(d) = ae^{-bd} \tag{1}$$

relative to indefinitely large distances, where kinship is the a priori probability that two genes are identical by descent [13]. This formula includes linearized selection as well as migration and mutation.

The Euclidean distance between populations i and j is

$$d_{ij} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}, \tag{2}$$

which implies $d_{ii} = 0$. Nevertheless, several years ago it seemed desirable to take $d_{ii} > 0$, for two reasons. First, Malécot [13] and Kimura and Weiss [14] had shown that the decline of kinship in the limit for large distance has a term d^{-c}, which is singular for $c \neq 0$, $d = 0$. Second, many available bodies of data pool the smallest demographic units into larger ones which have an appreciable radius. These two considerations suggested that d_{ii} should be taken as $\sqrt{A/n\pi}$, the mean radius of one of n populations in an array occupying an area of A km^2 [15]. However, subsequent investigation has shown these considerations to be invalid. In the absence of selective clines, $c \doteq 0$ for distances over which kinship is measurable [16, 17]. Use of the radius does not adequately compensate for pooling local populations, which in the extreme case for large distances results in measuring clines rather than local differentiation, and so is to be strictly avoided if interest centers on drift and migration [18]. Therefore we assume that the local population is small enough to be considered panmictic, and take $d_{ii} = 0$. If we wanted to estimate c, we could have taken $d_{ii} = 10^{-8}$, say, without affecting the other estimates.

The justification for Euclidean distance is not that migration is as the crow flies [19], but that non-Euclidean measures of distance are arbitrary, since the shortest road is not necessarily the one most traveled, and the mean travel time over many generations is hardly calculable. Euclidean distance has the two advantages of objectivity and ease of calculation, and an acceptable algorithm for computing non-Euclidean distance has not been proposed. Even when the number of populations is so small that the kinship matrix can be readily comprehended, and so no further condensation may seem necessary, comparison with other population structures is most easily made through the effect of Euclidean distance in equation (1).

The original method of bioassay from phenotype pairs has been shown to give a biased estimate of a [2, 20]. The current method consists in estimating gene frequencies for each commune by maximum likelihood, assuming Hardy-Weinberg equilibrium in each, and combining these estimates to give Y_{ij}, the estimated kinship between communes i and j, relative to random Swiss [5]. To convert this to a priori kinship [21], we must fit the equation

$$\phi(d) = (1 - L)\, ae^{-bd} + L \tag{3}$$

estimating a, b, and L simultaneously, where L is the kinship at the limit for large distances in Switzerland, relative to random Swiss. Thus L is a measure of deviation of contemporary from founder gene frequencies in Switzerland, although of course not including deviations of the founders from the rest of Europe or any other region greater than Switzerland.

In retrospect it was fortunate that the large number of communes led us to appreciate the economy of Malécot's theory for isolation by distance, whereby three parameters (a, b, L) suffice to describe population structure. However, we failed to recognize that, in other regions, the number of localities might be so small that interest would attach to each value of ϕ_{ij}, the kinship between populations i and j, an element of the matrix Φ which contains all the information about population structure. It was left to Bodmer and Cavalli-Sforza [22] to recall Malécot's earlier work which predicts ϕ_{ij} from migration [23]. We shall return to this in a later section.

Swiss communes range from Alpine isolates to large cities, and so it is desirable to divide them into more homogeneous classes. Morton and Hussels [4] regressed estimates of inbreeding from pedigrees on demographic parameters of the commune of origin, giving the discriminant $D = -.00055 + .00135\, X_1 + .00133\, X_2 + .00409\, X_3$, where $X_1 =$ altitude in kilometers, $X_2 =$ agricultural workers/total population in 1920, and $X_3 =$ population of origin resident in canton/total population of origin in 1920. This analysis allows us to divide communes into four classes:

Alpine isolates: $d > .0035, \quad N < 1{,}000;$

Alpine towns: $\begin{cases} .0025 < d < .0035, & N < 5{,}000 \\ d > .0035, & N > 1{,}000; \end{cases}$

lowland towns: $d < .0025, \quad N < 2{,}500;$

cities: $\begin{cases} d < .0025, & N > 2{,}500 \\ .0025 < d < .0035, & N > 5{,}000, \end{cases}$

where N is the resident population in 1920. Characteristics of these classes are shown in table 1, weighting each commune by N. The four demographic groups are clearly differentiated.

When the ABO samples are grouped in this way, and pair i,j is assigned to class k if either i or j or both belong to k, we find the expected differences in population structure (table 2). In this analysis we have pooled groups B and AB because the

TABLE 1

DEMOGRAPHIC CHARACTERISTICS BY CLASS OF COMMUNE

Characteristic	All Switzerland	Cities	Lowland Towns	Alpine Towns	Alpine Isolates	Saas
Altitude (km)	0.548	0.491	0.543	0.692	1.005	1.637
Agricultural workers / total population	0.118	0.039	0.185	0.195	0.382	0.295
Residents / km^2	97.6	404.6	92.4	38.4	13.0	5.9
1900 population / 1920 population	0.855	0.789	0.932	0.883	0.967	0.913
Proportion of population of origin resident in canton	0.748	0.750	0.704	0.828	0.908	0.933
Mean discriminant	.00218	.00221	.00167	.00307	.00413	.00415

significant deficiency of the latter argues for misclassification of AB as B [1]. The 37 distance classes used in the analysis have been grouped for simplicity of presentation.

The estimate of $\phi(0)$, for random individuals from the same commune, is greater than had been previously obtained from phenotype pairs (.00653 versus .00125). This must partly be due to the weights, which are the sample size N_i by the present method and N_i $(N_i - 1)/2$ from phenotype pairs; the latter give disproportionate effect to large samples. However, estimates for Alpine towns and isolates are so

TABLE 2

ESTIMATION OF KINSHIP FROM ABO FREQUENCIES ($\times 10^5$)

Distance Interval (km)	Class Value	Weight for All Swiss ($\times 10^{-4}$)	All Switzerland	Cities	Lowland Towns	Alpine Towns	Alpine Isolates
Same place	0	14	653	251	654	1,088	2,861
0–6	4	67	163	170	145	199	244
7–9	8	70	142	164	132	185	159
10–15	13	201	126	134	119	127	210
16–27	22	585	100	109	95	107	189
28–47	38	1,309	65	61	66	69	44
48–85	66	2,731	24	24	24	40	−7
86–225	141	6,597	−36	−49	−37	−28	−27
226–375	256	264	−28	−14	−23	−29	−56
Total weight $\times 10^{-4}$	...	...	11,836	2,931	10,214	3,674	1,395
L estimated simultaneously with $c = 0$:							
a			.0025	.0027	.0024	.0026	.0069
σ_a			.0002	.0001	.0001	.0002	.0019
b			.0185	.0177	.0175	.0106	.0643
σ_b			.0026	.0019	.0024	.0030	.0186
L			−.00056	−.00070	−.00058	−.00089	−.00027
σ_L			.00008	.00007	.00008	.00026	.00009

great, relative to adjacent communes, as to require further explanation. Sibs under universal military service are sampled contagiously, and to that extent sampling is not strictly random. Any clerical mistakes in tabulating phenotype frequencies (for example, double enumeration of the same individual) will tend to inflate $\phi(0)$, without systematic bias for $d > 0$. Finally, maximum-likelihood estimation of the A and B gene frequencies is biased upward by small samples consisting of only A and B phenotypes, and the mean sample size for an Alpine isolate as defined is about 25. Thus it is likely that the estimates of $\phi(0)$ are spuriously high, without any important consequence because the proportion of random pairs falling into this class is so small (.001).

Despite these possible errors, least-squares estimates of the Malécot parameters [5] are in reasonable agreement with phenotype pairs. As corrected by Morton [2], the latter estimates for $c = 0$ were $a = .0029 \pm .0001$, $b = .0231 \pm .0010$, and $L = -.0004$. There is little difference among cities and towns, but Alpine isolates have higher values of both a (.0069 by the present method, .0057 by corrected phenotype pairs) and b (.0643 here, .0967 by phenotype pairs), where Alpine isolates were somewhat differently defined in the two studies. These estimates are relative to large distances within Switzerland, and a is no greater if L is taken from the pooled data. Since the inbreeding coefficient under migration is less than kinship of random pairs within isolates, the conclusion of Hussels [3] from genealogical data is strengthened by bioassay: the inbreeding coefficient of Alpine isolates "relative to the whole of Switzerland could not much exceed .006, either now or in recent centuries." Values an order of magnitude higher have been found for oceanic islands and slash-and-burn agriculturists [24, 25].

ISONYMY

The pioneer study of G. Darwin [26] which related marital isonymy, or concordance of surnames, to consanguineous marriage was refined by Crow and Mange [27], who obtained an expression which may be combined with the Malécot equation as

$$I(d) = 4(1 - L)\, ae^{-bd} + L, \tag{4}$$

where $I(d)$ is the frequency of isonymous pairs at distance d, and L is the limiting frequency for large distance within the region. The principal error in this method is due to polyphyletic surnames, acquired contagiously by members of the same locality, such as Andenmatten in Saas and Jesus in northeastern Brazil [28]. The predilection of parish priests, in one case for place names and in the other for religious ones, may be an important factor in establishing a local polyphylon. Populations so characterized tend to give an unreliable estimate of kinship from isonymy, whereas in Britain, where names like Brown, Smith, and Johnson are less contagiously distributed, there is good agreement between estimates of kinship from isonymy and other evidence [16, 29].

We paired paternal and maternal surnames from the material on retinal degeneration in all possible ways with surnames from hemophilia and myotonic dystrophy,

TABLE 3

ISONYMY BY CLASS OF COMMUNE ($\times 10^5$)

Distance Interval (km)	All Switzerland	Cities	Lowland Towns	Alpine Towns	Alpine Isolates
Same place	2,890	1,306	4,982	7,500	0
0–6	737	599	790	644	1,408
7–9	446	383	512	141	0
10–15	313	287	303	336	313
16–27	167	145	182	121	143
28–47	108	113	118	61	69
48–85	73	78	79	51	23
86–225	26	32	27	21	8
226–375	21	13	25	6	0
Total weight $\times 10^{-3}$	1,681	758	1,289	547	150
Omitting same place:					
a	.0020	.0012	.0021	.0017	.0044
σ_a	.0002	.0003	.0003	.0006	.0020
b	.0756	.0499	.0750	.0830	.1262
σ_b	.0092	.0117	.0091	.0271	.0580
L	.0004	.0004	.0004	.0003	.0001
σ_L	.0001	.0001	.0001	.0001	.0001

assigning mothers the coordinates of their place of origin, but the discriminant and census size of their husband's commune of origin, so that each parent is demographically characterized by the child. As with ABO bioassay, a pair i,j is assigned to the kth class if either i or j or both belong to k (table 3). In estimating the Malécot parameters, pairs from the same place were omitted to reduce the bias due to polyphyletic surnames. The value of a is less, and b greater, than for blood groups, reflecting the greater mutability of surnames than genes. This mutability may not be random: for example, a common polyphylon may be more likely to differentiate (as Andenmatten, Anthamatten, and Indermatten in Saas, or Smith and Smythe in Britain), and an alien surname may undergo modification to conform to the local language. It is not surprising to find that genes and surnames have different systematic pressure, but the Malécot parameters are in rough agreement.

All paternal and maternal surnames for 1,260 marriages in Saas with both parental names recorded were paired in all possible ways, and the pairs classified by generation, assuming 25 years per generation (table 4). Thus a difference of 1–25 years in date of birth is defined as $t = 0$, omitting the class of zero difference which includes reflexive pairs. Fitting equation (4), we were unable to obtain convergence for a, b, and L simultaneously, because of the flatness of the likelihood surface for a range of L values. To obtain reliable estimates of a and b, we must know L. Hussels [3] regressed marital isonymy in Saas on pedigree inbreeding and obtained

TABLE 4

ISONYMY IN SAAS BETWEEN GENERATIONS

No. Generations Elapsed (t)	Difference in Mean Birth Years	Observed No. Pairs $\times 10^{-3}$	Isonymy
0	1–25	1,148	.0823
1	26–50	1,097	.0813
2	51–75	896	.0786
3	76–100	747	.0771
4	101–125	641	.0763
5	126–150	536	.0757
6	151–175	449	.0754
7	176–200	354	.0751
8	201–225	235	.0720
9	226–250	132	.0667
10	251–275	58	.0609
11	276–300	20	.0526
12	301–325	5	.0598

an intercept of .05775, corresponding to polyphyletic isonymy as kinship approaches zero. Taking $L = .05775$, we obtain $a = .0066 \pm .0002$ and $b = .0731 \pm .0107$.

The predicted evolution of kinship in time is

$$E\left[a - \frac{I(t) - L}{4(1 - L)} \right] = \left(\frac{1}{4 N_e m_e + 1} \right) \left(1 - e^{-(2m_e + 1/2N_e)t} \right), \quad (5)$$

where N_e is the evolutionary size of Saas and m_e is the systematic pressure. Fitting this equation by least squares to generation 8 and greater [30], we find the estimates of the evolutionary parameters are $N_e = 753 \pm 96$ and $m_e = .047 \pm .017$; and the kinship of random contemporary pairs, relative to indefinitely remote founders, is $\phi = 1/(4 N_e m_e + 1) = .00707 \pm .0089$. In evolution of isonymy the bias due to polyphyletic surnames is removed by the way we estimated L, so that this application of isonymy is not subject to the same limitations as isolation by distance.

CHAINS OF INBREEDING AND KINSHIP

Morton et al. [31] applied a theory which estimates the parameters of kinship or inbreeding in an evolving population. Assuming that the evolutionary size N_e and the systematic pressure m_e are constant, the cumulative kinship corresponding to a chain of length C from one member of a pair through a common ancestor of the other member of the pair to their hypothetical child is

$$\phi^{(t)} = \sum_{i=1}^{C} n_i 2^{-i}/N \doteq \left(\frac{1}{4 N_e m_e + 1} \right) \left(1 - e^{-(2m_e + 1/2N_e)t} \right), \quad (6)$$

where n_i is the number of chains of length i observed in a sample of N pairs and

$t = (C - 1)/2$ is the number of generations between the common ancestor and the mean of the pair.

Least-squares estimates of systematic pressure [30] have a mean of .2 and show no trend with increasing isolation, but evolutionary size is distinctly less for Alpine isolates (table 5). Assuming $m_e = .2$ for all groups, estimates of evolutionary size

TABLE 5

CHAINS OF INBREEDING BY CLASS OF COMMUNE IN CONTROL SAMPLE

Chain Length (C)	All Switzerland	Cities	Lowland Towns	Alpine Towns	Alpine Isolates
3	2	2	0	0	0
4	0	0	0	0	0
5	44	6	18	10	10
6	8	2	4	2	0
7	40	6	16	14	4
8	16	2	8	6	0
9	16	8	2	2	4
10	6	2	0	4	0
Total pairs	3,179	825	1,634	571	149
Estimates Using Cumulative Inbreeding for $C = 6$–8, Assuming $m_e = .2$					
N_e	1,394	1,428	1,971	1,163	388
ϕ	.00090	.00087	.00063	.00107	.00321

range from about 2,000 for lowland towns to about 400 for Alpine isolates, with asymptotic inbreeding increasing from .00063 to .00321. The mean inbreeding coefficient (.00090) is comparable to .00092 for Belgium [32] and .00076 for Denmark [33, p. 384] a generation ago. It agrees with an estimate of .00097 from data on mongolism in the Catholic cantons of eastern and central Switzerland [34], and exceeds the value ascertained from the incomplete pedigrees of this study, in which the mean inbreeding coefficient for degrees of relationship up to and including second cousins is .00065. Thus close inbreeding appears to account for at least two-thirds of the total inbreeding in Switzerland. The present estimates are about 10% higher than a previous report which included immigrants without a Swiss commune of origin, who in 1920 comprised 10.4% of the resident population [4].

In Saas the estimates of N_e and m_e for kinship show no marked change with time (table 6). Evolutionary size is about 300, and systematic pressure, about .1. In the nineteenth century the random kinship was about .009. Inbreeding then closely paralleled random kinship. However, before that time the evolutionary size was apparently much greater, reflecting avoidance of consanguineous marriage. There is other evidence that inbreeding increased in Europe during the nineteenth century [35–37].

TABLE 6

CHAINS OF INBREEDING AND KINSHIP IN SAAS

CHAIN LENGTH (C)	INBREEDING				KINSHIP			
	Before 1700	1700–1800	1800–1900	1900+	Before 1700	1700–1800	1800–1900	1900+
3	0	0	0	0	33	12	6	6
4	0	0	0	0	56	30	12	0
5	0	4	26	10	93	62	35	13
6	2	5	13	6	95	87	92	35
7	0	22	119	36	100	154	182	99
8	2	38	111	33	36	282	297	119
9	0	66	293	97	23	453	586	420
10	4	50	327	110	48	757	1,089	608
11	6	78	781	291	74	1,086	2,113	1,438
12	0	36	1,051	468	28	993	3,817	2,016
13	0	16	1,837	859	4	655	6,998	4,639
14	0	6	2,282	1,432	2	282	12,196	7,310
Total pairs	137	353	607	210	2,000	2,000	2,000	1,000
	Using Cumulative F for $C =$							
	6–7	7–9	9–11	11–13	3–5	7–9	12–14	12–14
	Assuming $m_e = .1$							
N_e	6,060	846	272	254	168	281	300	286
ϕ	.00041	.00294	.00912	.00973	.01466	.00883	.00828	.00868

MIGRATION

Distances between places of origin of mother and child are given in table 7. Only a minority of pairs come from the same place of origin. This corrects table 6 of Morton and Hussels [4], in which a peculiarity of exponentiation in our computer system assigned some pairs with different coordinates of origin to the same coordinate.*

The control sample is defined as the material on hemophilia and myotonic dystrophy, excluding pairs in which the child had no Swiss commune of origin. Since recessive inheritance is not involved, these pairs should be representative of the country at the time of this sample, in which the child's mean birth year was 1916.

The familial sample consists of index families for retinal degeneration, in which the child had a Swiss commune of origin, the parents were normal, there was no conclusive evidence of sex-linked or dominant inheritance, and there were at least two cases in the pedigree (not necessarily in the same sibship). The isolated sample agrees in all respects with the familial sample, except that there was only a single case in the pedigree.

* If $X \leqslant 0$, then the multiplication represented in FORTRAN by $X * X$ is correctly evaluated as X^2, but $X * X \neq X** 2 = 0$. Users of the DISTAN program in M mode should compute distance by multiplication, as the program does internally for I and S mode [15].

TABLE 7

MOTHER-CHILD DISTANCE BY TYPE OF RETINAL DEGENERATION

DISTANCE INTERVAL (km)	CLASS VALUE FOR MOTHER-CHILD CONTROL	MOTHER-CHILD		
		Control	Isolated	Familial
Same place	0	39	56	80
0–6	2	123	91	93
7–9	8	27	28	23
10–15	12	35	26	19
16–27	21	57	66	18
28–47	36	61	53	29
48–85	64	65	41	25
86–225	135	62	44	26
226–375	290	10	1	2
>375	400	16	8	3
Total	...	495	414	318
Mean distance, D	...	52.9	38.6	27.9

The mean mother-child distance increases from the familial to the control group, with a conspicuous reduction in the proportion of pairs at small distances. This confirms the suggestion [15] that parent-offspring distance is a sensitive test for recessive inheritance, which is not vitiated by incomplete ascertainment of inbreeding or affection.

Hussels et al. [9] were able to show that the familial sample is defined by at least seven recessive genes, without admixture of other mechanisms, and with a mean gene frequency of less than .00415. This result may be entered into equation (9) of [15], $r(d) = (Q + ae^{-bd})\mu(d)/(Q + \phi)$, where $\mu(d)$ is the distribution of parental distance, $r(d)$ is the distribution of distance in parents of affected, and

$$\phi = \sum_d \phi(d)\mu(d).$$

This equation is derived as follows. If $\phi(d)$ is the inbreeding coefficient at distance d, then the probability of affection is $A + B\phi(d)$, where A is the random load and B the inbred load. Then among parents of affected children, the posterior probability of d is $r(d) = [A + B\phi(d)]\mu(d)/(A + B\phi)$.

In genetic load theory, A/B estimates the mean gene frequency per contributing locus. If inbreeding and kinship are approximately the same for a given distance, we may write $\phi(d) \doteq ae^{-bd} + L$. Making these substitutions, we see that the first formula for $r(d)$ is satisfied by $Q = A/B + L$. Noting that the distance between places of origin of mother and child is the parental distance, we may evaluate $r(d)$ under the hypothesis that $\mu(d)$ follows a theoretical distribution with the same parameters as the control. Assuming an augmented gamma distribution [15], the control parameters are $h = .07879$, $z = .00526$, and $g = .68969$. At $Q = .00359$, $a = .0025$, $b = .0185$, and $c = 0$, the maximum-likelihood scores for Q give $U^2/K = 5.59$ for the familial sample and 4.15 for the isolated sample. Thus the value

of Q for the familial sample is significantly less than Q_0, which is based on an upper limit provided by genetic load theory. Iterating on Q, the maximum-likelihood estimate is .00137 ± .00042, which corresponds to a mean gene frequency per contributory locus of .00137 + .00056 = .00193. The total recessive gene frequency is .02908, and so the number of contributory loci is 15, or about twice as great as the lower limit [9]. Assuming a mean systematic pressure of .00145 per generation against a recessive gene for retinal degeneration [9], the mean mutation rate per locus per generation is $\mu = (.00145)\ (.00193) = 2.8 \times 10^{-6}$. Isolation by distance permits unbiased estimation of quantities for which genetic load theory gives only limits.

The value of Q for the isolated sample is significantly greater than Q_0 and the value from the familial sample. Iteration gives .01655 ± .01829 as the maximum-likelihood estimate. Let y be the proportion of nonrecessive cases in the isolated sample, estimated by $y = (.01655 + .00056 - .00193)/(.01655 + .00056) = .89$. Mean distances (table 7) yield the estimate $y = (38.6 - 27.9)/(52.9 - 27.9) = .43$. An alternative analysis uses the control distribution without assuming that it has any theoretical form [38]. Similar estimates are obtained: familial sample, $Q = .00081 \pm .00055$; isolated sample, $Q = .00347 \pm .00222$; mean gene frequency, .00137; number of loci, 21; mutation rate per locus, 2.0×10^{-6}; proportion of nonrecessive among isolated cases, .66. As first indicated by segregation and genetic load analyses, it appears that a large fraction of isolated cases is nonrecessive, and may well be nongenetic.

In principle, random pairs of parents of probands give information about the genetic basis of a rare trait [38]. In practice, the nonbinomial variation generated when n individuals are paired in all $n(n-1)/2$ nonreflexive ways, the loss of information due to nonallelic pairs if there is more than one locus causing the trait, and the possible distortion when incidence or ascertainment varies among localities make the distribution of random pairs less informative than that of parent-child pairs.

As a test of the method, all possible random pairs of father's place of origin were made within and between the diagnostic groups of familial myotonic dystrophy, hemophilia A, and hemophilia B. Since no locus is common to two diagnostic groups, we may use the pairs between groups as a control. There is a conspicuous excess of pairs at short distances within diagnostic groups, reflecting local identity by descent (table 8). Iterating Q, the maximum-likelihood estimate is $Q = .01088 \pm .00673$, a small fraction whose confidence interval includes the mean gene frequency per diagnostic class. The same procedure applied to retinal degeneration within diagnostic groups yields $Q = .00817 \pm .00289$, a similar result which is significantly less than the total gene frequency. Thus there is no doubt that more than one locus contributes to familial retinal degeneration, and that the genetic basis of the diagnostic categories is in part different. If we attempt to ask whether some loci contribute to two or more diagnostic categories (which is known from pedigree data), an appropriate test a priori is to take the second column in table 8 as control for the fourth column. Then $Q = .73241 \pm .03381$, as if there were no appreciable identity by descent between diagnostic categories. Other plausible con-

TABLE 8

RANDOM DISTANCE WITHIN AND BETWEEN DIAGNOSTIC GROUPS, BY FATHER'S PLACE OF ORIGIN

DISTANCE INTERVAL (km)	FAMILIAL MYOTONIC DYSTROPHY AND HEMOPHILIA		FAMILIAL RETINAL DEGENERATION	
	Within	Between	Within	Between
Same place	121	47	124	97
0–6	175	159	228	255
7–9	177	202	194	269
10–15	478	488	474	823
16–27	1,274	1,552	1,323	2,069
28–47	2,695	3,482	2,972	5,439
48–85	4,867	7,401	7,154	14,032
86–225	11,485	15,397	17,378	34,375
226–375	512	848	816	1,231
>375	0	0	0	0
Total	21,784	29,576	30,663	58,590
Mean distance, D	100.6	101.6	106.1	107.3

trols, such as isolated × familial retinal degeneration, gave similar results, which indicates that each contributory allele tends to produce a characteristic diagnostic type.

SIMULATION OF POPULATION STRUCTURE

With more than 3,000 communes, Switzerland cannot be simulated except by abstraction. We consider a circle on which 30 populations of equal size N are equally spaced and migration is symmetrically from the two adjacent populations, called stepping-stones. This model is a realistic representation of a linear continuum. The long-range migration rate m is separated from the short-range migration rate k by the point at which $(1 - L)\, ae^{-bd} + L = 0$ or $d = \ln[-a(1 - L)/L]/b = 81$ km.

Smaller distances are associated with greater kinship, and larger distances with less kinship, than for random Swiss. From table 7, this partition gives $m = 44/495 = .089$ and $k = 61.5/451 = .408$. For Alpine isolates there is not a sufficiently large random sample to estimate these parameters independently. The effective migration rate for Saas is less than for all Swiss, and so we halve the latter estimates to give $m = .045$, $k = .204$ for Alpine Swiss. Since Malécot's theory gives $b \doteq \sqrt{2m/k}/D$, where D is the distance assigned to stepping-stones, we take $D = \sqrt{2m/k}/b$ or 35.7 km for all Switzerland and 10.3 km for Alpine isolates. The effective size N must be guessed, since migration can make it much less than the evolutionary size N_e. Trial of the migration matrix with $1 - k$ on the diagonal and $k/2$ for each stepping-stone gave a ratio N_e/N of nearly three for all Swiss and nearly two for Alpine isolates, and so we took $N = 220$ for all Swiss and $N = 200$ for Alpine isolates.

As shown in table 9, these parameters provide a moderately good representation. The evolutionary size N_e is 594 for all Swiss and 390 for Alpine isolates, and the

TABLE 9

SIMULATION OF ISOLATION BY DISTANCE WITH A MIGRATION MATRIX

Parameter	All Switzerland	Alpine Isolates
Malécot intercept, a	.0025	.0069
Malécot exponent, b	.0185	.0643
Effective size, N	220	200
Long-range migration, m	.089	.045
Stepping-stone migration, k	.408	.204
Stepping-stone distance, D	35.7	10.3
At $t = \infty$:		
Kinship within commune, $\phi_0 = a$	.0026	.0071
Kinship with stepping-stone, ϕ_m	.0020	.0045
Malécot exponent, b	.0143	.0570
Inbreeding coefficient, α	.0021	.0060
At $t = 1, 5, 10, 20, 99$:		
Evolutionary size, N_e	594	390
Systematic pressure, m_e	.160	.092
At $t = 5$:		
Kinship within commune, $\phi_0 = a$	.0020	.0042
Kinship with stepping-stone, ϕ_m	.0014	.0020
Malécot exponent, b	.0182	.1010
Inbreeding coefficient, α	.0016	.0033

NOTE.—Thirty-point circle, road distance, equal population size, stepping-stone migration.

systematic pressure m_e is .160 and .093, respectively. The Malécot intercept a is .0026 for all Swiss and .0071 for Alpine isolates, and the value of b using perimeter distance is .0143 and .0570, respectively. In earlier generations a is of course less, and b is greater.

The estimate of the inbreeding coefficient for virilocal place of origin is $\alpha = [2k\phi_m + (1 - 2k)\,\phi_0]$, where ϕ_0 is the kinship within a commune and ϕ_m is the kinship with the stepping-stone. This gives .0060 for Alpine isolates, in reasonable agreement with table 6 for Saas and with the conclusion of Hussels [3]. However, inbreeding for all Swiss is estimated as .0021 asymptotically and .0016 in the fifth generation, or about twice as great as in table 5. Either the simulation is too approximate (for example, by assuming constant population size), or there is significant avoidance of consanguineous marriage in Switzerland. The latter hypothesis is likely, since the evolutionary size for kinship within a commune is about half as great as for inbreeding in table 5.

There are obviously other ways to simulate complex population structure by a reasonably small migration matrix, but our results with a circular model are encouraging. They suggest that close inbreeding ($t \leqslant 5$ generations) accounts for about 75% of total inbreeding in Switzerland, but only 55% in Alpine isolates.

SUMMARY

Phenotype bioassay of ABO blood groups in Switzerland agrees with the phenotype pair bioassay reported earlier. A priori kinship for two random Swiss from the same commune is estimated as $a = .0025$, with an exponential decline of $b = .0185$/km. Random Swiss have kinship $\phi_R = .00056$, relative to Swiss founders. Cities, lowland towns, and Alpine towns are consistent for these parameters. Alpine isolates have larger values of a (.0069) and b (.0643). Isonymy (surname concordance) gives an inflated estimate of local kinship, attributable to polyphyletic surnames. When pairs from the same place are omitted, estimates of a are less, and of b are greater than for the ABO blood groups, which we ascribe to the greater mutability of surnames than genes. Estimates of evolutionary size from inbreeding are of the order of 1,000 for cities, towns, and all Switzerland, and of hundreds for Alpine isolates. The mean inbreeding coefficient is .00090 for all Swiss, and several times this for Alpine isolates. The systematic pressure approaches .2 for all Swiss and about half of this for Alpine isolates, and is therefore dominated by migration rather than selection or mutation. Distances between places of origin of mother and child confirm (independently of segregation) that isolated retinal degeneration includes a large proportion of sporadic cases and that familial cases are due to more than one locus. Simulation of isolation by distance with a migration matrix suggests appreciable avoidance of consanguinity in all Switzerland, but not in Alpine isolates 50 years ago, and supports other evidence from genealogies, bioassay, and isonymy that the total coefficient of inbreeding in Alpine isolates relative to the founders of Switzerland is less than .006, or an order of magnitude less than some oceanic islands and slash-and-burn agriculturalists. Different methods of studying kinship are in good quantitative agreement.

REFERENCES

1. MORTON NE, YASUDA N, MIKI C, YEE S: Population structure of the ABO blood groups in Switzerland. *Amer J Hum Genet* 20:420–429, 1968
2. MORTON NE: Inference in population structure studies. *Amer J Hum Genet* 23:538–539, 1971
3. HUSSELS I: Genetic structure of Saas, a Swiss isolate. *Hum Biol* 41:469–479, 1969
4. MORTON NE, HUSSELS I: Demography of inbreeding in Switzerland. *Hum Biol* 42:65–78, 1970
5. MORTON NE, YEE S, HARRIS DE, LEW R: Bioassay of kinship. *Theor Pop Biol* 2:507–524, 1971
6. ROSIN S: Die Verteilung der ABO-Blutgruppen in der Schweiz. *Arch Klaus-Stift Vererbungsforsch* 31:1–127, 1956
7. AMMANN F, MARTY F: Une grande famille valaisanne avec plusieures branches atteintes de formes différentes de dégénérescence tapétorétinienne remontant au même couple ancestral. *J Génét Hum* 11:221–231, 1962
8. EIDGENÖSSISCHES STATISTISCHES BUREAU: *Eidgenössische Volkszählung* vom 1. Dezember 1920. Schweizerische statistische Mitteilungen, vol 6, Bern, Francke, 1925, pp 148–204
9. HUSSELS IE, MORTON NE, KLEIN D, AMMANN F: Retinal degeneration in Switzerland. In preparation
10. DODINVAL P, MORTON NE: Aspects génétiques des hémophilies A et B en Suisse. In preparation

11. TODOROV A: Analyse de la ségrégation dans la dystrophie myotonique. *J Génét Hum* 18:387–406, 1971
12. ZURBRIGGEN J: *Familienstatistik Saastal.* Visp, Neue Buchdruckerei, 1961
13. MALÉCOT G: *Les mathématiques de l'hérédité.* Paris, Masson, 1948
14. KIMURA M, WEISS GH: The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49:561–576, 1964
15. MORTON NE, MIKI C, YEE S: Bioassay of population structure under isolation by distance. *Amer J Hum Genet* 20:411–419, 1968
16. IMAIZUMI Y, MORTON NE, HARRIS DE: Isolation by distance in artificial populations. *Genetics* 66:569–582, 1970
17. MALÉCOT G: Génétique des populations naturelles dans le cas d'un seul locus. *Ann Génét Sélection Anim* 4:385–409, 1972
18. IMAIZUMI Y, MORTON NE: Isolation by distance in New Guinea and Micronesia. *Archeology Phys Anthrop Oceania* 5:218–235, 1970
19. CAVALLI-SFORZA LL, BODMER W: Discussion on population structure, in *Computer Applications in Genetics*, edited by MORTON NE, Honolulu, Univ. Hawaii Press, 1969, 68–70
20. HARPENDING H: Inference in population structure studies. *Amer J Hum Genet* 23:536–538, 1971
21. MALÉCOT G: Consanguinité panmictique et consanguinité systématique. *Ann Genet Sélection Anim* 1:237–242, 1969
22. BODMER WF, CAVALLI-SFORZA LL: A migration matrix model for the study of random genetic drift. *Genetics* 59:565–592, 1968
23. MALÉCOT G: Quelques schémas probabilistes sur la variabilité des populations naturelles. *Ann Univ Lyon Sci Sec A* 13:37–60, 1950
24. MORTON NE: Population structure of Micronesia. In preparation
25. FRIEDLAENDER JS: Isolation by distance in Bougainville. *Proc Nat Acad Sci USA* 68:704–707, 1971
26. DARWIN G: Marriage between first cousins in England and their effects. *J Statist Soc* 38:153–184, 1875
27. CROW JF, MANGE AP: Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen Quart* 12:199–203, 1965
28. AZEVEDO E, MORTON NE, MIKI C, YEE S: Distance and kinship in northeastern Brazil. *Amer J Hum Genet* 21:1–22, 1969
29. YASUDA N, MORTON NE: Studies on human population structure, in *Proceedings 3d International Congress of Human Genetics*, edited by CROW JF, NEEL JV, Baltimore, Johns Hopkins Press, 1967, pp 249–265
30. MORTON NE, ROISENBERG I, LEW R, YEE S: Pingelap and Mokil Atolls: genealogy. *Amer J Hum Genet* 23:350–360, 1971
31. MORTON NE, HARRIS DE, YEE S, LEW R: Pingelap and Mokil Atolls: migration. *Amer J Hum Genet* 23:339–349, 1971
32. TWIESSELMANN FR, MOUREAU P, FRANÇOIS J: Evolution du taux de consanguinité en Belgique de 1918 à 1959. *Population* 17:241–266, 1962
33. BARTELS ED: *Heredity in Grave's Disease.* Copenhagen, Munksgaard, 1941
34. HANHART E: 800 Fälle von Mongolismus in konstitutioneller Betrachtung. *Arch Klaus-Stift Vererbungsforsch* 35:1–312, 1960
35. MORONI A: Evoluzione della frequenza dei matrimoni consanguinei in Italia negli ultima cinquant' anni. *Atti Ass Genet Ital* (Pavia) 9:207–223, 1964
36. SUTTER J, TABAH L: L'évolution des isolats de deux départements français: Loire-et-Cher, Finistère. *Population* 10:645–674, 1955
37. DODINVAL P, KLEIN D: Caractères démographiques et génétiques d'un petit village savoyard. *J Génét Hum* 11:1–15, 1962
38. MORTON NE: Population structure, in *Computer Applications in Genetics,* edited by MORTON NE, Honolulu, Univ. Hawaii Press, 1969, pp 61–68