ANNALS OF HUMAN GENETICS VOL. 53 PART 4 OCTOBER 1989

CONTENTS

(*All rights reserved*)

© University College London 1989

PRINTED IN GREAT BRITAIN BY THE UNIVERSITY PRESS, CAMBRIDGE

Ann. Hum. Genet. (1989). **53**. 339–355
Printed in Great Britain

Inbreeding in the Utah Mormons: an evaluation of estimates based on pedigrees, isonymy, and migration matrices

L. B. JORDE
Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah 84132, USA

SUMMARY

Using a computerized genealogical database, inbreeding coefficients were calculated for a sample of 435777 Utah Mormons. The population was divided into ten ten-year birth cohorts (1846–1945) and 22 geographic subdivisions in order to assess temporal and spatial variation in inbreeding. The average inbreeding coefficient for this population is 0·000106. The average within-groups random kinship coefficient is 0·000312, reflecting consanguinity avoidance. Random kinship matrices were formed by estimating the average kinship within each spatial subdivision and between all pairs of subdivisions. These matrices were compared statistically with kinship matrices previously estimated using migration matrices and isonymy data. The isonymy approach consistently overestimates random and total inbreeding as well as Wright's F_{st}. This can be attributed primarily to the assumption of monophyletic origin of surnames. The migration matrix method underestimates random inbreeding and F_{st}. This is due mainly to the assumption that outside immigrants are derived from a genetically homogeneous population. While the absolute values of the kinship coefficients estimated by each method differ substantially, the patterns of between-groups kinship coefficients given by each method are highly congruent. Logistic and linear regression analyses of 85235 marriages demonstrate that consanguinity is significantly dependent upon year of marriage, geographic distance between husband's and wife's birthplaces, and the population size of husband's and wife's birthplaces.

INTRODUCTION

Several approaches have been devised to estimate inbreeding in natural populations. The most direct method involves using pedigrees, but adequate pedigree data are often difficult to obtain. Indirect estimates of inbreeding can be made from gene frequencies, isonymy data, and migration matrices. Since each of these methods involves certain assumptions, much can be learned about the weaknesses of each method by comparing their results in a single population. Earlier studies of the Utah Mormon population have presented results based on gene frequencies (McLellan *et al.* 1984), migration matrices (Jorde, 1982, 1984), and isonymy (Jorde & Morgan, 1987). This population is particularly useful for comparisons of this type because of the availability of a computerized genealogical database that includes 1·2 million individuals. With this large sample size, it is possible to subdivide the population in several informative ways (e.g. spatially and temporally) without losing statistical power. Also, several types of ancillary data are available, permitting an examination of causal factors related to inbreeding.

The objectives of the present study are twofold. First, spatial and temporal variation in inbreeding is investigated using the genealogies of 435777 Utah-born subjects, and causal

determinants of consanguineous marriage are explored. Second, kinship measured genealogically is compared with kinship measured by other methods in order to evaluate the assumptions of each method. While other studies have reported more limited comparisons of kinship data, no previous study has compared kinship based on genealogies, isonymy, migration matrices, and gene frequencies in the same population.

METHODS

The Utah Population Database includes information on the date and place of birth, marriage, and death for most individuals in the database. Nearly all members of the current database were members of the Church of Jesus Christ of Latter-day Saints (LDS or 'Mormons'). The analyses reported here are based on 435777 members of the database who were born in Utah from 1847 to 1945. Temporal variation in inbreeding was assessed by dividing the sample into ten-year birth cohorts. As in previous studies of this population, spatial variation was evaluated by dividing the sample into 22 spatial subdivisions. These are known as 'stakes,' a key organizational unit in the LDS Church (see Jorde (1982) for further details).

The Mormon colonization of Utah began in 1847. Population growth, spurred by immigration and a high birth rate, was rapid: nearly 100000 Mormons inhabited Utah in 1870, and the total population of Utah in 1890 was over 200000 (of whom about 70% were Mormons) (Wahlquist, 1978). Currently, 95% of Utah's population is Caucasian, and 70% are members of the LDS Church (Martin *et al.* 1986).

Inbreeding coefficients (F) were calculated in the standard fashion (Wright, 1922). These coefficients were then averaged over the individuals in each birth cohort and in each spatial subdivision. Random inbreeding (F_r) in each subdivision was calculated by averaging the kinship coefficients of all possible pairs of individuals in that subdivision. While some studies of random inbreeding exclude close relatives from this calculation (e.g. Brennan & Relethford, 1983), this was not done in this study in order to make the results comparable with those of the previous isonymy and migration matrix studies (both of which assume random mating *including* all relatives). For most individuals in the computerized database, ascending genealogies do not predate 1800. Thus, as with any genealogical estimate of inbreeding, the data are truncated. However, relative to the founding Utah population, the estimates of inbreeding should be relatively complete and accurate.

Since the sample sizes of some subdivisions became quite large in later time periods (over 20000), it was computationally not feasible to estimate random inbreeding on the complete sample (20000 individuals would require the calculation of nearly 200000000 kinship coefficients). Thus, in those populations exceeding 5000 individuals, a sample of 5000 was drawn randomly. In three subpopulations exceeding 5000 in size, the F and F_r estimates from the random sample were compared with those of the complete sample. In these six comparisons, five of the estimates were identical to five decimal places. In the one remaining comparison, the two values differed by 3×10^{-5}. Given this degree of agreement, the random sampling procedure appears to be quite reliable.

Random kinship coefficients were also estimated for each pair of subdivisions in each time period. Each between-groups kinship coefficient thus represents an average of $m \times n$ kinship coefficients, where m is the number of individuals in one subdivision and n is the number of

individuals in the other. Again, the random sampling process described above was used for subdivisions exceeding 5000 individuals. These values formed the off-diagonals of a random kinship matrix (denoted $\mathbf{\Phi}$), and the random within-groups kinship values formed the diagonals of this matrix.

$\mathbf{\Phi}$ is an *a priori* kinship matrix: it specifies kinship relative to an ancestral founder population. It is sometimes more appropriate to evaluate kinship relative to the contemporary population ('conditional' kinship). A conditional kinship matrix, **R**, can be obtained from $\mathbf{\Phi}$ using a transformation suggested by Harpending & Jenkins (1974):

$$r_{ij} = \frac{\phi_{ij} + \phi_{..} - \phi_{i.} - \phi_{.j}}{1 - \phi_{..}}, \tag{1}$$

where $\phi_{i.} = \Sigma_k w_k \phi_{ik}$ (w_k is the proportion of the total population which lives in subdivision k); $\phi_{.j} = \phi_{i.}$, due to symmetry in $\mathbf{\Phi}$; $\phi_{..} = \Sigma_{i,k} w_i w_k \phi_{ik}$.

The overall level of genetic differentiation among subdivisions, F_{st} (Wright, 1943), can be obtained from **R** as:

$$F_{st} = \Sigma_i w_i r_{ii}. \tag{2}$$

This quantity has also been labelled R_{st} (Harpending & Jenkins, 1974) and r_0 (Rogers & Harpending, 1986). The quantity estimated in (2) is a 'reduced' estimate of F_{st} (Felsentstein, 1982), which has been shown to converge to equilibrium more quickly than other estimates (Rogers & Harpending, 1986; Wood, 1986). A simpler transformation is sometimes used to obtain **R**. This involves replacing the numerator of (1) with the quantity $\phi_{ij} - \phi_{..}$ (Harpending & Jenkins, 1974; Relethford, 1988). However, as Harpending and Jenkins point out, the latter transformation is valid only if the subdivisions are of equal size and arrayed in a regular geometric arrangement such that $\phi_{i.} = \phi_{.j} = \phi_{..}$. This is seldom the case in human populations, so the transformation given in (1) is preferable. This same transformation was applied to the $\mathbf{\Phi}$ matrices derived earlier from migration matrices (Jorde, 1982) and isonymy (Jorde & Morgan, 1987) in order to obtain comparative F_{st} measures.

The relationship between kinship and between-subdivision geographic distance was assessed by applying a centroid transformation to $\mathbf{\Phi}$ (as in equation (1)) and then plotting the first two eigenvectors of this matrix against one another (Lalouel, 1973). This plot, which provides a 2-dimensional representation of the genetic distances among subdivisions, was then rotated to maximum congruence with the actual geographic positions of each subdivision using least-squares estimation. The product-moment correlation (R_c) between the first two eigenvectors of $\mathbf{\Phi}$ and those of the transformed geographic distance matrix provides a measure of the fit between kinship and geographic distance. The relationship between $\mathbf{\Phi}$ and geographic distance often conforms to a negative exponential distribution (Malécot, 1948). Thus, a natural logarithmic transformation was applied to $\mathbf{\Phi}$ prior to centroid transformation and eigenvector extraction.

Since the elements of $\mathbf{\Phi}$ are not statistically independent, conventional statistical tests of significance between this matrix and geographic distance are not appropriate. Therefore, statistical significance was estimated empirically (Mantel, 1967; Smouse *et al.* 1986; Dow & Cheverud, 1985). A test statistic, S, is given by $\Sigma_{i<j} \Sigma X_{ij} Y_{ij}$, where **X** and **Y** are the two matrices being compared. This statistic was estimated first for $\mathbf{\Phi}$ and the geographic distance matrix. A

product-moment correlation coefficient, r, was also estimated for **X** *vs.* **Y**. Then a distribution of 1000 additional test statistics (denoted here as S^*) was generated, randomly permuting the rows and columns of one of the matrices in each comparison. The empirical significance level was obtained by comparing the distribution of S^* with S. The expectation and variance of S were derived by Mantel (1967); a Z score was obtained in the standard manner using these quantities and S. Since the distribution of Z is very similar to a t distribution for large n, the *approximate* standard error of r was obtained using the relationship S.E. $(r) = r/Z$ (Sokal & Rohlf, 1981).

The Mantel technique was also used to compare the random kinship matrices based on genealogies, isonymy, and migration matrices in each time period. The Mantel procedure uses only the upper triangles of each matrix, excluding the diagonal elements. Since the diagonal elements of kinship matrices contain important information, each kinship matrix was first transformed to a distance matrix using the formula:

$$d_{ij} = \phi_{ii} + \phi_{jj} - 2\phi_{ij}.$$

In comparing random kinship estimates based on different types of data, it is particularly important to ascertain that the cohorts being considered are indeed comparable (Rogers, 1987). In this study, the same birth cohorts were used for each random kinship estimate. Thus, the 'offspring' used in forming parent-offspring migration matrices are the same individuals used in estimating random kinship from isonymy and from genealogical data.

In a previous analysis of isonymy data from Utah, logistic regression analysis was used to 'predict' whether a marriage was isonymous or not (Jorde & Morgan, 1987). The independent variables used in that analysis were year of marriage, endogamous *vs.* exogamous marriage (using the stake of birth as the unit of subdivision), geographic distance between husband's and wife's city of birth, and the population sizes of the husband's and wife's stakes of birth. In the present study, a similar analysis was carried out. The dependent variable was consanguineous *vs.* non-consanguineous marriage, and the independent variables were the same as in the previous analysis. Also, a stepwise multiple linear regression was performed, using the actual kinship coefficient between husband and wife as the dependent variable. This analysis was performed on a subset of 85235 marriages of Utah-born couples for whom all of the above data were available.

RESULTS

The sample sizes used for each time period are given in Table 1. The first column gives the total number of individuals available for sampling in each time period, and the second column gives the actual number used after randomly sampling subdivisions with sizes greater than 5000. In total, 76% of the available subjects were actually used in calculating kinship coefficients.

The average inbreeding coefficient, F, for the entire sample (435777 of 572971 individuals) is 0·000106. The average within-groups random inbreeding coefficient, F_r, is also quite low (0·000312), reflecting the relatively large population sizes of most subdivisions. Avoidance of consanguinity is indicated by the fact that F_r is approximately three times greater than F. The average within-groups F_r value obtained from isonymy data is several times higher than that

Table 1. *Sample sizes for each birth cohort*

(The total sample size available in the database is given in column 1, and the actual sample used for calculating kinship coefficients is given in column 2.)

Birth cohort	Total	Sampled
1846–1855	6005	6005
1856–1865	25516	25516
1866–1875	41580	37331
1876–1885	57481	47484
1886–1895	68434	52749
1896–1905	74021	55936
1906–1915	80233	58532
1916–1925	81963	58602
1926–1935	72578	50484
1936–1945	65160	43138
Total	572971	435777

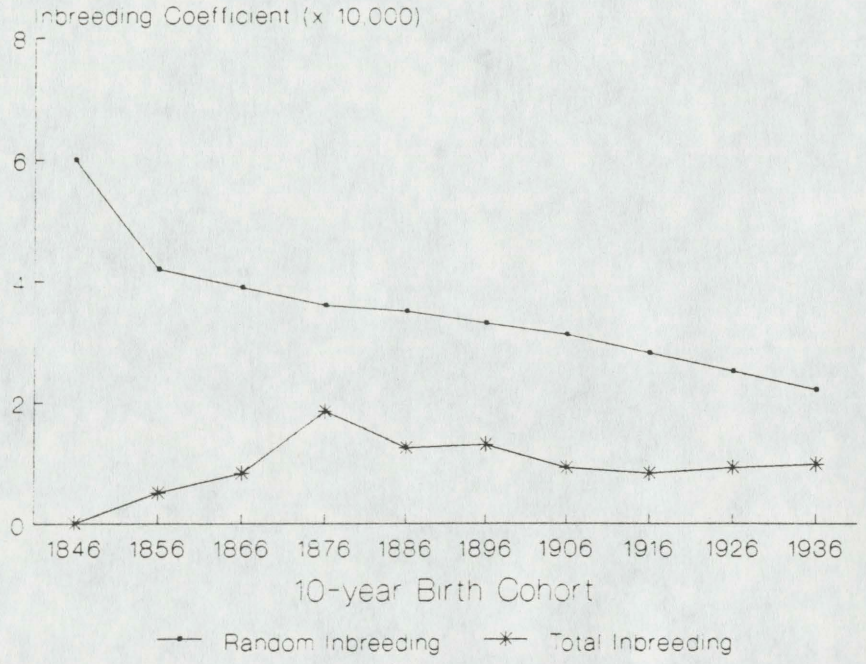

Fig. 1. Average random and total inbreeding coefficients in each 10-year birth cohort.

obtained from genealogies: 0·001186. The corresponding value for the migration data is an order of magnitude lower than the pedigree-obtained value: 0·000038.

Temporal variation in random and nonrandom inbreeding is illustrated in Fig. 1. Total inbreeding has a value of nearly zero in the first birth cohort, reflecting in part a lack of pedigree depth. The value increases to approximately 2×10^{-4} in the 1876–85 cohort, after which it gradually declines through most of the time period. Random inbreeding is highest in the earliest cohort, due to relatively low population sizes and the consequent greater weight of siblings and other close relatives in the computation of average random inbreeding. As population sizes increase through time, random inbreeding decreases. It should be pointed out that the F and F_r values shown in this graph are not strictly comparable. This is because the F values are based

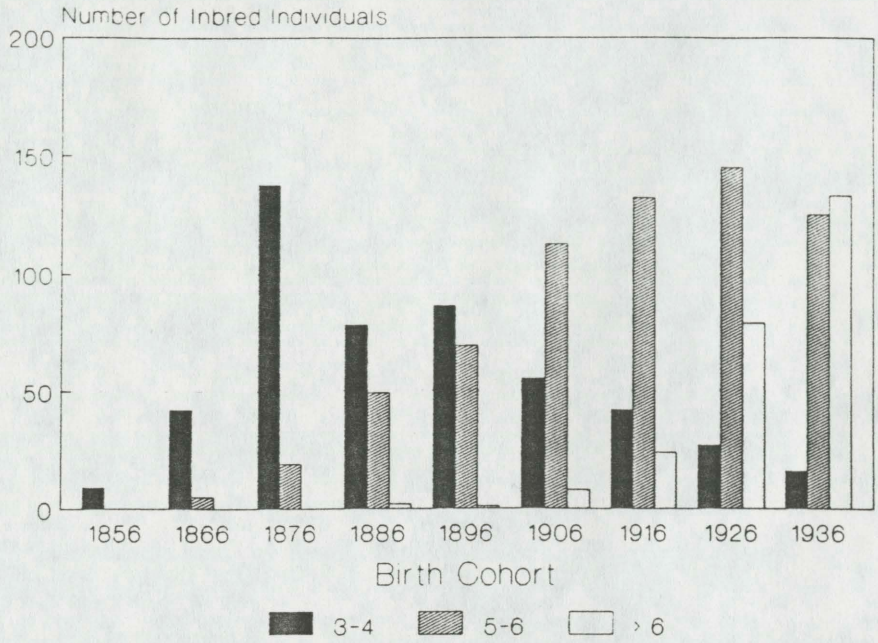

Fig. 2. Frequency of inbred individuals in each 10-year birth cohort, subdivided into three levels of inbreeding: $0{\cdot}5^3$ and $0{\cdot}5^4$, $0{\cdot}5^5$ and $0{\cdot}5^6$, and $< 0{\cdot}5^6$.

on kinship coefficients between the parents of the individuals in each birth cohort, while the F_r values are based on kinship coefficients calculated between the individuals themselves in each birth cohort.

Figure 2 shows the temporal distribution of inbreeding after dividing inbred subjects into three levels of inbreeding coefficients: $0{\cdot}5^3$ and $0{\cdot}5^4$, $0{\cdot}5^5$ and $0{\cdot}5^6$, and less than $0{\cdot}5^6$. While 'close' inbreeding tends to decrease through time (after reaching a maximum in the 1876 birth cohort), the more remote inbreeding classes continue to increase through time. As expected for a genealogy growing in complexity, the most remote category of inbreeding increases substantially only after 1916, but then grows rapidly.

In order to assess urban–rural differences in inbreeding patterns, average inbreeding rates were calculated separately for the four stakes containing urban centres (Salt Lake stake, Utah stake, Cache stake, and Weber stake) and for the remaining 18 subdivisions. Random and total inbreeding for these two groups in each time period are presented in Table 2. As expected, both categories of inbreeding are higher for the rural subdivisions than for the urban ones (with the exception of total inbreeding in the 1876–85 cohort). The urban–rural difference is greater for random inbreeding than for total inbreeding, however. The relatively high degree of random inbreeding reflects the small average sample sizes for the rural stakes. For all time periods combined, the F and F_r values for the urban stakes are 0·00009 and 0·00011, while those for the rural stakes are 0·00013 and 0·00055. Thus, the relative degree of departure from random mating is substantially greater in the rural subdivisions. As anticipated, the product-moment correlation between stake population size (averaged over all time periods) and F_r is negative and significant ($r = -0{\cdot}539$, $P < 0{\cdot}01$). The correlation between F and population size, while in the expected direction, is not significant ($r = -0{\cdot}235$, $P > 0{\cdot}29$). This result is consistent with the finding that the urban and rural subdivisions differ more with regard to F_r than F.

Table 2. *Average stake population sizes, total inbreeding (F), and random inbreeding (F_r) in urban and rural stakes*

(Inbreeding values are multiplied by 10^4.)

	Urban			Rural		
Birth cohort	Avg. size	F	F_r	Avg. size	F	F_r
1846–1855	1098·8	0·00	2·75	89·4	0·00	14·75
1856–1865	3870·0	0·48	1·94	557·6	0·56	7·56
1866–1875	5545·5	0·67	1·73	1077·7	1·01	6·30
1876–1885	7405·8	2·10	1·47	1547·7	1·55	5·82
1886–1895	8809·0	1·03	1·28	1844·3	1·48	5·76
1896–1905	9366·0	1·19	1·16	2030·9	1·40	5·48
1906–1915	10421·0	0·70	1·01	2141·6	1·14	5·26
1916–1925	10840·3	0·62	0·88	2144·6	1·05	4·95
1926–1935	10523·5	0·70	0·76	1693·6	1·21	4·85
1936–1945	10505·5	0·74	0·66	1285·4	1·41	4·98

Table 3. *F_{st} values estimated from migration matrices, isonymy, and pedigrees*

(Values are multiplied by 10^4.)

Birth cohort	Migration matrix	Isonymy	Pedigree
1876–1885	0·029	8·781	3·384
1886–1895	0·155	8·308	3·182
1896–1905	0·340	7·726	3·137
1906–1915	0·468	7·047	2·827
1916–1925	0·428	6·428	2·514
1926–1935	0·350	5·925	2·197
1936–1945	0·274	6·162	2·316
Total	0·304	7·383	2·837

F_{st} values obtained from pedigree data, isonymy, and migration matrices are presented in Table 3. For comparability with the migration matrix results published previously, only the birth cohorts beginning with 1876 are included here. Examination of the values for the total time period shows that isonymy gives the largest F_{st} value (0·00074), pedigree data give the next largest value (0·00028), and migration matrices give a much smaller value (0·00003). The pedigree and isonymy estimates display very similar temporal patterns; both gradually decrease through time, with a very slight increase in the final time period. The migration matrix F_{st} values, on the other hand, are lowest in the first time period, reach a maximum in the 1906–15 birth cohort, and decline thereafter. To a large extent, the low values in the first three time periods reflect the extremely high immigration rates during the colonization phase (most parents having been born outside Utah). Since the migration matrix model assumes that outside immigrants come from a genetically homogeneous population, a high proportion of such immigrants results in a substantially lowered F_{st} estimate.

The relationship between random kinship and geographic distance is depicted graphically in Fig. 3. In general, the genetic relationships between subdivisions are fairly concordant with their geographic distances, indicating an important isolation by distance effect in this population. The R_c value for kinship *vs.* geographic distance is 0·775. Without a logarithmic transformation of Φ, the R_c value is substantially smaller (0·453). Figure 3 bears a remarkable

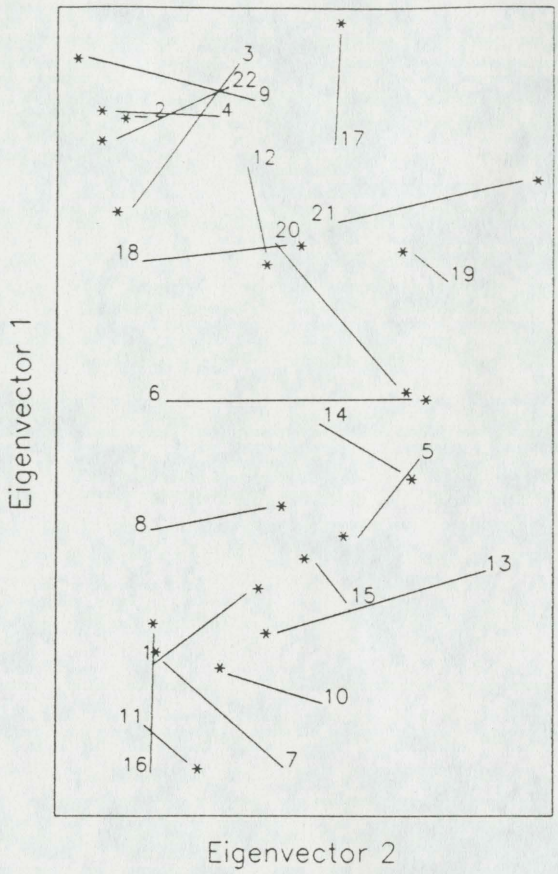

Fig. 3. A 'genetic map' in which the first two eigenvectors of ln Φ are plotted against one another and rotated to maximum congruence with the geographic locations of the populations. Each number indicates the geographic location of a population. The line next to each number connects it to the * which indicates the coordinates of the population as defined by its loadings on the first two eigenvectors of Φ. Thus, each line is a graphical measure of the goodness of fit between Φ and geographic distance.

resemblance to the eigenvector plot published earlier for kinship based on migration matrices (Jorde, 1982). In fact, the R_c value for the first two eigenvectors of Φ based on pedigree data versus Φ based on the migration matrix is very high indeed: 0·979. Again, this figure is based on logarithmically transformed kinship matrices. Without the transformation, the R_c value falls to 0·740.

Table 4 gives correlation values for each type of kinship compared with geographic distance. The significance levels were obtained using the Mantel technique. For the total sample, the Φ matrix based on migration data yields the highest correlation with geographic distance, while the Φ matrix based on isonymy data yields the lowest correlation. As with the F_{st} estimates, the temporal patterns based on isonymy and pedigree data are very similar, while the pattern based on migration matrices is somewhat divergent.

A direct comparison of each type of kinship matrix is presented in Table 5. Here, the correlations and significance values are given for pedigree versus migration data, pedigree versus isonymy data, and isonymy versus migration data. As expected from the results

Table 4. *Correlation coefficients, with standard errors, for kinship matrices (transformed into distance matrices) versus geographic distance*

Birth cohort	Migration matrix	Isonymy	Pedigree
1876–1885	0·283 ± 0·114*	0·253 ± 0·117*	0·265 ± 0·118*
1886–1895	0·132 ± 0·116	0·281 ± 0·117*	0·302 ± 0·118**
1896–1905	0·186 ± 0·116	0·287 ± 0·116**	0·304 ± 0·118**
1906–1915	0·195 ± 0·117*	0·259 ± 0·117*	0·298 ± 0·118**
1916–1925	0·227 ± 0·117*	0·263 ± 0·117**	0·290 ± 0·118*
1926–1935	0·298 ± 0·116**	0·264 ± 0·117**	0·287 ± 0·118*
1936–1945	0·347 ± 0·116**	0·063 ± 0·118	0·141 ± 0·118
Total	0·390 ± 0·117***	0·245 ± 0·117*	0·282 ± 0·118*

*$P < 0·05$; **$P < 0·01$; ***$P < 0·001$.

Table 5. *Correlation coefficients, with standard errors, for pairs of kinship matrices*

Birth cohort	Migration/ Isonymy	Migration/ Pedigree	Isonymy/ Pedigree
1876–1885	0·746 ± 0·205**	0·757 ± 0·207**	0·976 ± 0·216**
1886–1895	0·313 ± 0·209	0·272 ± 0·212	0·939 ± 0·214***
1896–1905	0·425 ± 0·209*	0·306 ± 0·213	0·927 ± 0·214***
1906–1915	0·654 ± 0·212***	0·585 ± 0·214**	0·945 ± 0·215***
1916–1925	0·609 ± 0·213**	0·623 ± 0·215**	0·960 ± 0·216***
1926–1935	0·624 ± 0·212**	0·653 ± 0·214**	0·958 ± 0·216***
1936–1945	0·204 ± 0·212	0·327 ± 0·213	0·963 ± 0·218**
Total	0·807 ± 0·213***	0·873 ± 0·215***	0·959 ± 0·216***

*$P < 0·05$; **$P < 0·01$; ***$P < 0·001$.

presented thus far, the greatest level of concordance is seen between the isonymy and pedigree data: all correlations exceed 0·920, and nearly all of the significance values reach the 0·001 level. The migration- and pedigree-derived Φ matrices exhibit somewhat lower correlations, but the correlation for all time periods is still quite high ($r = 0·873$, $P < 0·001$). The lowest overall correlation is obtained for isonymy versus migration, but it is again relatively high and significant ($r = 0·807$, $P < 0·001$).

For the total time period, a multivariate comparison of these matrices was carried out by regressing the pedigree-derived kinship matrix on three independent variables: geographic distance, kinship based on migration, and kinship based on isonymy. This regression yielded an R^2 value of 0·947 ($P < 0·001$). An evaluation of all possible subsets of the independent variables gave R^2 values of 0·765, 0·923, and 0·947 for the combinations of geographic distance/migration, geographic distance/isonymy, and migration/isonymy, respectively. The R^2 values for the independent variables singly were 0·080, 0·762, and 0·920 (geographic distance, migration, and isonymy, respectively). These statistics show that, once isonymy was entered as a predictive variable, migration yielded little additional predictive power, and geographic distance yielded almost none at all.

Another way to compare isonymy and pedigree data involves a direct evaluation of the degree of association between consanguineous and isonymous marriages. From this data set, 132093 marriages between Utah-born couples were evaluated in terms of isonymy and

Table 6. *Cross-tabulation of isonymous vs. consanguineous marriages*

	Consanguineous	Non-consanguineous
Isonymous	47	523
Non-isonymous	593	130930

$\chi^2 = 704{\cdot}3$; $P < 0{\cdot}001$.

consanguinity. The average F value for the 570 isonymous marriages in this sample was 0·00370, while the average F value among the 131523 non-isonymous marriages was 0·00009. This difference was highly significant ($P < 0{\cdot}001$) using both a t test and a non-parametric median test. A crosstabulation of isonymy versus consanguinity for these marriages is given in Table 6. These values reveal a highly significant association between consanguinity and isonymy ($\chi^2 = 704{\cdot}3$, $P < 0{\cdot}001$). Among the isonymous marriages 8·3% are consanguineous, while only 0·45% of the non-isonymous marriages are consanguineous. These figures, while substantiating an association between isonymy and consanguinity, are not closely in accord with the theory of isonymy. The isonymy method assumes that $F = P/4$, where P is the proportion of isonymous marriages in the population (Crow & Mange, 1965). Since the average kinship coefficient in this sample of marriages is 0·000104, the proportion of isonymous marriages should be 0·000416, or 55 marriages. The actual number of same-name marriages, 570, is ten times greater than the isonymy method predicts.

When stepwise multiple linear regression was applied to these data, three independent variables entered the equation predicting consanguinity levels in marriages. The first variable to enter the equation was year of marriage, which, as expected from Fig. 1, was negatively associated with consanguinity. The second variable to enter the equation was the population size of the wife's birthplace. As anticipated from population genetic theory and the results of the urban-rural comparison, the relationship between population size and consanguinity was negative. Finally, geographic distance between husband's and wife's birthplace entered the equation, again showing a negative association with consanguinity. While all of these variables had a highly significant relationship with consanguinity ($P < 0{\cdot}001$), the multiple R value was only 0·027, indicating that they do not explain much of the variance in consanguinity.

In the logistic regression analysis, 'stake endogamy' was the first variable to enter the equation. The odds ratio here is 0·45 (95% confidence limits = 0·37, 0·54), indicating that exogamous couples are about half as likely to be consanguineous as are endogamous couples. The population sizes of wife's and husband's birthplaces were the next two variables to enter the equation. The odds ratios for these two variables were 1·36 (95% confidence limits = 1·20, 1·54) and 1·26 (95% confidence limits = 1·12, 1·43), respectively. As expected, these odds ratios show that couples born in larger stakes are less likely to be consanguineous. The last variable to enter the equation was year of marriage (odds ratio = 0·73, 95% confidence limits = 0·68, 0·79). This odds ratio indicates that couples married more recently have a higher probability of being consanguineous, which seems contradictory to the results of the multiple linear regression. However, the logistic regression specifies simply the probability that couples will be related to one another at any level, while the multiple linear regression predicts the actual kinship coefficient. As Fig. 2 shows, the actual number of inbred individuals (and therefore consanguineous marriages) does increase through time, but this increase is due to more remote levels of consanguinity. Thus, while the probability of consanguinity does increase through

time, the average kinship coefficient decreases through time. The χ^2 goodness-of-fit test for the logistic model yields a P value of 0·50, indicating that the logistic model fits these data very well.

DISCUSSION

Pedigree, isonymy, and migration data all show that inbreeding rates in the Utah Mormon population are very low. Since random inbreeding exceeds total inbreeding in this population, a slight excess of heterozygotes should be observed (Allen, 1965). A gene frequency analysis showed that all loci investigated were in Hardy–Weinberg equilibrium (McLellan *et al.* 1984) and that some loci exhibited an excess of heterozygotes while others had a deficiency. Genotype proportions are the product of multiple evolutionary forces (Workman, 1969), and the Hardy–Weinberg procedure can be quite insensitive to individual factors such as inbreeding (Neel *et al.* 1964; Jenkins *et al.* 1985). Nonetheless, these gene frequency results are at least consistent with expectations for an outbred population.

In an earlier study of consanguinity in Utah, Woolf *et al.* (1956) reported average kinship values for a total of 36909 Utah marriages. For marriages occurring between 1847 and 1929, they obtained average kinship coefficients ranging from a maximum of 0·00088 (1870–89 marriage cohort) to a minimum of 0·00026 (1910–29 marriage cohort). While these values again demonstrate a low level of consanguinity in this population, they are somewhat higher than the values reported in the present study. This may reflect in part the more complete sample of marriages used here. In addition, the computerized database is truncated in earlier years because of the inclusion criterion that a nuclear family must have had a vital event (birth or death) occurring in Utah or along the 'pioneer trail' leading to Utah. In Woolf's study, families could be traced further back in time, accounting for a greater difference between the two studies in the earlier time periods.

Several comprehensive reviews of consanguinity levels in human populations have been published (Cavalli-Sforza & Bodmer, 1971; Freire-Maia, 1957; Lebel, 1983; McCullough & O'Rourke, 1986; Reid, 1973). Comparison of the Utah Mormon inbreeding rate of approximately 10^{-4} with the rates published in these studies shows that inbreeding in this population is relatively quite low. A particularly useful comparison is provided by the Wisconsin Roman Catholic population studied by Lebel (1983). Using dispensation records, this study documents a gradual rise in average kinship coefficients to a maximum of about 4×10^{-4} at the turn of the century, followed by a gradual decline to slightly over 10^{-5} in recent years. In general, these values are quite similar to those obtained for the Utah Mormon population. This pattern of temporal decline in inbreeding, particularly during the 20th century, has been seen in most human populations (e.g. Brennan & Relethford, 1983; Imaizumi, 1986; Khlat, 1988; O'Brien *et al.* 1988; Pettener, 1985; Saugstad, 1977; Sutter & Goux, 1962) and can usually be ascribed to increased migration rates and population mixture as transportation and communication facilities improve.

While total inbreeding generally decreases through time, there is a gradual buildup of remote consanguinity. Most other genealogical studies of inbreeding show a similar pattern, and in many studies remote consanguinity can lead to a rather large inbreeding coefficient (Bear *et al.* 1988; Hussels, 1969; Leslie *et al.* 1981; O'Brien *et al.* 1988; Roberts, 1969; Spuhler & Kluckhohn, 1953). In this population, however, continued population growth and high migration rates caused total inbreeding to remain nearly constant after 1886, while random

inbreeding continued to decrease. The great majority of marriages in this population were contracted between individuals who were unrelated at any level. Other studies have also shown that population growth and migration can mitigate the effects of consanguinity buildup through time (O'Brien *et al.* 1988; Relethford, 1986; Ward *et al.* 1980).

The rural–urban comparison of inbreeding rates showed that rural inbreeding was roughly 50% greater than that of the urban stakes, while random inbreeding in rural areas was approximately 5 times higher than in the urban stakes. Woolf *et al.* (1956) also analysed a separate sample of rural Utah marriages and obtained an average kinship coefficient of 0·001 89 for 625 marriages in nine small communities. This figure is more than an order of magnitude higher than the average figure obtained in this study for 259430 individuals born in rural stakes. The communities chosen in Woolf's study, however, were selected *a priori* on the basis that they appeared to manifest high inbreeding levels and are thus probably not representative of the entire rural population. The values reported here reflect more accurately the overall inbreeding rate in rural Utah. The elevated inbreeding rate in the rural portion of this population is consistent with the findings of several other studies (Freire-Maia *et al.* 1983; Gedde-Dahl, 1973; Rao *et al.* 1972).

One of the most instructive aspects of this study is the comparison of random kinship based on migration matrices, isonymy, and pedigree data. Deficiencies in each approach are revealed by this comparison. The primary weakness of pedigree data is that inbreeding is underestimated in the early time periods due to truncation of the genealogies. In the present study, this is probably not a serious deficiency because the founders of the population came from diverse parts of the United States and northern Europe. Most were thus not likely to be related to one another.

The isonymy data consistently overestimated random inbreeding and F_{st}. Also, total inbreeding estimated by isonymy varied between 0·005 and 0·001 for the birth cohorts considered here (Jorde & Morgan, 1987), while total inbreeding measured from pedigree data varied between 0·00005 and 0·0002. Rogers (1987) compared isonymy and pedigree estimates of inbreeding in nine populations and showed that isonymy estimates exceeded pedigree estimates in every case but one. The magnitudes of these overestimates varied from 2-fold to 200-fold. Several additional studies have revealed similar results (Hurd, 1983; Roberts & Roberts, 1983; Robinson, 1983). A common explanation for these inflated estimates is the polyphyletic origin of surnames (i.e. the same surname can be derived from multiple, unrelated ancestors). This is undoubtedly an important source of error in the Utah Mormon population, since many of the Scandinavians used patronyms and many other members of the population had occupational surnames. Table 6, which shows that 92% of isonymous marriages are not consanguineous, substantiates the polyphyletic origin of surnames in this population.

Another source of overestimation has been explored by Tay & Yip (1984). They show theoretically that inbreeding estimated from isonymy is exaggerated in populations with low inbreeding values. The overestimation is even greater when random inbreeding estimated by isonymy is much larger than total inbreeding estimated from pedigrees. Both of these attributes are seen in the Mormon population as well as many other human populations. Another important assumption of the isonymy method is that males and females migrate in equal proportions (Crow & Mange, 1965). This assumption is also violated in the Utah Mormon population: males have been considerably more mobile than females (Jorde, 1982). Finally, the

isonymy method assumes that the variances in the number of offspring born to males and females are equal, an assumption which holds in strictly monogamous societies (Crow, 1983). Polygyny was practised in Utah during the 19th century. While only a small minority of males had multiple wives, polygyny substantially increased the variance of progeny size among males (Jorde & Durbize, 1986). Considering these assumptions and results, it is clear that isonymy methods are most reliable when applied to small, closed populations in which a limited number of distinct surnames were present among the founders (Crow, 1980).

While isonymy tends to overestimate inbreeding and genetic differentiation in this population, the migration matrix approach underestimates these quantities. This result stands in contrast to two other studies that compared these two approaches and found *higher* estimates of F_{st} when migration data were used (Fuster, 1986; Relethford, 1986). In the present study, the differences are clearly attributable to assumptions inherent in the migration matrix model. One of these is that immigrants are derived from a homogeneous outside population. A collateral assumption is that the initial founding population is genetically homogeneous. Both of these assumptions are inaccurate for the Utah population and have led to underestimates of random inbreeding and genetic differentiation. A third assumption, to be discussed in greater detail below, is that migration patterns among subdivisions are at equilibrium (i.e. they do not change from one generation to the next). In this rapidly colonizing population, migration patterns changed considerably through time. While the migration matrix, isonymy, and pedigree methods yielded somewhat divergent estimates of F_{st}, it should be emphasized that all of these estimates are in the low range for human populations (see Jorde (1980) and Relethford (1988) for comparative values).

In evaluating inbreeding estimates based on different types of data, it is important to emphasize that inbreeding is always measured *relative* to a given reference population (Wright, 1969). For the genealogy data, the reference population is the large set of founders who initially came to Utah. Isonymy estimates, like gene frequency estimates, reflect the effects of events occurring many generations in the past (Crow, 1983). The reference populations represented by the isonymy and genealogy measures are therefore quite different, and this may account for some of the differences observed in the estimates.

The comparison of each type of kinship matrix with geographic distance showed that, considering all time periods, the migration matrix estimates yielded the highest correlation with geographic distance, while those of isonymy had the lowest correlation with geographic distance. This result was obtained using both the Mantel technique and the eigenvector technique. This pattern did not hold, however, in each individual time period. It was suggested previously that the lower correlation between isonymy and geographic distance is due to the nonrandom settlement of Utah by different northern European population groups (Jorde & Morgan, 1987). In a recent analysis of a French Canadian population, Gradie *et al.* (1988) also found low concordance between isonymy and geographic distance while obtaining good concordance between migration-derived kinship and geographic distance. Smith (1988) obtained a similar result in an analysis of a British population. After reviewing the literature on such comparisons, Jorde & Morgan (1987) concluded that the concordance between isonymy and geographic distance seems to be low in recently founded populations and higher in well-established populations. This reflects the accumulation of an isolation by distance effect over time.

Direct comparison of all three types of kinship matrices using the Mantel technique showed substantially greater congruence between isonymy- and pedigree-derived kinship estimates than between the other 2 pairs of estimates. It is very interesting, though, that the migration matrix estimates, while not correlating highly with the other estimates in individual time periods, correlated very highly with both estimates when all time periods were combined. In addition, the eigenvector plots based on pedigree and migration data are extremely similar. This suggests that, by combining data over 70 years, a pattern more similar to 'equilibrium' is being obtained, yielding a more reliable estimate of between-groups kinship. While there were rather large discrepancies among the differet data types in estimates of random and total inbreeding coefficients, the *patterns* of between-subdivision relationships are remarkably similar. This seems reasonable, since most of the assumptions discussed above tend to bias these estimates in the same direction in each subdivision (overestimation of values derived from isonymy, underestimation of values derived from migration matrices). Provided that the biases are fairly consistent among subdivisions, they will not distort *between*-groups relationships.

The regression analyses are useful in helping to determine the causes of variation in inbreeding patterns. For the most part, the linear and logistic regression analyses yielded similar results. Both indicated that population size is negatively correlated with inbreeding. The linear regression showed that geographic distance between husband's and wife's birthplaces is negatively associated with inbreeding, while the logistic regression showed that stake endogamy is positively associated with inbreeding. Since geographic distance and endogamy show a strong inverse correlation ($r = -0{\cdot}64$), these results are congruent. As explained above, the differing signs for the regression coefficient for 'year of marriage' in the two analyses reflect the buildup of remote consanguinity which is measured in the linear regression but not in the logistic regression. In a previous logistic regression analysis of isonymy data, year of marriage and geographic distance both correlated negatively with probability of isonymous marriage (Jorde & Morgan, 1987). These results are also consistent with those of the present study and with population genetic theory. The main difference between the regression analyses based on pedigree and isonymy data is that the isonymy analysis did not indicate a population size effect. The strong dependency of consanguinity upon population size in individual marriages is consistent with the correlation between subdivision size and average within-subdivision inbreeding levels. Considering that the great majority of isonymous marriages are not in fact consanguineous, one would not expect isonymy to exhibit as high a degree of dependency upon population size.

In summary, the pedigree results presented here largely corroborate previous results based on migration matrices, isonymy, and gene frequencies: the Utah Mormon population is outbred, homogeneous, and has experienced little genetic drift since its founding. As a consequence, one would expect the distribution and prevalence of genetic diseases in this population to be quite similar to those of other U.S. populations. While an extensive inventory of genetic diseases has not yet been carried out in Utah, the prevalence rates of certain genetic diseases (or diseases with genetic components) have been estimated. Among 192083 Utah births from 1983 to 1987, 15 were affected with classical PKU (C. O. Leonard, M.D., personal communication). This gives a birth prevalence estimate of 1/12806, a figure which is in accord with the commonly cited estimates of 1/10000 to 1/15000 for PKU in Caucasian populations (Scriver & Clow, 1980). In addition, hemochromatosis (Edwards *et al.* 1988), neural tube defects

(Jorde *et al.* 1983), and autism (Ritvo *et al.* 1989) all occur with frequencies similar to those found in U.S. and European populations.

The comparison of kinship estimates based on genealogies, isonymy, and migration matrices showed that the assumptions underlying each method can produce divergent estimates of total inbreeding and random kinship. If reliable estimates of inbreeding are to be obtained with isonymy and migration data, investigators must devote considerable attention to possible violations of these assumptions. On a more sanguine note, these methods yielded fairly consistent results in terms of between-groups kinship patterns. Further studies in other populations are needed to confirm this optimism.

This research was supported by NIH grant HD-16109 and NSF grants BNS-8703841 and BNS-8720330. Drs Margaret Gradie, Richard Kerber, Paul Leslie, Jeffrey Long, Kenneth Morgan, Elizabeth O'Brien and Alan Rogers provided assistance and useful comments on the manuscript. Dr Claire Leonard provided data on PKU births from the Utah State Health Department's Newborn Metabolic Screening Program.

REFERENCES

Allen, G. (1965). Random and nonrandom inbreeding. *Eugen. Quart.* **12**, 181–198.

Bear, J. C., Nemec, T. F., Kennedy, J. C., Marshall, W. H., Power, A. A., Kolonel, V. M. & Burke, G. B. (1988). Inbreeding in outport Newfoundland. *Am. J. Med. Genet.* **29**, 649–660.

Brennan, E. R. & Relethford, J. H. (1983). Temporal variation in the mating structure of Sanday, Orkney Islands. *Ann. Hum. Biol.* **10**, 265–280.

Cavalli-Sforza, L. L. & Bodmer, W. F. (1971). *The Genetics of Human Populations*. San Francisco: W. H. Freeman.

Crow, J. F. (1980). The estimation of inbreeding from isonymy. *Hum. Biol.* **52**, 1–14.

Crow, J. F. (1983). Discussion. *Hum. Biol.* **55**, 383–397.

Crow, J. F. & Mange, A. P. (1965). Measurement of inbreeding from the frequency of marriages between persons of the same surname. *Eugen. Quart.* **12**, 199–203.

Dow, M. M. & Cheverud, J. M. (1985). Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. *Am. J. Phys. Anthrop.* **68**, 367–373.

Edwards, C. Q., Griffin, L. M., Goldgar, D., Drummond, C., Skolnick, M. H. & Kushner, J. P. (1988). Prevalence of hemochromatosis among 11,065 presumably healthy blood donors. *New Engl. J. Med.* **318**, 1355–1362.

Felsenstein, J. (1982). How can we infer geography and history from gene frequencies? *J. Theor. Biol.* **96**, 9–20.

Freire-Maia, N. (1957). Inbreeding levels in different countries. *Eugen. Quart.* **4**, 127–138.

Freire-Maia, N., Chautard-Freire-Maria, E. A., De Aguiar-Wolter, I. P., Azevedo-Fialho, M., Barros De Azevedo, M., Krieger, H. & Barbosa, C. A. A. (1983). Inbreeding studies in Brasilian schoolchildren. *Am. J. Med. Genet.* **16**, 331–355.

Fuster, V. (1986). Relationship by isonymy and migration pattern in Northwest Spain. *Hum. Biol.* **58**, 391–406.

Gedde-Dahl, T. (1973). Population structure in Norway. *Hereditas* **73**, 211–231.

Gradie, M., Jorde, L. B. & Bouchard, G. (1988). Genetic structure of the Saguenay, 1852–1911: evidence from migration and isonymy matrices. *Am. J. Phys. Anthrop.* **77**, 321–333.

Harpending, H. C. & Jenkins, T. (1974). !Kung population structure. In *Genetic Distance* (ed. J. F. Crow and C. Denniston), pp. 137–165. New York: Plenum Press.

Hurd, J. P. (1983). Comparison of isonymy and pedigree analysis measures in estimating relationships between three 'Nebraska' Amish churches in central Pennsylvania. *Hum. Biol.* **55**, 349–355.

Hussels, I. (1969). Genetic structure of Saas, a Swiss isolate. *Hum. Biol.* **41**, 469–479.

Imaizumi, Y. (1986). A recent survey of consanguineous marriages in Japan: religion and socioeconomic class effects. *Ann. Hum. Biol.* **13**, 317–330.

Jenkins, T., Beighton, P. & Steinberg, A. G. (1985). Serogenetic studies on the inhabitants of Tristan da Cunha. *Ann. Hum. Biol.* **12**, 363–371.

Jorde, L. B. (1980). The genetic structure of subdivided human populations: a review. In *Current Developments in Anthropological Genetics*, vol 1 (ed. J. H. Mielke and M. H. Crawford), pp. 135–208. New York: Plenum Press.

JORDE, L. B. (1982). The genetic structure of the Utah Mormons: migration analysis. *Hum. Biol.* **54**, 583–597.
JORDE, L. B. (1984). A comparison of parent–offspring and marital migration data as measures of gene flow. In *Migration and Mobility* (ed. A. J. Boyce), pp. 83–96. London: Taylor & Francis.
JORDE, L. B. & DURBIZE, P. (1986). Opportunity for natural selection in the Utah Mormons. *Hum. Biol.* **58**, 97–114.
JORDE, L. B., FINEMAN, R. M. & MARTIN, R. A. (1983). Epidemiology and genetics of neural tube defects: an application of the Utah genealogical data base. *Am. J. Phys. Anthrop.* **62**, 23–31.
JORDE, L. B. & MORGAN, K. (1987). Genetic structure of the Utah Mormons: isonymy analysis. *Am. J. Phys. Anthrop.* **72**, 403–412.
KHLAT, M. (1988). Social correlates of consanguineous marriage in Beirut: a population based study. *Hum. Biol.* **60**, 541–548.
LALOUEL, J. M. (1973). Topology of population structure. In *Genetic Structure of Populations* (ed. N. E. Morton), pp. 139–149. Honolulu: University of Hawaii Press.
LEBEL, R. R. (1983). Consanguinity studies in Wisconsin. I. Secular trends in consanguineous marriage, 1843–1981. *Am. J. Med. Genet.* **15**, 543–560.
LESLIE, P. W., MORRILL, W. T. & DYKE, B. (1981). Genetic implications of mating structure in a Caribbean isolate. *Am. J. Hum. Genet.* **33**, 90–104.
MALÉCOT, G. (1948). *Les Mathématiques de l'hérédité*. Paris: Masson.
MANTEL, N. A. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.
MARTIN, T. K., HEATON, T. B. & BAHR, S. J. (eds.) (1986). *Utah in Demographic Perspective*. Salt Lake City: Signature Books.
MCCULLOUGH, J. M. & O'ROURKE, D. H. (1986). Geographic distribution of consanguinity in Europe. *Ann. Hum. Biol.* **13**, 359–367.
MCLELLAN, T., JORDE, L. B. & SKOLNICK, M. H. (1984). Genetic distances between the Utah Mormons and related populations. *Am. J. Hum. Genet.* **36**, 836–857.
NEEL, J. V., SALZANO, F. M., JUNQUEIRA, P. C., KEITER, F. & MAYBURY-LEWIS, D. (1964). Studies on the Xavante Indians of the Brazilian Mato Grosso. *Am. J. Hum. Genet.* **16**, 52–140.
O'BRIEN, E., JORDE, L. B., RÖNNLÖF, B., FELLMAN, J. O. & ERIKSSON, A. W. (1988). Inbreeding and genetic disease in Sottunga, Finland. *Am. J. Phys. Anthrop.* **75**, 477–486.
PETTENER, D. (1985). Consanguineous marriages in the Upper Bologna Appennine (1565–1980): Micro-geographic variations, pedigree structure and correlation of inbreeding secular trend with changes in population size. *Hum. Biol.* **57**, 267–288.
RAO, P. S. S. S., INBARAJ, S. G. & JESUDIAN, G. (1972). Rural-urban differentials in consanguinity. *J. Med. Genet.* **9**, 174–178.
REID, R. M. (1973). Inbreeding in human populations. In *Methods and Theories of Anthropological Genetics* (ed. M. H. Crawford and P. L. Workman), pp. 83–116. Albuquerque: University of New Mexico Press.
RELETHFORD, J. H. (1986). Microdifferentiation in historical Massachusetts: A comparison of migration matrix and isonymy analyses. *Am. J. Phys. Anthrop.* **71**, 365–375.
RELETHFORD, J. H. (1988). Estimation of kinship and genetic distance from surnames. *Hum. Biol.* **60**, 475–492.
RITVO, E. R., FREEMAN, B. J., PINGREE, C., MASON-BROTHERS, A., JORDE, L. B., JENSON, W. R., MCMAHON, W. M., PETERSEN, B., MO, A. & RITVO, A. (1989). The UCLA–University of Utah epidemiologic survey of autism: prevalence. *Am. J. Psychiat.* **146**, 194–199.
ROBERTS, D. F. (1969). Consanguineous marriage and calculation of the genetic load. *Ann. Hum. Genet.* **32**, 407–410.
ROBERTS, D. F. & ROBERTS, M. J. (1983). Surnames and relationships: an Orkney study. *Hum. Biol.* **55**, 341–347.
ROBINSON, A. P. (1983). Inbreeding as measured by dispensations and isonymy on a small Hebridean island, Eriskay. *Hum. Biol.* **55**, 289–295.
ROGERS, A. R. & HARPENDING, H. C. (1986). Migration and genetic drift in human populations. *Evolution* **40**, 1312–1327.
ROGERS, L. A. (1987). Concordance in isonymy and pedigree measures of inbreeding: the effects of sample composition. *Hum. Biol.* **59**, 753–767.
SAUGSTAD, L. F. (1977). Inbreeding in Norway. *Ann. Hum. Genet.* **40**, 481–491.
SCRIVER, C. R. & CLOW, C. L. (1980). Phenylketonuria and other phenylalanine hydroxylation mutants in man. *Ann. Rev. Genet.* **14**, 179–202.
SMITH, M. T. (1988). The Isle of Wight in 1851: historical migration and genetic structure. *Actes des 3èmes Journées Anthropologiques* **24**, 113–122.
SMOUSE, P. E., LONG, J. C. & SOKAL, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* **35**, 627–632.
SOKAL, R. R. & ROHLF, F. J. (1981). *Biometry* (2nd ed.). San Francisco: W. H. Freeman and Company.

Spuhler, J. N. & Kluckhohn C. (1953). Inbreeding coefficients of the Ramah Navaho population. *Hum. Biol.* **25**, 295–317.

Sutter, J. & Goux, J. M. (1962). Decline in consanguineous marriages in France from 1926 to 1958. *Eugen. Quart.* **11**, 127–140.

Tay, J. S. H. & Yip, W. C. L. (1984). The estimation of inbreeding from isonymy: relationship to the average inbreeding coefficient. *Ann. Hum. Genet.* **48**, 185–194.

Wahlquist, W. L. (1978). Population growth in the Mormon core area: 1847–90. In *The Mormon Role in the Settlement of the West* (ed. R. H. Jackson), pp. 107–134. Provo: Brigham Young University Press.

Ward, R. H., Raspe, P. D., Ramirez, M. E., Kirk, R. L. & Prior, I. A. M. (1980). Genetic structure and epidemiology: the Tokelau study. In *Population Structure and Genetic Disorders* (ed. A. W. Eriksson, H. R. Forsius, H. R. Nevanlinna, P. L. Workman and R. K. Norio), pp. 301–325. New York: Academic Press.

Wood, J. W. (1986). Convergence of genetic distances in a migration matrix model. *Am. J. Phys. Anthrop.* **71**, 209–219.

Woolf, C. M., Stephens, F. E., Mulaik, D. D. & Gilbert, R. E. (1956). An investigation of the frequency of consanguineous marriages among the Mormons and their relatives in the United States. *Am. J. Hum. Genet.* **8**, 236–252.

Workman, P. L. (1969). The analysis of simple genetic polymorphisms. *Hum. Biol.* **41**, 97–114.

Wright, S. (1922). Coefficients of inbreeding and relationship. *Am. Nat.* **56**, 330–338.

Wright, S. (1943). Isolation by distance. *Genetics* **28**, 114–138.

Wright, S. (1969). *Evolution and the Genetics of Populations*, vol. 2. Chicago: University of Chicago Press.