Genetic Structure of the Utah Mormons: Comparison of Results Based on RFLPs, Blood Groups, Migration Matrices, Isonymy, and Pedigrees

ELIZABETH O'BRIEN,[1] ALAN R. ROGERS,[2] JUDY BEESLEY,[3] AND LYNN B. JORDE[1]

Abstract The genetic structure of the Utah Mormon population is examined using 25 blood group and 47 RFLP alleles obtained from 442 subjects living in 8 geographic subdivisions. Nei's G_{ST} was 0.013 ($p < 0.002$) for the RFLP data and 0.012 ($p > 0.4$) for the blood group data, showing that only 1% of the genetic variance in this population can be attributed to subdivision effects. A comparison of intersubdivision distance matrices based on blood groups, RFLPs, migration matrices, isonymy, and pedigrees shows that genetic distances have relatively low and nonsignificant correlations with the other three types of data. However, the correlations based on RFLPs are considerably higher than those based on blood groups. Relationship matrices based on interindividual allele sharing were compared with known genealogical kinship coefficients between each pair of individuals. The correlation between the blood group and RFLP relationship matrices was small but marginally significant using the Mantel test ($r = 0.014$, $p < 0.06$). The RFLP relationship matrix correlated more highly with genealogical kinship than did the blood group relationship matrix ($r = 0.023$, $p < 0.0001$ and $r = 0.012$, $p < 0.001$, respectively). These correlations increased by approximately one order of magnitude when pairs of subjects having zero kinship coefficients were excluded. These results show that genetic distances derived from RFLPs correlate more strongly with other types of kinship than do distances based on blood groups. This probably reflects the fact that RFLPs are more neutral, have frequencies that are more accurately estimated, and contain more information about DNA sequence variation.

The genetic structure of human populations can be estimated from several different types of data, including migration matrices, isonymy, pedigrees, anthropometrics, and gene frequencies. The use of each type of data involves certain assumptions, advantages, and disadvantages. Much

[1]Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT 84112.
[2]Department of Anthropology, University of Utah, Salt Lake City, UT 84112.
[3]Blood Bank, Latter-Day Saints Hospital, Salt Lake City, UT 84103.

Human Biology, October 1994, v. 66, no. 5, pp. 743–759.

can be learned about the effects of these assumptions and the relative merits of each data type by comparing their results in the same population. Because DNA polymorphisms are now an important source of information about human evolution (Balazs et al. 1992; Bowcock et al. 1991; Cann et al. 1987; Deka et al. 1991; Edwards et al. 1992; Harpending et al. 1993; Horai et al. 1993; Kidd et al. 1991; Martinson et al. 1993; Stoneking et al. 1990; Torroni et al. 1990), it is particularly relevant to establish their utility for estimating genetic structure by comparing them with other types of data.

The Utah Mormon population has been the subject of genetic structure studies based on migration matrices (Jorde 1982), isonymy (Jorde and Morgan 1987), pedigrees (Jorde 1989), and gene frequencies estimated from blood groups and protein electrophoresis (McLellan et al. 1984). These studies have shown that the Utah Mormon population has a low inbreeding rate and is genetically quite homogeneous. Gene frequency analyses have demonstrated that the Utah Mormon population is genetically similar to the northern European populations from which it was derived, with little indication of genetic drift. A recent study revealed no evidence of a founder effect (O'Brien, Kerber et al. 1994). These results are all consistent with the demographic history of the population, which was founded in the mid-nineteenth century by over 100,000 individuals, most of whom were derived from northern Europe. High birth rates and a high rate of immigration contributed to rapid growth and a high degree of population mobility.

Given the extent of previous studies, the Utah Mormon population presents a useful opportunity for comparing DNA polymorphism results with those derived from other types of data. Here, we report an analysis of nuclear RFLP (restriction fragment length polymorphism) and blood group variation in 442 Utah males from 8 geographic subdivisions. To our knowledge this is the first time that genetic kinship coefficients based on RFLPs, blood groups, migration matrices, isonymy, and pedigrees have been estimated and compared in the same human population.

Materials and Methods

The study population consists of 442 males born in 8 Utah counties [geographically equivalent to "stakes," the unit of subdivision used in previous analyses (Jorde 1982)]. The counties, shown in Figure 1, were chosen to maximize between-group genetic diversity, based on the results of previous genetic structure analyses. All the study subjects are members of the Utah Population Data Base (UPDB), a computerized genealogical database that consists of 1.1 million individuals linked into large pedigrees. A random sample of UPDB members born between 1941

Figure 1. Map of Utah showing the locations of the eight counties from which subjects were sampled.

and 1955 was ascertained for each county. Analysis of these subjects' genealogies showed that their collective ancestry is 61% British (including Welsh, Irish, and Scottish) and 31% Scandinavian, with Swiss and German ancestry accounting for most of the remainder [see O'Brien, Zenger et al. (1994) for further details]. The ethnic heritage of these subjects closely reflects the known historical origins of Utah's founders (Arrington and Bitton 1979; McLellan et al. 1984).

Each subject was located by obtaining records from the Utah Division of Motor Vehicles. Initial contact was made by mail, with telephone followup. Approximately 50% of those subjects who were contacted agreed to participate in the study. Because ascertainment was random and no financial remuneration was offered, this response rate should produce no systematic bias.

After obtaining informed consent (the protocol was approved by the Institutional Review Board of the University of Utah Health Sciences Center), a 50-ml sample of venous blood was drawn from each study subject. Blood typing for the ABO, Rh, MNS, Lewis, P, Kell, Duffy, and Kidd blood groups was carried out at the Latter-Day Saints Hospital, Salt Lake City, Utah.

To assay RFLPs, 5 μg of DNA was digested to completion with specific restriction enzymes in 35-μl reaction volumes. Digests were prepared in 96-well trays using a Biomek 1000 workstation and left overnight at a temperature recommended by the enzyme manufacturers. Digested DNA fragments were separated by size in 0.8% agarose gels, which were electrophoresed overnight at approximately 35 V. DNA fragments were then transferred to Hybond+ nylon membrane (Amersham) in 4 *N* NaOH for 6 hr or overnight. Membranes were rinsed (2× SSC), lightly blotted, and prehybridized in SSPE (5×) hybridization solution. Probes, which were obtained from the laboratory of Raymond White at the University of Utah, were labeled with ^{32}P (specific activity = 1.5×10^6 cpm/ml hybridization volume), added to 65°C hybridization solution, and incubated with the membranes overnight at 65°C. Membranes were washed once in 2× SSC/0.1% SDS, once in 0.1× SSC/0.1% SDS for approximately 15 min at room temperature, and a final time in 0.1× SSC/0.1% SDS at 60°C. Membranes were lightly blotted, wrapped in Saran Wrap, and exposed to x-ray film with intensifying screens for several hours or overnight.

Gene frequencies for each subpopulation were estimated by maximum likelihood for the systems in which there is dominance (ABO, Rh, MNS, P, and Lewis). The other blood group systems and all RFLP systems are codominant, so gene frequencies for these systems were estimated by direct gene counting. Among the codominant systems, none of the genotype frequencies deviated significantly from Hardy-Weinberg proportions.

Genetic variation in the population was estimated using the G_{ST} statistic (Nei 1987). This statistic provides an estimate of the proportion of genetic variance in the population attributable to subdivision, both for diallelic and multiallelic systems. To test the null hypothesis that $G_{ST} = 0$, subjects were randomly assigned to each of the eight subdivisions. Gene frequencies and G_{ST} were then reestimated 1000 times to generate a null distribution of G_{ST} values. Statistical significance was assessed by comparing the estimated G_{ST} value with this distribution. If, for example, the estimated G_{ST} exceeded 95% of the values in the distribution, a type I error level of 0.05 was assigned.

Kinship coefficients among subdivisions, r_{ij}, were estimated using the *R* matrix technique (Harpending and Jenkins 1973). Genetic distances

d_{ij} were obtained from the kinship coefficients using the standard formula

$$d_{ij} = r_{ii} + r_{jj} - 2r_{ij}. \quad (1)$$

Distance matrices estimated from RFLPs and blood groups were compared using the Mantel matrix comparison method (Mantel 1967; Smouse et al. 1986). This method randomly permutes columns of one of the matrices to generate a null distribution of correlation coefficients. The estimated correlation coefficient for a pair of matrices is then compared with this distribution to obtain a significance level. The blood group and RFLP distance matrices were also compared with distance matrices based on migration matrices, isonymy, pedigrees, and geographic location. These data have been published previously (Jorde 1982, 1989; Jorde and Morgan 1987).

In addition to examining genetic similarity among subdivisions, we assessed genetic similarity among individuals by estimating the proportion of identical alleles for each possible pair of individuals. This procedure is direct for codominant systems, including all the RFLPs. For systems in which there is dominance it is not possible to distinguish the homozygous dominant genotype from the heterozygote. There are many ways of dealing with such missing data problems (Little and Rubin 1987). Karlin et al. (1982) advocate a Monte Carlo approach in which individuals are assigned a second allele randomly, based on the gene frequency in the population. Thus, if a subject had one copy of the dominant allele and the frequency of this allele in the population was 0.70, the subject would have a 70% chance of being randomly assigned the dominant allele as the second allele. The effects of this procedure on our results will be discussed later.

Two relationship matrices were generated: one for blood groups and one for RFLPs. Because each study subject was a member of the genealogical database, it was also possible to estimate the genealogical kinship coefficient for each pair of subjects. This provided a third relationship matrix, based on known kinship coefficients. The three relationship matrices were compared using the Mantel matrix comparison technique. Ten thousand random permutations were carried out for each comparison to achieve a high degree of accuracy in estimating the significance levels. Despite the large size of these matrices (442 × 442), an optimized computer algorithm performed the permutations in approximately 1 hr of dedicated time on a Sun IPC workstation. [This C program is in a file named mantel.tar, available by means of anonymous FTP from anthro.utah.edu.]

Results

Table 1 gives probe-enzyme combinations, chromosome location, and heterozygosity for each RFLP system. Twenty of the 22 RFLPs were

Table 1. Probe-Enzyme Combinations, Chromosome Locations, and Heterozygosity for Each RFLP

Probe	*Enzyme*	*Chromosome*	*Heterozygosity*
4c11	*Bgl*II	6p	0.43
pMS3-18	*Bgl*II	22q	0.39
WC64	*Bgl*II	13q	0.40
MCR3	*Eco*RI	1q	0.37
pW236b	*Eco*RI	21q	0.42
DV1.9	*Hin*dIII	17q	0.48
HcoII	*Hin*dIII	12q	0.48
EFZ33	*Hin*dIII	15	0.38
EFZ31	*Msp*I	22	0.47
EFD122	*Msp*I	2p	0.49
pPW228c	*Msp*I	21q	0.43
WC25	*Msp*I	13q	0.42
T593	*Pst*I	7p	0.43
PSL	*Pst*I	17q	0.41
EFD70.2	*Pvu*II	10	0.51
EKX3B	*Pvu*II	16	0.28
pHHH163	*Pvu*II	18p	0.47
TB10.171	*Pvu*II	10	0.43
CRAF-1	*Taq*I	3p	0.35
EKH7.4	*Taq*I	1q	0.50
9F11	*Taq*I	12q	0.44
M5.12E-3	*Taq*I	5q	0.43

diallelic systems. The WC25 polymorphism had three alleles, and the EFD70.2 system had four alleles. Gene frequencies for each blood group and RFLP system in each subdivision are given in Table 2. This table also provides the sample size from each subdivision; these ranged from 52 to 58.

Table 3 presents the average gene diversity (or heterozygosity) within subdivisions (H_s), total gene diversity (H_t), gene identity between subdivisions (D_{ST}), and the standardized measure G_{ST} for the blood group and RFLP systems. Significance levels for G_{ST}, based on the random assignment procedure, are also given. The table shows that a relatively small proportion of the genetic variation in this population, slightly over 1%, is due to subdivision effects. The G_{ST} estimate from blood groups, 0.013, is similar to the value estimated from RFLPs, 0.012. Although the statistic based on RFLPs is highly significant ($p < 0.002$), the G_{ST} value based on blood groups is nonsignificant ($p > 0.4$), possibly because RFLP alleles outnumber blood group alleles by nearly 2 to 1 (47 vs. 25). To test this possibility, we computed a second G_{ST} using only 10 randomly chosen RFLPs (20 alleles). The value obtained from this data set was similar to the previous one (0.10) and was still marginally

significant ($0.08 < p < 0.09$). If a correction for sampling variance is applied to the G_{ST} estimates (Workman et al. 1973), they become substantially lower (0.005 and 0.003 for the blood groups and RFLPs, respectively).

The results of comparing the intersubdivision blood group and RFLP distance matrices to those based on isonymy, migration matrices, pedigrees, and geographic distance are shown in Table 4. None of the correlations are statistically significant. However, all correlations involving the RFLP distance matrix substantially exceed those involving the blood group distance matrix. The correlation between the blood group and RFLP distance matrices was low and nonsignificant ($r = 0.064$, $p > 0.4$).

Table 5 presents the results of Mantel matrix comparisons at the individual level. The RFLP and blood group similarity matrices show a low and marginally significant correlation. A comparison of the blood group and genealogical kinship matrices also reveals a low but more significant correlation. The correlation between the RFLP matrix and genealogical kinship is about twice as high as that of the blood group matrix and is highly significant. The highest correlation is seen when the RFLP and blood group data are combined and then compared with genealogical kinship.

A possible explanation for the higher correlations involving the RFLP data is that they include more independent alleles than do the blood groups. To test this, we randomly divided the RFLP systems into two halves and then compared them with genealogical kinship. Although both correlations were lower than the correlation involving the entire RFLP data set, both were still higher than the blood group correlation (Table 5). A second explanation is that accurately estimating the proportion of shared alleles may be difficult when using blood groups because direct estimation was possible for only four codominant systems (MN, Kell, Duffy, and Kidd). To evaluate this possibility, we formed a similarity matrix based on only these four systems. This matrix exhibited a higher correlation with the RFLP matrix than did the total blood group matrix ($r = 0.026$, $p < 0.008$), whereas the correlation between the other blood group systems and the RFLP matrix was low and nonsignificant ($r = 0.006$, $p > 0.20$). However, the similarity matrix based on codominant blood groups demonstrated a correlation with genealogical kinship even lower than the correlation using the total blood group matrix ($r = 0.0097$, $p < 0.01$). Comparisons at the subdivision level also yielded very low ($r < 0.01$) nonsignificant correlations between a distance matrix based on the codominant blood groups and all other distance matrices (RFLP, migration, isonymy, and pedigree).

Because the Utah Mormon population is an outbred population, most of the genealogical kinship coefficients (98%) are zero. Most of the remaining coefficients are relatively small (Figure 2). Because these ge-

Table 2. Blood Group Gene Frequencies and RFLP Frequencies in Each Subdivision

	County and Sample Size							
Blood Group or RFLP	*Beaver* (n = 55)	*Box Elder* (n = 55)	*Cache* (n = 55)	*Davis* (n = 52)	*Salt Lake* (n = 56)	*Sanpete* (n = 58)	*Wasatch* (n = 56)	*Washington* (n = 55)
Blood group[a]								
A	0.2274	0.2317	0.3778	0.2744	0.2511	0.3416	0.2200	0.2765
B	0.0566	0.0920	0.0561	0.0717	0.0761	0.0398	0.0380	0.0590
O	0.7160	0.6763	0.5661	0.6539	0.6728	0.6186	0.7420	0.6645
CDE	0.0000	0.0167	0.0000	0.0000	0.0000	0.0000	0.0165	0.0000
CDe	0.4167	0.4833	0.4302	0.4412	0.3455	0.3958	0.4435	0.4327
cDE	0.1481	0.0550	0.1323	0.1765	0.1091	0.1042	0.1189	0.1250
cDe	0.0000	0.0000	0.0253	0.0000	0.0000	0.0615	0.0293	0.0000
cdE	0.0000	0.0183	0.0421	0.0000	0.0000	0.0000	0.0245	0.0000
cde	0.4352	0.4267	0.3701	0.3823	0.5454	0.4385	0.3673	0.4423
MS	0.2946	0.1918	0.1724	0.1676	0.2863	0.2534	0.1719	0.1775
Ms	0.2145	0.3915	0.3504	0.3424	0.3740	0.2792	0.3474	0.3706
NS	0.0327	0.0790	0.1117	0.1224	0.0156	0.0727	0.0493	0.1687
Ns	0.4582	0.3377	0.3655	0.3676	0.3241	0.3947	0.4314	0.2832
Le	0.7303	0.8000	0.7389	0.7172	0.6985	0.6464	0.6331	0.6899
P1	0.5329	0.5528	0.4161	0.4000	0.4955	0.4796	0.6331	0.4629
K	0.0636	0.0400	0.0455	0.0200	0.0636	0.0208	0.0385	0.0385
*FY*A*	0.3545	0.5510	0.3630	0.5500	0.4455	0.4896	0.4184	0.4231
*JK*A*	0.5545	0.5000	0.4659	0.6100	0.5000	0.4271	0.5288	0.4904

RFLPs[b]								
4c11	0.6909	0.7222	0.7000	0.6635	0.7255	0.7091	0.6546	0.6563
MS3	0.8039	0.6667	0.7453	0.7308	0.7549	0.7222	0.6981	0.7500
WC64	0.3426	0.2685	0.2830	0.2157	0.2353	0.2232	0.3519	0.3269
MCR3	0.7353	0.7157	0.7925	0.7596	0.6887	0.8276	0.6727	0.7778
W236	0.2963	0.2115	0.3396	0.2700	0.3241	0.3796	0.3019	0.3192
DV1.9	0.3889	0.3889	0.3207	0.4020	0.4091	0.4569	0.4630	0.4020
*Hco*II	0.6852	0.5648	0.5943	0.5100	0.5000	0.6207	0.5865	0.6111
EFZ33	0.1667	0.1981	0.2404	0.2857	0.2963	0.3727	0.2037	0.3039
EFZ31	0.6944	0.6100	0.6389	0.5588	0.6442	0.6518	0.5943	0.6018
EFD12	0.5833	0.4712	0.4528	0.4808	0.4352	0.4483	0.5364	0.4519
PW228	0.2453	0.2778	0.3396	0.2900	0.3207	0.3793	0.3273	0.3611
WC25 (1)	0.1981	0.1509	0.2359	0.1731	0.1509	0.2155	0.2736	0.2130
WC25 (2)	0.7453	0.7453	0.6604	0.7596	0.8207	0.6897	0.6981	0.6944
T593	0.3404	0.3019	0.2963	0.2788	0.3333	0.2364	0.6182	0.3774
PSL	0.6091	0.6759	0.7453	0.7500	0.8365	0.6552	0.7091	0.6727
EFD70 (1)	0.0106	0.0865	0.0000	0.0200	0.0660	0.0182	0.0189	0.0200
EFD70 (2)	0.4681	0.3173	0.3100	0.3900	0.3207	0.3636	0.3491	0.3500
EFD70 (3)	0.5213	0.5865	0.6700	0.5600	0.5943	0.6000	0.6132	0.6000
EKX3B	0.1698	0.1944	0.1321	0.1827	0.1389	0.1810	0.1852	0.1090
HHH	0.3909	0.4722	0.3868	0.2885	0.4074	0.3621	0.4182	0.3636
TB10	0.6727	0.6132	0.7358	0.7692	0.6961	0.7091	0.6636	0.5943
CRAF	0.8019	0.7308	0.8061	0.7551	0.7059	0.8036	0.7925	0.8061
EKH7	0.5000	0.5094	0.4327	0.4412	0.4362	0.4825	0.5377	0.4808
9F11	0.7358	0.6574	0.6204	0.6471	0.6827	0.6053	0.7358	0.6939
M5.12	0.3039	0.3617	0.3173	0.3137	0.2547	0.3333	0.3300	0.2778

a. Only one allele is shown from each of the diallelic systems.
b. One allele is omitted from each system.

Table 3. Gene Diversity Measures

	H_s	H_t	D_{ST}	G_{ST}
Blood groups	0.467	0.474	0.007	0.013 ($p > 0.4$)
RFLPs	0.427	0.432	0.005	0.012 ($p < 0.002$)
RFLPs (10 loci only)	0.430	0.434	0.004	0.010 ($p < 0.08$)

Table 4. Correlation Coefficients and Significance Levels (in Parentheses) for RFLP and Blood Group Intersubdivision Genetic Distances versus Other Distance Matrices

	Migration Matrix	*Isonymy*	*Genealogies*	*Geographic Distance*
Blood groups	0.036 (>0.43)	0.111 (>0.33)	0.008 (>0.48)	−0.189 (>0.15)
RFLPs	0.280 (>0.20)	0.170 (>0.28)	0.216 (>0.27)	−0.270 (>0.13)

Table 5. Correlation Coefficients and Significance Levels for Individual-Level Comparisons

	All Pairs (N = 97,461)		*Pairs with Nonzero Kinship* (N = 1512)	
Comparison	r	p	r	p
RFLP/blood group	0.014	<0.060	–	–
RFLP/blood group (4 codominant systems)	0.026	<0.001	–	–
RFLP/blood group (noncodominant systems)	0.006	>0.200	–	–
Genealogy/blood group	0.012	<0.001	0.089	<0.0001
Genealogy/blood group (4 codominant systems)	0.010	<0.008	0.078	<0.0001
Genealogy/blood group (noncodominant systems)	0.011	<0.003	0.078	<0.0001
Genealogy/RFLP	0.023	<0.0001	0.168	<0.0001
Genealogy/RFLP (first 11 RFLP systems)	0.014	<0.0001	0.119	<0.0001
Genealogy/RFLP (second 11 RFLP systems)	0.018	<0.0001	0.131	<0.0001
Genealogy/RFLP + blood group	0.027	<0.0001	0.204	<0.0001

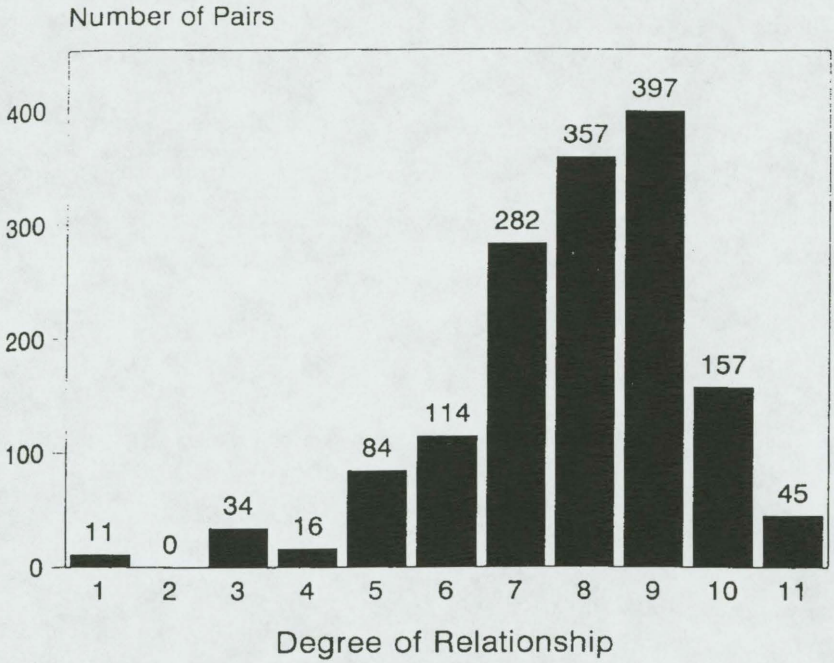

Figure 2. Distribution of pairs of individuals with nonzero kinship coefficients. The *x* axis gives the degree of relationship for each pair (e.g., a value of 1 corresponds to sibling pairs, 2 corresponds to uncle-nephew or double first cousin pairs, 3 corresponds to first cousins, 4 to first cousins once removed, and so on). A small number of individuals were related through multiple paths and thus did not fit into one of these categories; they are not shown here.

nealogies are truncated (like all pedigrees), pairs of individuals who are relatives may still yield a kinship coefficient of zero. It is also possible that some genealogical links may be missing. Accordingly, we performed a second set of analyses in which only the nonzero kinship coefficients were included. The results of this analysis are also shown in Table 5. In every comparison the correlations increased by approximately one order of magnitude. The general patterns, particularly the larger correlations for RFLP data sets than for blood groups, remained the same.

To ensure that our computer algorithm was providing meaningful significance levels, we generated a matrix of random numbers and compared it with the RFLP, blood group, and genealogical kinship matrices. All three correlations were nonsignificant, with p values ranging from 0.07 to 0.41.

Discussion

The analysis of both blood group and RFLP data demonstrates that intersubdivision genetic variance in this population is relatively low. This is consistent with the results of previous studies based on migration matrices, isonymy, and genealogies. These studies yielded F_{ST} values of 0.00003, 0.0007, and 0.0003, respectively (F_{ST} is essentially the same measure as G_{ST}). These results are also consistent with those of a recent study of VNTR (variable number of tandem repeat) variation in the Utah CEPH (Centre d'Étude du Polymorphisme Humain) kindreds, which, although based on a less systematically sampled population, show little evidence of subdivision effects in the Utah population (Chakraborty and Daiger 1991). The G_{ST} values obtained here, particularly those corrected for sampling variance, are well within the ranges seen in other continental populations and are substantially lower than those of small isolated populations (Jorde 1980). This result is expected, given this population's historically large size (which limits genetic drift) and high rates of population mobility and immigration.

Although the G_{ST} values based on gene frequencies are relatively low, they are nonetheless substantially higher than the previously obtained F_{ST} values. This can be attributed to the fact that gene frequencies reflect many generations of past evolution, whereas genealogical kinship data are truncated. In this population few of the genealogies include individuals born before 1800. The migration matrix approach assumes equilibrium migration patterns and uniformity of source populations, and both of these assumptions are often incorrect (Jorde 1982). Isonymy data, like gene frequencies, tend to reflect more of a population's past evolution; it is thus interesting that the isonymy-based F_{ST} value is closest in value to the G_{ST} values obtained here.

The relative lack of between-subdivision heterogeneity helps to account for the low correlation between genetic distances and those based on isonymy, migration matrices, and genealogies. Our previous studies have shown that the last three types of data yield highly significant correlations among one another. This may be partly because these kinship matrices were based on large sample sizes—thousands of individuals in each subdivision. The sample size for the genetic data, approximately 55 per subdivision, is necessarily much smaller.

The especially poor correlation between geographic distance and genetic distance, with negative correlation coefficients for both the blood group and RFLP matrices, can be explained by a nonrandom settlement pattern of Utah by different European population groups. For example, the Scandinavians, who are genetically similar to one another, tended to settle in the geographically removed far-northern and central parts of the state, represented here by Box Elder and Sanpete counties, respectively

(Figure 1) (Bitton and Irving 1976; Mulder 1976; O'Brien, Zenger et al. 1994). These settlement patterns and the relatively few generations for the buildup of an isolation-by-distance effect after settlement contribute to a lack of correlation between genetic and geographic distance. Previous analyses of this population have shown that geographic distance is highly correlated with kinship estimated from migration matrices and pedigrees but less correlated with kinship estimated by isonymy. Again, this can be attributed to nonrandom settlement of Utah by founding groups with different surname distributions (Jorde 1989; Jorde and Morgan 1987). Other studies have shown a similar lack of correlation between geographic distance and isonymy in recently founded populations (Gradie et al. 1988; Smith 1988; Swedlund et al. 1985).

These results can be compared with those of one other study in which genealogical kinship and genetic kinship have been analyzed in a subdivided population. O'Brien (1987) compared genetic distances based on four blood group loci with genealogical kinship in the Hutterite S-leut. In contrast to the present study and despite a smaller number of loci, O'Brien demonstrated a significant correlation between genealogical kinship and genetic distance. This can be ascribed to several factors. First, the number of individuals analyzed was 3171, much larger than the present study. Second, 44 subdivisions were included, providing more data points for statistical analysis. Finally, and perhaps most important, the Hutterite population has a history of genetic isolation (both from the outside world and between subdivisions within the population) and a small founding population. It is thus considerably more structured than the Utah Mormon population.

Despite a lack of significant correlation at the subdivision level, both the blood group and RFLP data showed significant, albeit low, correlations with genealogical kinship at the individual level. As for the subdivision-level correlations, the RFLP data provided higher correlations than did the blood group data. To the extent that genealogical kinship represents known genetic relationships, this result indicates that the RFLP data provide a more accurate portrayal of genetic kinship than do the blood group data.

Although the significance levels of these correlations are encouraging, the low values of the correlation coefficients indicate that the genetic data could not accurately predict genealogical kinship. These results are expected on the basis of previous theoretical analyses. It is well known that goodness-of-fit tests based on Hardy-Weinberg expectations have low power to detect inbreeding (Jenkins et al. 1985; Ward and Sing 1970; Weir 1990). Yasuda (1968) showed that maximum-likelihood estimates of inbreeding coefficients from genotypes become reliable only when the inbreeding coefficient is large (0.02 or more). As the inbreeding coefficient becomes smaller, larger samples of subjects and loci are

required for an accurate estimate. The average inbreeding coefficient (f) in the present sample is 10^{-4}, which is consistent with previous estimates based on much larger samples (Jorde 1989). Yasuda's results, based on an assumption of 22 codominant loci, indicate that inbreeding coefficients in this low range could be statistically distinguished from zero only when over 18,000,000 subjects are assayed genetically. Similarly, Thompson (1976) showed that the relationship between f and heterozygosity measured by gene frequencies is unpredictable, particularly when the number of generations since founding is small (as in this population). Like Yasuda, Thompson indicates that the relationship between f and heterozygosity becomes more stable as more loci are sampled. In this regard, it is instructive that the correlations obtained here continued to increase as more loci were compared with genealogical kinship (e.g., adding the blood group and RFLP information together).

It is also noteworthy that Robertson and Hill (1984) showed that rare alleles provide more reliable estimates of low f values than do common ones. This implies that minisatellite and microsatellite repeat polymorphisms, which usually have large numbers of alleles, may be more useful for predicting f values in human populations.

Why do the RFLP data yield higher correlations with other types of data than do the blood groups? Several reasons can be offered. First, gene frequency estimates are inexact in systems with dominance. This may explain why in some analyses the four codominant blood group systems correlated more highly with the RFLPs than did the entire blood group data set. Second, RFLPs are more likely than blood groups to be selectively neutral. (However, this factor is probably of limited importance in a study such as this, in which there is relatively little time depth.) Finally, RFLPs are a direct reflection of DNA sequence variation at single sites, whereas the relationship between most blood group alleles and DNA sequences is more complex [e.g., Mouro et al. (1993) and Yamamoto et al. (1990)].

There is no question that blood group variation has played an important role in helping to understand human genetic evolution. Because of their abundance and wide distribution, blood group data continue to be useful in many contexts (Cavalli-Sforza et al. 1988; Nei and Roychoudhury 1993; Sokal et al. 1991, 1993). Nevertheless, our results indicate that RFLP data may provide a more accurate and complete picture of human genetic variation and evolution.

Acknowledgments We are grateful for aid and comments from Sharon Austin, Michael Bamshad, John Bird, Mary Hadley, Elaine Hillas, Leslie Jerominski, Rich Kerber, Mark Leppert, Judy Lindblom, Jean Nash, Steve Randall, Joli Schwellenbach, Scott Watkins, Ray White, and Keith Willmore. This re-

search was supported by the National Institutes of Health through grants HG-00347 (awarded to L.B. Jorde) and CA42014 (awarded to the Utah Cancer Center) and by the National Science Foundation through grants BNS-8720330 (awarded to L.B. Jorde) and DBS-9211255 (awarded to A.R. Rogers).

Received 18 November 1993; revision received 8 February 1994.

Literature Cited

Arrington, L.J., and D. Bitton. 1979. *The Mormon Experience: A History of the Latter-Day Saints*. New York: Alfred A. Knopf.

Balazs, I., J. Neuweiler, P. Gunn et al. 1992. Human population genetic studies using hypervariable loci. I. Analysis of Assamese, Australian, Cambodian, Caucasian, Chinese, and Melanesian populations. *Genetics* 131:191–198.

Bitton, D., and G. Irving. 1976. The continental inheritance. In *The Peoples of Utah*, H.Z. Papanikolas, ed. Salt Lake City, UT: Utah State Historical Society, 221–250.

Bowcock, A.M., J.R. Kidd, J.L. Mountain et al. 1991. Drift, admixture, and selection in human evolution: A study with DNA polymorphisms. *Proc. Natl. Acad. Sci. USA* 88:839–843.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36.

Cavalli-Sforza, L.L., A. Piazza, P. Menozzi et al. 1988. Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85:6002–6006.

Chakraborty, R., and S.P. Daiger. 1991. Polymorphisms at VNTR loci suggest homogeneity of the white population of Utah. *Hum. Biol.* 63:571–587.

Deka, R., R. Chakraborty, and R.E. Ferrell. 1991. A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics* 11:83–92.

Edwards, A., H.A. Hammond, L. Jin et al. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12:241–253.

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

Harpending, H., and T. Jenkins. 1973. Genetic distance among southern African populations. In *Methods and Theories of Anthropological Genetics*, M.H. Crawford and P.L. Workman, eds. Albuquerque, NM: University of New Mexico Press, 177–199.

Harpending, H.C., S.T. Sherry, A.R. Rogers et al. 1993. The genetic structure of ancient human populations. *Curr. Anthropol.* 34:483–496.

Horai, S., R. Kondo, Y. Nakagawa-Hattori et al. 1993. Peopling of the Americas, founded by four major lineages of mitochondrial DNA. *Molec. Biol. Evol.* 10:23–47.

Jenkins, T., P. Beighton, and A.G. Steinberg. 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*, v. 1, J.H. Mielke and M.H. Crawford, eds. New York: Plenum Press, 135–208.

Jorde, L.B. 1982. The genetic structure of the Utah Mormons: Migration analysis. *Hum. Biol.* 54:583–597.

Jorde, L.B. 1989. Inbreeding in the Utah Mormons: An evaluation of estimates based on pedigrees, isonymy, and migration matrices. *Ann. Hum. Genet.* 53:339–355.

Jorde, L.B., and K. Morgan. 1987. Genetic structure of the Utah Mormons: Isonymy analysis. *Am. J. Phys. Anthropol.* 72:403–412.

Karlin, S., D. Carmelli, and B. Bonné-Tamir. 1982. Analysis of biochemical data on Jewish populations. III. The application of individual phenotype measurements for population comparisons. *Am. J. Hum. Genet.* 34:50–64.

Kidd, J.R., F.L. Black, K.M. Weiss et al. 1991. Studies of three Amerindian populations using nuclear DNA polymorphisms. *Hum. Biol.* 63:775–794.

Little, R.J.A., and D.B. Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.

Martinson, J.J., R.M. Harding, G. Philippon et al. 1993. Demographic reductions and genetic bottlenecks in humans: Minisatellite allele distributions in Oceania. *Hum. Genet.* 91:445–450.

McLellan, T., L.B. Jorde, and M.H. Skolnick. 1984. Genetic distances between the Utah Mormons and related populations. *Am. J. Hum. Genet.* 36:836–857.

Mouro, I., Y. Colin, B. Cherif-Zahar et al. 1993. Molecular genetic basis of the human Rhesus blood group system. *Nature Genet.* 5:62–65.

Mulder, W. 1976. Scandinavian saga. In *The Peoples of Utah*, H.Z. Papanikolas, ed. Salt Lake City, UT: Utah State Historical Society, 141–185.

Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.

Nei, M., and A.K. Roychoudhury. 1993. Evolutionary relationships of human populations on a global scale. *Molec. Biol. Evol.* 10:927–943.

O'Brien, E. 1987. The correlation between population structure and genetic structure in the Hutterite population. In *Mammalian Dispersal Patterns: The Effects of Social Structure on Population Genetics*, B.D. Chepko-Sade and Z.T. Halpin, eds. Chicago, IL: University of Chicago Press, 193–224.

O'Brien, E., R. Zenger, and L.B. Jorde. 1994. Genetic structure of the Utah Mormons: genetic kinship and ancestral relationships. Unpublished.

O'Brien, E., R.A. Kerber, L.B. Jorde et al. 1994. Founder effect: Assessment of variation in genetic contribution among founders. *Hum. Biol.* 66(2):185–204.

Robertson, A., and W.G. Hill. 1984. Deviations from Hardy-Weinberg proportions: Sampling variances and use in estimation of inbreeding coefficients. *Genetics* 107:703–718.

Smith, M.T. 1988. The Isle of Wight in 1851: Historical migration and genetic structure. *Actes des 3emes Journees Anthropologiques* 24:113–122.

Smouse, P.E., J.C. Long, and R.R. Sokal. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35:627–632.

Sokal, R.R., N.L. Oden, and C. Wilson. 1991. Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351:143–145.

Sokal, R.R., G.M. Jacquez, N.L. Oden et al. 1993. Genetic relationships of European populations reflect their ethnohistorical affinities. *Am. J. Phys. Anthropol.* 91:55–70.

Stoneking, M., L.B. Jorde, K. Bhatia et al. 1990. Geographic variation of human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–733.

Swedlund, A.C., A.B. Anderson, and A.J. Boyce. 1985. Population structure in the Connecticut valley. II. A comparison of multidimensional scaling solutions of migration matrices and isonymy. *Am. J. Phys. Anthropol.* 68:539–547.

Thompson, E.A. 1976. Population correlation and population kinship. *Theor. Popul. Biol.* 10:205–226.

Torroni, A., O. Semino, R. Scozzari et al. 1990. Y chromosome DNA polymorphisms in human populations: Differences between Caucasoids and Africans detected by 49a and 49f probes. *Ann. Hum. Genet.* 54:287–296.

Ward, R.H., and C.F. Sing. 1970. A consideration of the power of the chi-square test to detect inbreeding effects in natural populations. *Am. Natur.* 104:355–365.

Weir, B.S. 1990. *Genetic Data Analysis.* Sunderland, MA: Sinauer Associates.

Workman, P.L., H.C. Harpending, J.M. Lalouel et al. 1973. Population studies on Southwestern Indian tribes. VI. Papago population structure: A comparison of genetic and migration analyses. In *Genetic Structure of Populations*, N.E. Morton, ed. Honolulu, HI: University of Hawaii Press, 166–194.

Yamamoto, F., H. Clausen, T. White et al. 1990. Molecular genetic basis of the histo-blood group ABO system. *Nature* 345:229–233.

Yasuda, N. 1968. Estimation of the inbreeding coefficient from phenotype frequencies by a method of maximum likelihood scoring. *Biometrics* 24:915–935.