

Population Structure and Genetic Heterogeneity in the Upper Markham Valley of New Guinea

JAMES W. WOOD

Department of Anthropology, University of Michigan, Ann Arbor, Michigan 48109

KEY WORDS Population structure · Genetic variation · Wahlund's f · Blood groups · New Guinea

ABSTRACT An analysis is presented of the standardized Wahlund's variances (f) in gene frequencies of the ABO, Rh and MNS blood group systems among 19 villages of the Atsera isolate of the upper Markham Valley, Papua New Guinea. In the past, there has been some disagreement over the relative importance of population structure and natural selection in the determination of these variances. The Lewontin-Krakauer test is presented as a means of resolving this disagreement. According to this test, selectively neutral variation in gene frequencies should generate essentially homogeneous values of f for all loci, a homogeneity which can be tested by comparing the value of s_f^2 to a theoretical σ_f^2 expected when variations in \hat{f} are due solely to sampling error. The observed value of s_f^2 for the Atsera isolate is 2.9×10^{-5} , which is not significantly different from the expected values that range from 1.23×10^{-5} to 2.46×10^{-5} depending on the constant used in calculating σ_f^2 . Therefore it appears that nonselective aspects of population structure such as genetic drift and intervillage migration are responsible for the recorded genetic variation in this isolate.

The genetic studies of the Atsera and Waffa isolates of Papua New Guinea's Upper Markham Valley conducted by Eugene Giles and his associates (Giles et al., '66a,b, '70) occupy an important place in the development of our understanding of human biology in Melanesia. They were among the earliest field investigations in the region to record genetic data on the level of the local breeding population and to provide some of the demographic information necessary for an evolutionary interpretation of those data. By far the most remarkable result of this research has been the discovery of huge stores of genetic variation within these small populations. Gene frequencies often differ significantly from one village to the next within each isolate despite the superficial similarity of village environments and the moderating effects of intervillage migration. According to the χ^2 method of Neel and Schull ('54), the probabilities that villages of the Atsera isolate were drawn from the same genetic universe range from less than 0.05 for the MNS blood group system to less than 0.001 for the ABO and Rh systems

(Giles et al., '70: p. 65). The Waffa population displays comparable heterogeneity (Giles et al., '66b). It has, however, proven far easier to demonstrate the existence of this variation than to explain it. Because of the smallness of these isolates and the lack of obvious environmental differences within them, it is generally believed that random genetic drift, the founder effect, migration or some other "non-selective" aspect of population structure is responsible (Giles et al., '66b, '70; Simmons, '73). But to date the proper test of this belief has not been forthcoming.

One attack on this problem was made by Cadien ('71) who sought to explain the observed variation in the Atsera isolate as a balance between drift and intervillage migration. Cadien used Sewall Wright's island model of migration to generate a set of expected variances among Atsera villages for gene frequencies of the ABO, Rh and MNS systems. The approximate form of this model is

$$\sigma_q^2 = \frac{\bar{q}(1 - \bar{q})}{4N_{em} + 1} \quad (1)$$

where \bar{q} is the mean village gene frequency, σ_q^2

is its expected variance, N_e is the mean effective population size of villages and m is the fraction of each village which is replaced every generation by migrants drawn at random from the whole population (Wright, '31). This is not, however, the sort of migration that characterizes the Atsera isolate. Cadien treated m as if it were simply the mean inter-village migration rate per generation, which is not at all what Wright intended. As the data of Giles ('70) show, there is considerable differential migration and some isolation by distance within the isolate, factors which could substantially increase the expected value of σ_q^2 . Moreover Eq. (1) is applicable only at genetic equilibrium, deviations from which would be very nearly impossible to preclude in such a small population.

A different approach, one not based on the assumption of genetic equilibrium, has been adopted in the present paper. This approach is not concerned with the gene frequency variance *per se*, but rather with "Wahlund's variance" f , which is

$$f = \frac{\sigma_q^2}{\bar{q}(1-\bar{q})}$$

That is, f is equal to the variance in the frequency of a given allele standardized by the mean frequency of that allele as computed over all the villages of the isolate.

In a classic paper, Wahlund ('28) demonstrated that f is the variance in gene frequency contributed by the deviations from panmixis that result when a population is subdivided. Moreover, f (which is equivalent to Wright's F_{ST}) is a measure of the average inbreeding within population subdivisions relative to the population as a whole and of the rate at which genetic variation is lost through random drift (Crow and Kimura, '70). In general, f is determined by all the nonselective elements of population structure, including past and present population size, the distribution of population among subdivisions, migration and the like.

For the purposes of this paper, the most important property of f is the following: If several loci display geographical variation within a subdivided population and if this variation is due to random drift or some other "structural" factor like differential migration, then the value of \hat{f} (the unbiased estimate of f) should be approximately the same for each locus. Population structure affects all loci indiscriminately. The homogeneity or heterogeneity of \hat{f} values therefore provides a test of the

selective neutrality of polymorphic loci. If there is significant heterogeneity among \hat{f} values, at least some of the loci are likely to be subject to natural selection. (However, this test will not reveal which loci are undergoing selection, nor whether high \hat{f} values represent diversifying selection or low \hat{f} values represent stabilizing selection.) If, on the other hand, the \hat{f} values are reasonably homogeneous, it can be inferred that the observed variation is not due to natural selection. This test has already been used by Cavalli-Sforza ('66) and by Lewontin and Krakauer ('73) to analyze geographical variation in the distribution of gene frequencies.

MATERIALS AND METHODS

This approach can easily be applied to the problem of genetic variation in the Upper Markham Valley. Table 1 presents the appropriate data for the ABO, Rh and MNS blood group systems gathered from 19 villages of the Atsera language group (Giles et al., '66a, '70). The Atsera isolate as a whole can be considered a single breeding population, while the villages represent its most important subdivisions. In certain instances, villages which have high rates of intermarriage and similar gene frequencies have been pooled, so that a total of only 15 subdivisions is recognized. This pooling of similar villages will tend to inflate our estimates of f , but it will affect all loci proportionately and hence will not change the results of the analysis.

It should be noted that a fair amount of intervillage migration occurs each generation within the Atsera isolate (Giles, '70). Although Wahlund's original formulation of f was based on the assumption that the population subdivisions were entirely isolated from each other, more recent work has shown that intra-population migration, like all other aspects of population structure, is expected to influence all loci equally (Bodmer and Cavalli-Sforza, '68). This, although migration will tend to dampen the \hat{f} values, it will not in theory affect their homogeneity or heterogeneity.

Before the homogeneity of \hat{f} values can be tested statistically, we must deal with the fact that the gene frequencies in tri-allelic systems like the ABO, Rh and MNS are correlated and that their corresponding \hat{f} values are therefore not independent. I have compensated for this problem in two ways: (1) by subtracting one degree of freedom for each tri-al-

TABLE 1

Gene frequencies of the ABO, Rh and MNS blood group systems among villages of the Atsera isolate, Morobe District, Papua New Guinea¹

Village	Sample size	A _i	B	O	R _i (cDe)	R ₂ (cDE)	R ₀ (cDe)	Ms	Ns	NS
A. Awan	103	0.3032	0.2567	0.4396	0.9951	0.0049	0	0.0583	0.8020	0.1398
B. Bampa-Antir-Siats ²	235	0.1963	0.1600	0.6437	0.9745	0.0255	0	0.0340	0.8511	0.1149
C. Gnarowein	111	0.2226	0.2111	0.5662	0.9775	0.0225	0	0.0856	0.8253	0.0891
D. Guruf	130	0.2398	0.2047	0.5551	0.9423	0.0538	0.0038	0.0577	0.7711	0.1712
E. Intoap	87	0.1782	0.2597	0.5619	0.9425	0.0575	0	0.0575	0.8792	0.0633
F. Itsingants	61	0.1411	0.2318	0.6271	0.9180	0.0820	0	0.1066	0.7843	0.1092
G. Kaiapit	211	0.2346	0.2471	0.5183	0.8981	0.0718	0.0301	0.0521	0.8500	0.0978
H. Onga-Naruboin ²	227	0.2220	0.1998	0.5782	0.9758	0.0220	0.0022	0.0793	0.8267	0.0940
I. Puguap	95	0.2172	0.1780	0.6048	0.9632	0.0316	0.0053	0.0947	0.7650	0.1402
J. Singas	91	0.3418	0.1928	0.4653	0.9670	0.0330	0	0.0330	0.8259	0.1411
K. Sukurum-Dumlinan ²	164	0.2222	0.0857	0.6921	0.9421	0.0427	0.0152	0.0518	0.8530	0.0951
L. Wankum	160	0.2028	0.3362	0.4610	0.9469	0.0500	0.0031	0.0344	0.8067	0.1589
M. Wompul	97	0.1105	0.2503	0.6391	0.9536	0.0258	0.0206	0.0928	0.7768	0.1304
N. Yanuf	45	0.1727	0.1864	0.6406	0.9444	0.0556	0	0.0222	0.8695	0.1083
O. Yatsing	68	0.2172	0.3077	0.4749	0.9412	0.0515	0.0074	0.0882	0.7408	0.1709

¹ From Giles et al. ('66a, '70).

² Pooled data from two or more villages.

lelic system in the following statistical tests, and (2) by arbitrarily excluding the least common allele or haplotype of each system from the analysis. Thus the \hat{f} values for the B allele the the R₀ and Ms haplotypes do not enter into the following computations. Because of the manner in which \hat{f} is standardized, its value is independent of gene frequency, and this correction will in no way bias the results of the analysis.

How homogeneous must the \hat{f} values be before they can be attributed to selectively neutral variation? The precise answer to this question requires an explicit expression for $\sigma_{\hat{f}}^2$, the theoretical sampling variance of \hat{f} . If the observed variance $s_{\hat{f}}^2$ is significantly larger than $\sigma_{\hat{f}}^2$, the \hat{f} values cannot be considered homogeneous. Lewontin and Krakauer ('73) have attempted to derive $\sigma_{\hat{f}}^2$ as follows: Assume that \bar{q} and \bar{f} (the mean value of \hat{f}) have been estimated from a random sample taken across n population subdivisions. Also assume for the moment that the underlying distribution of q is normal, that \bar{f} is small and that n is so large that \bar{q} is very close to the true mean, μ_q , and consequently there is only negligible sampling variance in the denominator of \hat{f} . Given these conditions, \hat{f} has an expected variance over repeated samples of

$$\sigma_{\hat{f}}^2 = \frac{2\bar{f}^2}{n-1}, \tag{2}$$

the standard equation for the sampling variance of the variance of a normal variate based on n independent observations. Obviously the assumptions behind this equation—normal

distribution of q , small \bar{f} , large n —are unlikely to all be true in most empirical applications. But Lewontin and Krakauer were unable to derive expressions analytically for cases in which these assumptions are relaxed. Instead they resorted to a large series of Monte Carlo simulations of the behavior of neutral alleles with various distributions of q and several values of n (Lewontin and Krakauer, '73: table 2). The simulations yielded the following results:

(1) As \bar{f} increases, the general expression $\sigma_{\hat{f}}^2 = \kappa\bar{f}^2/(n-1)$ still holds true, with κ decreasing as \bar{f} increases. However $\kappa = 2$ is approximately correct for values of \bar{f} in the range 0 to 0.05, a range which includes the value of \bar{f} considered in the present paper. The value of κ appears to be independent of n .

(2) The expression $\sigma_{\hat{f}}^2 = 2\bar{f}^2/(n-1)$ applies to both symmetrical and asymmetrical binomial distributions of q as well as the normal distribution of q .

(3) Uniform and U-shaped distributions of q require smaller values of κ , the former calling for $\kappa \approx 1.0$ and the latter for $\kappa \approx 0.5$. However the simulations using these distributions also generated higher values of \bar{f} , thus confounding the effects of distribution and large \bar{f} on κ .

(4) The denominator of \hat{f} , namely $\bar{q}(1-\bar{q})$, appears to have very little sampling variance compared to the sampling variance of the numerator, $s_{\hat{f}}^2$. Thus the expression $\kappa\bar{f}^2/(n-1)$ applies even for small values of n .

The expression for $\sigma_{\hat{f}}^2$ depends in large part

on the underlying distribution of q among population subdivisions. This distribution is difficult to infer for the Atsera isolate with an n of only 15. But inspection of table 1 suggests that the distributions of the nine alleles and haplotypes in question are as a rule unimodal and convex, more dispersed than the binomial but considerably less dispersed than the uniform. (The only exception to this generalization is the R_O haplotype which takes on a strongly asymmetrical J-shaped distribution. But this haplotype has already been excluded from the analysis.) Therefore κ equal to one and two would appear to set the lower and upper limits respectively of $\sigma_{\hat{f}}$, and both values have been used in the following test.

Nei and Maruyama ('75) suggested that certain forms of migration and mutation may demand values of κ greater than two. For example, they carried out their own Monte Carlo simulations for neutral alleles in circular steppingstone models of migration and no mutation, and computed average values of κ ranging from 4.42 to 6.14 depending on the number of populations subdivisions in the models. A second set of simulations was conducted with an island model of ten colonies of size 10 and mutation rates varying from 0.001 to 0.1. These simulations also required a κ greater than two (although Nei and Maruyama, '75, failed to indicate the precise value of κ needed for these experiments). Thus the use of κ equal to one or two represents a conservative test of selective neutrality.

Robertson ('75a,b) has argued that Eq. (2) may seriously underestimate the sampling variance of \hat{f} in hierarchically structured populations. Hierarchical structure refers to the organization of population subdivisions into clusters within which, due to common history or migration, genetic relatedness is significantly greater than the relatedness between clusters. The structure of relatedness can be described in terms of a matrix of genetic correlation coefficients between pairs of subdivisions. The correlation between the i^{th} and j^{th} subdivisions, r_{ij} , can be estimated as $\text{cov}(i,j)/\hat{\sigma}_{qi}\hat{\sigma}_{qj}$ where $\hat{\sigma}_{qi}$ is the observed standard deviation in gene frequencies across the m loci studied in the i^{th} subdivision, and

$$\text{cov}(i,j) = \sum_x (q_{ix} - \bar{q}_x)(q_{jx} - \bar{q}_x)/\bar{q}_x(1 - \bar{q}_x)/m,$$

q_{ix} being the frequency of allele x in the i^{th} subdivision.

Robertson ('75b) reasoned as follows in deriving an expression for the sampling variance

of \hat{f} in a hierarchically structured population. Assume that the distribution of gene frequencies across a set of hierarchically structured population subdivisions can be described by a normal variate $q \sim N(\mu_q, \sigma_q^2)$. Since r_{ij} is the genetic correlation between subdivisions i and j , the expected mean squared difference between them is $2\sigma_q^2(1 - r_{ij})$, and the total average mean squared difference is $2\sigma_q^2(1 - \bar{r})$ where \bar{r} is the mean correlation between paired subdivisions. Thus, if a random sample is taken across all the subdivisions, the effects of hierarchical structure are such that the expected inter-subdivision variance, V_q , as measured from the sample is $\sigma_q^2(1 - \bar{r})$.

Suppose now that several samples are drawn from this set of population subdivisions. The estimates of V_q will themselves vary from sample to sample with variance equal to $V(\hat{V}_q)$. Define q_i as the gene frequency in the i^{th} subdivision and write $\Delta_i = q_i - \bar{q}$. The value of \hat{V}_q in a single sample of n subdivisions can then be computed as $[\sum \Delta_i^2 - (\sum \Delta_i)^2/n]/(n - 1)$. Since the expectation of this variance is $\sigma_q^2(1 - \bar{r})$, the expected variance over all samples (assuming all have the same structure) is

$$E(V_q) = E[\sum \Delta_i^2 - (\sum \Delta_i)^2/n]/(n - 1)^2 - [\sigma_q^2(1 - \bar{r})]^2 = E[(\sum \Delta_i^2)^2 - (2/n)(\sum \Delta_i^2)(\sum \Delta_i)^2 + (1/n^2)(\sum \Delta_i)^4]/(n - 1)^2 - \sigma_q^4(1 - \bar{r})^2.$$

It can be shown that

$$E(\sum \Delta_i^2)^2 = \sigma_q^4 [n(n + 2) + 2 \sum_{i \neq j} r_{ij}^2],$$

$$E[\sum \Delta_i^2 (\sum \Delta_i)^2] = \sigma_q^4 [n(n + 2) + n(n - 1)(n + 4) + \bar{r} + 2 \sum_{i \neq j} r_{ij}^2], \text{ and}$$

$$E(\sum \Delta_i)^4 = \sigma_q^4 [3n^2 + 6n^2(n - 1)\bar{r} + 3n^2(n - 1)\bar{r}^2].$$

Define $\delta_{ij} = (r_{ij} - \bar{r})$. Then

$$\sum_{i \neq j} r_{ij}^2 = n(n - 1)\bar{r}^2 + \sum \delta_{ij}^2 \text{ and}$$

$$\sum_i (\sum_{j \neq i} r_{ij})^2 = n(n - 1)\bar{r}^2 + \sum_i (\sum_{j \neq i} \delta_{ij})^2.$$

The expression for the sampling variance of \hat{V}_q then reduces to

$$V(\hat{V}_q) = 2\sigma_q^4 \left[(1 - \bar{r})^2/(n - 1) + \frac{\sum \delta_{ij}^2 - 2 \sum_i (\sum_{j \neq i} \delta_{ij})^2/n}{(n - 1)^2} \right] \quad (3)$$

According to Lewontin and Krakauer's simulations, the sampling variance of $\bar{q}(1 - \bar{q})$ is very small relative to $V(\hat{V}_q)$. Therefore $\sigma_{\hat{f}}^2 \approx \kappa V(\hat{V}_q)$.

Consider now a set of "internal" correlations, ρ_{ij} , related to r_{ij} but adjusted to have mean zero:

$$\rho_{ij} = (r_{ij} - \bar{r})/(1 - \bar{r}).$$

It follows that $\sum \delta_{ij}^2 = n(n - 1)(1 - \bar{r})^2 \sigma_{\rho}^2$ where σ_{ρ}^2 is the variance of the internal correlations. Substituting in Eq. (3) and remember-

ing that $E(\hat{V}_q) = \sigma_q^2(1 - \bar{r})$ and $\sigma_f^2 \cong \kappa V(\hat{V}_q)$, it is approximately the case that

$$\sigma_{\hat{f}}^2 = \frac{\kappa \bar{f}^2}{(n-1)} + \kappa \bar{f}^2 \sigma_\rho^2 \tag{4}$$

Stated differently, the effect of hierarchical structure is to inflate the estimated sampling variance of \hat{f} by a term proportional to the variance in the internal genetic correlations between population subdivisions. Relaxing the assumption of normality, the behavior of κ should be approximately the same as in Lewontin and Krakauer's simulations, and κ of one and two should still provide lower and upper limits for a conservative expectation of $\sigma_{\hat{f}}^2$.

Strictly, Robertson's correction is applicable only when the samples for each locus are independent. In the present case, the same individuals have been sampled for all three loci. Under some forms of migration, this sampling scheme could introduce a covariance between loci, measurable as $\text{cov}(a, b) = \sum_i (q_{ia} - \bar{q}_a)(q_{ib} - \bar{q}_b)/(n - 1)$, which would invalidate the assumption of independently sampled loci. The data in table 1 give the following covariances:

$$\begin{aligned} \text{cov(ABO, Rh)} &= -0.0004 \pm 0.0077 \\ \text{cov(ABO, MNS)} &= 0.0005 \pm 0.0086 \\ \text{cov(Rh, MNS)} &= -0.0001 \pm 0.0039. \end{aligned}$$

Fortunately none of these values differs significantly from zero, and the application of Eq. (4) to the Atsera data appears to be justified.

It is not immediately obvious that Robertson's correction of Lewontin and Krakauer's

original variance has much relevance to a population like the Atsera with small, closely related villages, substantial intervillage migration and thus little opportunity for the development of hierarchical population structure. In fact the data on the Atsera suggest the absence of significant hierarchical structure. Table 2, which presents the matrix of ρ_{ij} values between Atsera villages for the ABO, Rh and MNS loci, clearly shows that there is little variation around a mean internal correlation of zero. The estimated value of σ_ρ^2 is equal to only 3.4×10^{-3} with 95% confidence limits of 2.6×10^{-3} to 4.5×10^{-3} . Thus, although Robertson's correction has been used in the following analysis, it produces an expected value of $\sigma_{\hat{f}}^2$ that is only slightly different from the one given by Lewontin and Krakauer's equation.

RESULTS

In table 3, the weighted means and variances of the relevant village gene frequencies are shown, along with the estimates of f and their 95% confidence limits. While the observed values of \hat{f} are not strictly identical, they are remarkably similar. The value of \hat{f} from table 3 is 0.0128 and the estimated variance $s_{\hat{f}}^2$ is 2.9×10^{-5} . The latter figure does not differ significantly from either of the two theoretically expected variances:

$$\text{with } \kappa = 1, \sigma_{\hat{f}}^2 = 1.23 \times 10^{-5} (F = 2.36, 0.05 < P < 0.10),$$

$$\text{with } \kappa = 2, \sigma_{\hat{f}}^2 = 2.46 \times 10^{-5} (F = 1.18, 0.25 < P < 0.50).$$

Thus the observed \hat{f} values appear to be homo-

TABLE 2

Pairwise internal genetic correlations (ρ_{ij}) between villages of the Atsera isolate based on the A and O alleles and the R₁, R₂ Ns and NS haplotypes^{1,2}

	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	-0.084	-0.005	0.005	-0.117	-0.168	-0.022	-0.017	-0.050	0.061	0.151	0.028	-0.173	-0.123	0.028
B		0.050	0.045	0.039	0.045	0.017	0.056	0.050	-0.084	0.056	0.011	0.039	0.056	-0.005
C			0.061	0.050	0.017	0.050	0.072	0.050	-0.005	0.028	0.028	0	0.034	0.028
D				0.017	0.006	0.017	0.061	0.061	-0.011	0.006	0.039	0.017	0.017	0.050
E					0.028	0.056	0.050	0.011	-0.056	0.022	0.011	-0.005	0.050	-0.011
F						-0.017	0.022	0.028	-0.173	0.039	-0.056	0.061	0.056	-0.045
G							0.045	-0.005	-0.005	0	0.028	-0.050	0.022	0.006
H								0.056	-0.011	0.028	0.028	0.011	0.039	0.028
I									-0.056	0.028	-0.005	0.039	0.028	0.022
J										-0.134	0.011	-0.179	-0.117	0.006
K											-0.078	0.022	0.056	-0.073
L												-0.050	-0.022	0.061
M													0.039	-0.034
N														-0.034

¹ Data from table 1.

² Villages listed alphabetically as in table 1.

TABLE 3

Weighted mean frequency, variance and Wahlund's \hat{f} (with 95% confidence limits for \hat{f}) among Atsera villages for genes of the ABO, Rh and MNS blood group systems¹

Gene	\bar{q}	σ_q^2	\hat{f}	95% c.l. ²
A ₁	0.218	0.0022	0.0129	0.0069-0.0321
O	0.567	0.0054	0.0220	0.0118-0.0547
R ₁	0.953	0.0007	0.0156	0.0083-0.0388
R ₂	0.040	0.0004	0.0104	0.0056-0.0259
Ns	0.821	0.0011	0.0075	0.0040-0.0186
NS	0.119	0.0009	0.0086	0.0046-0.0215

¹ Data from table 1.

² Based on χ^2/df distribution with $df = 14$.

geneous, and it can be concluded that the gene frequency variations found in the Atsera isolate are indeed a reflection of population structure alone and not of natural selection.

It is important to consider possible biases in this test. Spuriously low s_i^2/σ_i^2 ratios could be caused by linkage disequilibrium (which would establish correlations among loci) or by similar selection operating on all loci. The former is not a problem in the present case: MNS is unmapped, but the Rh locus is on chromosome 1 whereas the ABO locus is on chromosome 9. The latter bias is more difficult to dismiss on empirical grounds but appears to be unlikely for theoretical reasons. The selection coefficients of the genotypes of all three blood group systems would have to be more or less identical for selection to produce homogeneous \hat{f} values. It is probable that the selection coefficient of each genotype is determined by its antigenic specificity, which varies from one genotype to another not only within each system but *a fortiori* between systems as well.

It is difficult to imagine any selective pressure operating in the same way on all three systems.

Spuriously high s_i^2/σ_i^2 ratios could be produced by several demographic factors even in the absence of selection. For example, the founder effect could increase the heterogeneity of \hat{f} values. If a village with Δ_i^2 very large for one system (say, locus *a*) but small for the other two were to expand and produce several new villages, the \hat{f} for locus *a* would be greater than that for the other loci. A similar effect could be produced by large, uncorrelated fluctuations in the population size of villages and by the selective migration of genotypes. But any attempt to correct for these processes would only reduce the ratio of s_i^2 to σ_i^2 and consequently would not change our conclusions about the homogeneity of the observed \hat{f} values.

DISCUSSION

I believe this is as conclusive a demonstration as possible, given the data and methods at hand, that natural selection is not acting to maintain the geographical variation in gene frequencies of the ABO, Rh and MNS systems that has been observed in the Upper Markham Valley. This should not be interpreted to mean that natural selection does not affect these loci at all. On the contrary, a considerable amount of selection is known to operate on the ABO and Rh loci in the form of maternal-fetal incompatibility, and it may be that selection has affected the mean gene frequencies of all three systems, in the Upper Markham region. Selection does not, however, appear to explain microgeographical variation between villages.

TABLE 4

Population size of selected Atsera villages and average heterozygosity for genes of the ABO, Rh and MNS blood group systems¹

Village	Total pop.	Average heterozygosity ²										
		A ₁	B	O	R ₁	R ₂	R ₀	Ms	Ns	NS	\bar{H}_i	θ_i^3
1. Bampa + ⁴	352	0.3155	0.2688	0.4587	0.0467	0.0467	0	0.0657	0.2534	0.2034	0.1843	0.5913
2. Gnarowein	186	0.3461	0.3331	0.4912	0.0440	0.0440	0	0.1565	0.2884	0.1623	0.2073	0.6355
3. Guruf	185	0.3646	0.3256	0.4939	0.1087	0.1018	0.0076	0.1087	0.3530	0.2838	0.2386	0.6978
4. Kaiaput	440	0.3591	0.3721	0.4993	0.1830	0.1333	0.0584	0.0988	0.2550	0.1765	0.2373	0.6952
5. Onga + ⁴	437	0.3454	0.3197	0.4878	0.0472	0.0430	0.0044	0.1460	0.2865	0.1703	0.2056	0.6322
6. Sukurum + ⁴	254	0.3456	0.1567	0.4262	0.1091	0.0818	0.0299	0.0982	0.2508	0.1721	0.1856	0.5938
7. Wankum	344	0.3233	0.4463	0.4970	0.1006	0.0950	0.0062	0.0664	0.3119	0.2673	0.2349	0.6903
8. Wompul	126	0.1966	0.3753	0.4613	0.0885	0.0503	0.0404	0.1684	0.3468	0.2268	0.2172	0.6549

¹ Data from table 1.

² $H_{ij} = 2q_{ij}(1 - q_{ij})$.

³ $\theta_i = -1/(\ln \bar{H}_i)$.

⁴ Pooled data from two or more villages.

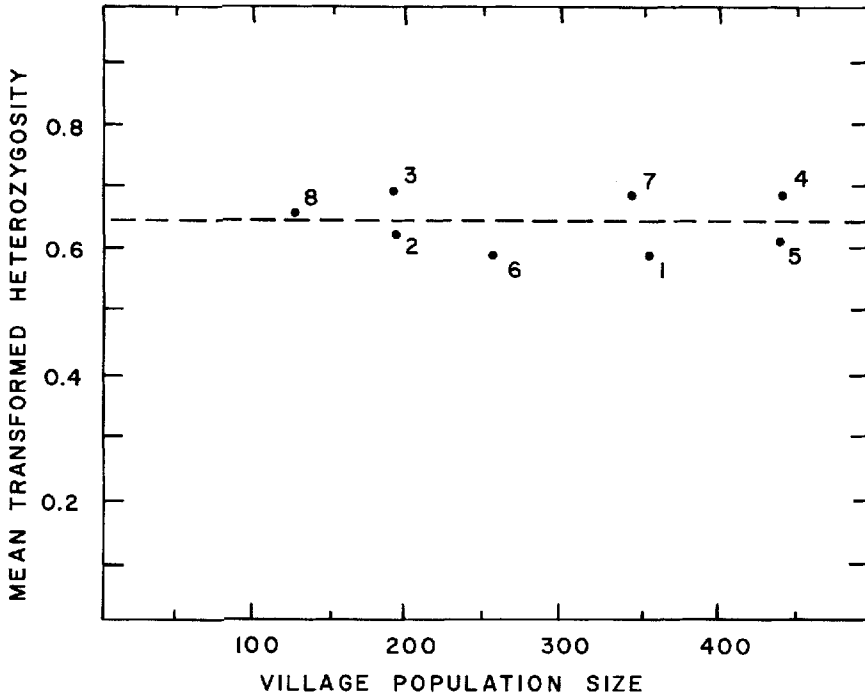


Fig. 1 Relationship between population size, N_i , and mean transformed heterozygosity, $\theta_i = -1/(\ln \bar{H}_i)$, for eight villages of the Atsera isolate based on genes of the ABO, Rh and MNS blood group systems. Single-digit numbers refer to villages as listed in table 4. Broken line represents the grand mean transformed heterozygosity, $\bar{\theta} = 0.649$.

But the evidence against selection adduced herein is purely circumstantial, and perhaps all that can be concluded is that we need not invoke the hypothesis of selection in order to account for the facts of the case. Giles et al. ('66a) were in any event correct in believing that these three blood group systems display a level of heterogeneity in the Atsera isolate that is consistent with the population structure of the isolate.

Unfortunately this analysis cannot reveal what aspects of population structure are important in determining this heterogeneity. It is unlikely, however, that random genetic drift in the strict sense can, by itself, account for the observed variations, since there is no apparent increase in average heterozygosity with village population size. The average heterozygosity of the x^{th} allele in the i^{th} village can be estimated (assuming random mating within villages) as $H_{ix} = 2q_{ix}(1 - q_{ix})$. Table 4 presents the mean village heterozygosity, \bar{H}_i , averaged over the ABO, Rh and MNS loci from eight Atsera villages from which data on population size are available. In theory, there

should be an approximately linear increase in $\theta_i = -1/(\ln \bar{H}_i)$ with N_i , the size of the i^{th} village, if drift alone is operating. This follows from the equation for average heterozygosity at equilibrium between the loss of variation due to drift and its gain due to mutation:

$$\bar{H}_i = 1 - \frac{1}{4\alpha N_i \mu + 1}$$

where α is a constant of proportionality that converts N_i to N_{ei} , the effective size of village i , and μ is the rate per locus at which selectively neutral alleles are produced by mutation (Kimura, '68). This equation really applies only when the number of allelic states is large and the population is in genetic equilibrium, but it suffices to describe a general relationship between \bar{H}_i and N_i . To an approximation,

$$\begin{aligned} \bar{H}_i &= e^{-1/(4\alpha N_i \mu + 1)}, \\ \ln \bar{H}_i &= -1/(4\alpha N_i \mu + 1) \text{ and} \\ \theta_i &= 1 + \xi N_i \end{aligned}$$

where $\xi = 4\alpha\mu$. In fact, as figure 1 clearly shows, there is no systematic increase in θ_i with N_i . Rather, all the transformed village heterozygosities are closely clustered around

the grand mean transformed heterozygosity, $\theta = 0.649$. (This parallels the finding of Wiesenfeld and Gajdusek, '76, that average heterozygosity and population size are uncorrelated among several neighboring isolates in New Guinea's Eastern Highlands.) Additional factors must be operating to obscure the patterns of differentiation that would result from drift alone.

Doubtless the balance between intra-isolate migration and drift is of some importance in the Upper Markham Valley, and Cadien's analysis is at least useful in drawing attention to this process. A more pertinent study might be undertaken using the pairwise intervillage migration rates provided by Giles ('70) and subjecting them to a migration matrix projection of the sort developed by Bodmer and Cavalli-Sforza ('68) or by Imaizumi et al. ('71). But these methods require knowledge of N_{ei} , which can be only crudely estimated from the available data on the Atsera. In addition the migration matrix approach requires information on migration into the isolate from the outside world, information which has not to date been published for the Upper Markham Valley. In short, further research on this subject is contingent on more comprehensive demographic field data being made available by the original investigators.

To end on a note of caution, the fact that selective forces do not seem to be of any appreciable significance in the differentiation of a set of small, closely related villages living under essentially identical ecological conditions tells us nothing whatsoever about the larger issue of the relative importance of selection and drift in the maintenance of polymorphism on a global scale. The methodology used in this paper may have wider applicability, but the specific conclusions apply only to a single, minuscule part of the world.

LITERATURE CITED

- Bodmer, W. F., and L. L. Cavalli-Sforza 1968 A migration matrix model for the study of random genetic drift. *Genetics*, *59*: 565-592.
- Cadien, J. D. 1971 A note on genetic drift in New Guinea. *Hum. Biol. Oceania*, *1*: 140-143.
- Cavalli-Sforza, L. L. 1966 Population structure in human evolution. *Proc. Roy. Soc. London (Ser. B)*, *164*: 362-379.
- Crow, J. F., and M. Kimura 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- Giles, E. 1970 Culture and genetics. In: *Current Directions in Anthropology*. A. Fischer, ed. Am. Anthropol. Assoc., Washington.
- Giles, E., E. Ogan, R. J. Walsh and M. A. Bradley 1966a Blood group genetics of natives of the Morobe District and Bougainville, Territory of New Guinea. *Arch. and Phys. Anthropol. Oceania*, *1*: 135-154.
- Giles, E., R. J. Walsh and M. A. Bradley 1966b Micro-evolution in New Guinea: the role of genetic drift. *Ann. N. Y. Acad. Sci.*, *134*: 655-665.
- Giles, E., S. Wyber and R. J. Walsh 1970 Micro-evolution in New Guinea: additional evidence for genetic drift. *Arch. and Phys. Anthropol. Oceania*, *5*: 60-72.
- Imaizumi, Y., N. E. Morton and D. E. Harris 1971 Isolation by distance in artificial populations. *Genetics*, *66*: 569-582.
- Kimura, M. 1968 Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.*, *11*: 247-269.
- Lewontin, R. C., and J. Krakauer 1973 Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, *74*: 175-195.
- Neel, J. V., and W. J. Schull 1954 *Human Heredity*. Univ. of Chicago Press, Chicago.
- Nei, M., and T. Maruyama 1975 Lewontin-Krakauer test for neutral genes. *Genetics*, *80*: 395.
- Robertson, A. 1975a Remarks on the Lewontin-Krakauer test. *Genetics*, *80*: 396.
- 1975b Gene frequency distributions as a test of selective neutrality. *Genetics*, *81*: 775-785.
- Simmons, R. T. 1973 Blood group genetic patterns and heterogeneity in New Guinea. *Hum. Biol. Oceania*, *2*: 63-71.
- Wahlund, S. 1928 Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas*, *11*: 65-106.
- Wiesenfeld, S. L., and D. C. Gajdusek 1976 Genetic structure and heterozygosity in the kuru region, Eastern Highlands of New Guinea. *Am. J. Phys. Anthropol.*, *45*: 177-190.
- Wright, S. 1931 Evolution in Mendelian populations. *Genetics*, *16*: 97-159.