MULTIVARIATE ANALYSIS OF GAMETIC DISEQUILIBRIUM IN THE YANOMAMA[1]

PETER E. SMOUSE AND JAMES V. NEEL

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109

Manuscript received July 2, 1976
Revised copy received November 29, 1976

ABSTRACT

The gametic disequilibria between all possible pairs of loci were examined for a set of eight codominant loci in each of fifty Yanomama villages, using a multivariate correlation analysis which reduces the results to a single measure of departure from multiple-locus-gametic equilibrium. Thirty-two of the fifty villages departed significantly from multiple-locus gametic equilibrium. The largest contributions to the departure from multiple-locus equilibrium were due to the disequilibria between *MN* and *Ss* and between *Rh(Cc)* and *Rh(Ee)*, indicating the effects of tight linkage. After removing the effects of these obvious sources of disequilibrium, sixteen of the fifty villages still remained significantly out of equilibrium. The disequilibrium between any particular pair of loci was highly erratic from village to village, and (with the exception of the *MN-Ss* and *Cc-Ee* disequilibria) averaged out very close to zero overall, suggesting a lack of systematic forces (epistatic selection). The departure from equilibrium in any one village is in excess of that expected from random sampling alone, and is attributed primarily to the fission-fusion mode of village formation operative in the Yanomama and the fact that a single village consists of a few extended lineages. Village allele frequencies are highly correlated across loci, and most of the non-independence is accounted for by large correlations in the average allelic frequencies of different loci for related villages. It is suggested that these correlations also are due to territorial expansion and population growth. For the tribe as a whole, all but the tightly linked markers of the *MNSs* and *Rh* complexes are approximately uncorrelated, and large departures from multiple-locus Hardy-Weinberg expectation are primarily due to substantial Wahlund variance within the tribe. There is no need to postulate a role for selection in these disequilibria.

THE genetic structure of any human population should reflect the demographic forces which operate on that population. For several years, members of our group have been involved in the study of the Yanomama and other Indian tribes of South America. The objective of these multidisciplinary studies has been to elucidate the genetic organization of relatively undisturbed tribal populations engaged in hunting, gathering, and early agriculture. Most of these tribal populations are highly fragmented into small demes (villages), each consisting primarily of a few extended lineages(NEEL 1967). Our major objective has been to

[1] Supported by National Science Foundation Grant BMS-74-11823.

delineate the magnitude of interdemic microdifferentiation, as measured by single-locus genetic indicators such as WRIGHT's (1943 *et seq.*) F_{ST} measure (NEEL and WARD 1972) and various statistical analogues (SPIELMAN, NEEL and LI 1977), and by multivariate analysis of anthropometric (SPIELMAN *et al.* 1972; SPIELMAN 1973b) dermatoglyphic (ROTHHAMMER *et al.* 1973), and dental (BREWER-CARIAS, LE BLANC and NEEL 1976) criteria. The amount of microdifferentiation encountered has been considerable; the various types of characters (SPIELMAN 1973a; SPIELMAN, MIGLIAZZA and NEEL 1974) and different sorts of measures (NEEL, ROTHMAMMER and LINGOES 1974; SPIELMAN, NEEL and LI 1977) have consistently yielded comparable magnitudes and patterns for this variation. Our initial expectation was that the very dispersive forces leading to marked microdifferentiation should also lead to considerable departures from Hardy-Weinberg equilibrium within single villages. As has been forcefully pointed out by NEEL and WARD (1972), however, this is not at all what one observes. It appears that mating is sufficiently random within any one village that most villages are reasonably close to the Hardy-Weinberg equilibrium. Inasmuch as a single generation of random mating is sufficient to restore single-locus H-W equilibrium, it appears that a test of departure from this condition is not a very sensitive indicator of the internal genetic disruption of these highly stochastic gene pools.

The present paper is an attempt to construct a more sensitive gauge of the internal disruption of these demes. The sort of fission-fusion dynamics mentioned above should lead to gametic disequilibrium within single villages, due to mixture of differentiated gene pools (NEI and LI 1973). Even under random mating within villages, the decay of this disequilibrium occurs only asymptotically. With periodic disruptions, one might expect to find individual villages in a semi-continuous state of gametic disarray, even while single-locus frequencies conform fairly closely to Hardy-Weinberg expectations, as they are known to do (NEEL and WARD 1972).

The objectives of this paper are threefold. We wish: (1) to determine the degree of disequilibrium within the Yanomama, a large unacculturated tribe of southern Venezuela and northern Brazil, a people very little disturbed since pre-Columbian times; (2) to partition this disequilibrium among various components (within villages, among villages, etc.), so as to separate the effects of subdivision *per se* from those due to demographic forces within single villages; and (3) to gauge the extent to which the within-village disequilibrium for any pair of loci is consistent from village to village, in order to determine whether there are any directional forces operative.

STATISTICAL METHODS

The procedures we shall use are somewhat unusual, and a few words about general strategy are in order at the outset. The usual procedure is to test the disequilibrium for each pair of loci in each population, usually in the hope of uncovering some indication of selection. For the 50 villages and 8 loci reported here, this represents $[50(8)(7) \div 2 = 1400]$ separate test criteria, no more than 50 of which are independent. Our objective here is to use disequilibrium to gauge the internal disruption of the village gene pools, and we are only secondarily concerned with

any specific pair of loci. By reducing the whole problem to a smaller set of test criteria, we hope to obtain a measure of the magnitude of the demographic forces operative on these populations, which forces are not locus-specific.

We find it convenient to collate the notation at the outset. All but the more obvious terms are listed in Table 1.

Correlation structure

Consider a locus with two codominant alleles (A_1 and A_2). We define a variable Y_1, which takes the values (1, ½, 0) for the genotypes (A_1A_1, A_1A_2, A_2A_2). A second locus with codominant alleles (B_1 and B_2) yields a second variable Y_2, which takes the values (1, ½, 0) for the genotypes (B_1B_1, B_1B_2, B_2B_2). A similar variable can be defined for each codominant locus. The genotype of an individual can thus be represented by a vector $\mathbf{Y}' = (Y_1, \ldots, Y_H)$; for example, the genotype (A_1A_1, B_2B_2, C_1C_2, D_1D_1) would yield the vector $\mathbf{Y}' = (1, 0, ½, 1)$.

If there are J_i individuals sampled from the i^{th} village, we can compute a mean vector $\bar{\mathbf{Y}}_i$ and a covariance matrix $\mathbf{S}_i$

$$\bar{\mathbf{Y}}_i = \frac{\sum_{j=1}^{J_i} \mathbf{Y}_{ij}}{J_i}; \quad \mathbf{S}_i = \frac{\sum_{j=i}^{J_i} (\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)\,(\mathbf{Y}_{ij} - \bar{\mathbf{Y}}_i)'}{(J_i - 1)}. \tag{1}$$

TABLE 1

Partial list of symbols used and their meanings

$\mathbf{Y}_{ij}$ = the genotypic vector representation of the j^{th} individual in the i^{th} population
$\bar{\mathbf{Y}}_i$ = the mean vector for the i^{th} population
$\bar{\mathbf{Y}}_c$ = the mean vector for the c^{th} cluster
$\bar{\mathbf{Y}}_T$ = the mean vector for the total population

$\mathbf{S}_i$ = the estimated covariance matrix for the i^{th} population
$\mathbf{S}_W$ = the pooled within-village covariance matrix
$\mathbf{S}_A$ = the among-villages covariance matrix
$\mathbf{S}_C$ = the among-clusters covariance matrix
$\mathbf{S}_V$ = the among-villages, within-clusters covariance matrix
$\mathbf{S}_T$ = the total covariance matrix

$\mathbf{V}_i = E(\mathbf{S}_i)$ $\mathbf{V}_W = E(\mathbf{S}_W)$ $\mathbf{V}_C = E(\mathbf{S}_C)$ $\mathbf{V}_V = E(\mathbf{S}_V)$ $\mathbf{V}_T = E(\mathbf{S}_T)$
$\boxminus$ = a measure of departure from random union of multiple-locus gametes
D_{hk} = the gametic disequilibrium between the h^{th} and k^{th} loci
r_{hk} = the estimated correlation between the h^{th} and k^{th} loci

$\mathbf{R}_i$ = the estimated correlation matrix derived from $\mathbf{S}_i$
$\mathbf{R}_W$ = the estimated correlation matrix derived from $\mathbf{S}_W$
$\mathbf{R}_A$ = the estimated correlation matrix derived from $\mathbf{S}_A$
$\mathbf{R}_C$ = the estimated correlation matrix derived from $\mathbf{S}_C$
$\mathbf{R}_V$ = the estimated correlation matrix derived from $\mathbf{S}_V$
$\mathbf{R}_T$ = the estimated correlation matrix derived from $\mathbf{S}_T$

$\mathbf{R}_e$ = the correlation matrix with a constant value r_e in all off-diagonal positions
r_e = a value chosen so that $\det \mathbf{R}_e = \det \mathbf{R}$, used as a measure of general departure from multiple-locus independence
Z = a test criterion for independence of sets of loci

Considering all $I = 50$ villages, the overall mean vector $\bar{\mathbf{Y}}_T$ and covariance matrix $\mathbf{S}_T$ are given by

$$\bar{\mathbf{Y}}_T = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{Y}_{ij}}{\sum_{i=1}^{I} J_i}; \quad \mathbf{S}_T = \frac{\sum_{i=j}^{I}\sum_{j=1}^{J_i}(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_T)(\mathbf{Y}_{ij}-\bar{\mathbf{Y}}_T)'}{(N-1)}, \tag{2}$$

where $N = \sum_{i=1}^{I} J_i$.

It is also convenient to define a variety of other matrices and mean vectors. We may compute a covariance matrix $\mathbf{S}_W$, which is the average within-village matrix

$$\mathbf{S}_W = \frac{\sum_{i=1}^{I}(J_i-1)\mathbf{S}_i}{(N-I)}, \tag{3}$$

and a matrix $\mathbf{S}_A$ describing the variation pattern among villages

$$\mathbf{S}_A = \frac{\sum_{i=1}^{I} J_i(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_T)(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_T)'}{(I-1)} \tag{4}$$

The relationship between $\mathbf{S}_T$, on the one hand, and $\mathbf{S}_A$ and $\mathbf{S}_W$, on the other, is given by

$$(N-1)\mathbf{S}_T = (I-1)\mathbf{S}_A + (N-I)\mathbf{S}_W. \tag{5}$$

For certain purposes it is convenient to recognize clusters of closely related villages and it is thus expedient to compute the weighted mean $\bar{\mathbf{Y}}_c$ of a cluster of villages (see p. 740 for definition of clusters). This allows further subdivision of the "among-villages" component $\mathbf{S}_A$ into a component describing the village to village variation $\mathbf{S}_V$ within a cluster and a component describing the cluster to cluster variation $\mathbf{S}_C$. The matrix $\mathbf{S}_A$ is related to $\mathbf{S}_V$ and $\mathbf{S}_C$ by

$$\begin{aligned}(I-1)\mathbf{S}_A &= \sum_{i=1}^{I} J_i(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_T)(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_T)' \\ &= \sum_{i=1}^{I} J_i(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_c)(\bar{\mathbf{Y}}_i-\bar{\mathbf{Y}}_c)' \\ &+ \sum_{c=1}^{C} J_c(\bar{\mathbf{Y}}_c-\bar{\mathbf{Y}}_T)(\bar{\mathbf{Y}}_c-\bar{\mathbf{Y}}_T)' \\ &= (I-C)\mathbf{S}_V + (C-1)\mathbf{S}_C,\end{aligned} \tag{6}$$

where C is the number of clusters and J_c is the sample size from the c^{th} cluster.

These mean vectors and covariance matrices have some useful relationships to familiar parameters. Consider the genotypic frequencies for a pair of loci (Figure 1).
The elements of the vector $\mathbf{Y}' = Y_1, Y_2)$ are (ignoring the population subscript)

$$\begin{aligned}\bar{Y}_1 &= 1(X_{1.}) + \tfrac{1}{2}(X_{2.}) + 0(X_{3.}) = \tilde{P}_1 = \tilde{f}(A_1) \\ \bar{Y}_2 &= 1(X_{.1}) + \tfrac{1}{2}(X_{.2}) + 0(X_{.3}) = \tilde{P}_2 = \tilde{f}(B_1)\end{aligned} \tag{7}$$

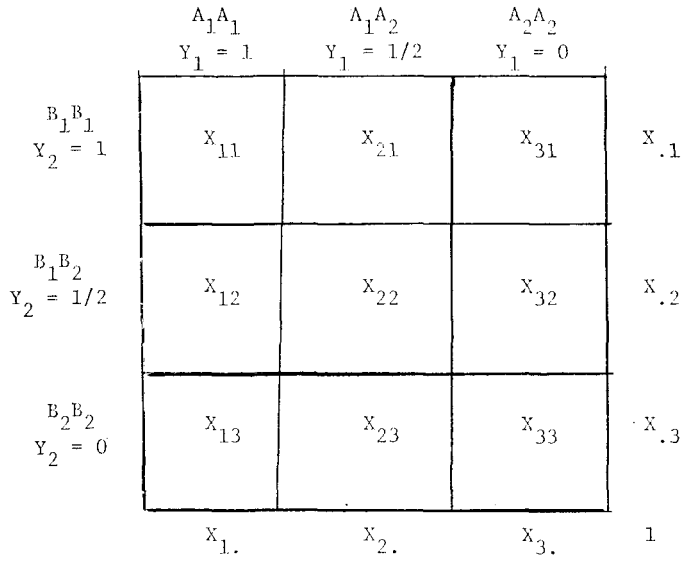

FIGURE 1.—Two-locus zygotic frequencies (X_{ij}) and genetic variables (Y_1, Y_2).

In general, $\bar{\mathbf{Y}}_i = \tilde{\mathbf{P}}_i$, the vector of observed allelic frequencies. The elements of the covariance matrix $\mathbf{S}_i$ can be elaborated in much the same fashion. Considering only the A and B loci, and ignoring the population subscript, these elements are

$$
\begin{aligned}
s_{11} &= X_{1.}\,(1-\tilde{P}_1)^2 + X_{2.}\,(\tfrac{1}{2}-\tilde{P}_1)^2 + X_{3.}\,(0-\tilde{P}_1)^2 \\
&= \ldots = \tilde{P}_1\,(1-\tilde{P}_1) - \frac{X_{2.}}{4} \\
s_{12} &= X_{11}\,(1-\tilde{P}_1)\,(1-\tilde{P}_2) + \ldots + X_{33}\,(0-\tilde{P}_1)\,(0-\tilde{P}_2) \\
&= \ldots = [X_{11} + \tfrac{1}{2}X_{12} + \tfrac{1}{2}X_{21} + \tfrac{1}{4}X_{22}] - \tilde{P}_1\tilde{P}_2 \\
s_{22} &= X_{.1}\,(1-\tilde{P}_2)^2 + X_{.2}\,(\tfrac{1}{2}-\tilde{P}_2)^2 + X_{.3}\,(0-\tilde{P}_2)^2 \\
&= \ldots = \tilde{P}_2\,(1-\tilde{P}_2) - \frac{X_{.2}}{4}\ .
\end{aligned}
\tag{8}
$$

In order to simplify these expressions further, it is necessary to deal with the implications of nonrandom mating. The two-locus zygotic array is generated from a two-locus gametic array (A_1B_1, A_1B_2, A_2B_1, A_2B_2) with gametic frequencies (P_{11}, P_{12}, P_{21}, P_{22}). The process of zygote formation may be viewed as being composed of two parts; (1) a portion $(1 - \boxminus)$ of zygotes is formed by random union of gametes, and (2) a portion $\boxminus$ is formed by union of identical gametes. Under these conditions, we expect the zygotic frequencies shown in Table 2. A similar line of argument has been employed by FALCONER (1961) to deal with inbreeding at a single locus. The parameter $\boxminus$ is not to be confused, however, with an inbreeding coefficient. It simply measures a departure from random mating, which might be due to population subdivision, assor-

TABLE 2

Two-locus zygotic frequencies, as observed and as expected from two locus gametic frequencies

Zygotic genotype	Observed frequencies	Expected frequencies
$A_1A_1B_1B_1$	X_{11}	$P^2_{11}(1-\square) + \square P_{11}$
$A_1A_1B_1B_2$	X_{12}	$2P_{11}P_{12}(1-\square)$
$A_1A_1B_2B_2$	X_{13}	$P^2_{12}(1-\square) + \square P_{12}$
$A_1A_2B_1B_1$	X_{21}	$2P_{11}P_{21}(1-\square)$
$A_1A_2B_1B_2$	X_{22}	$2(P_{11}P_{22} + P_{12}P_{21})(1-\square)$
$A_1A_2B_2B_2$	X_{23}	$2P_{12}P_{22}(1-\square)$
$A_2A_2B_1B_1$	X_{31}	$P^2_{11}(1-\square) + \square P_{21}$
$A_2A_2B_1B_2$	X_{32}	$2P_{21}P_{22}(1-\square)$
$A_2A_2B_2B_2$	X_{33}	$P^2_{22}(1-\square) + \square P_{22}$

A portion $\square$ of the zygotes lead to two locus homozygotes, and a portion $(1-\square)$ are formed by random mating.

tative mating or other features of the breeding system. Substituting from Table 2 into (8), we obtain the observed elements of the covariance matrix

$$s_{11} = \frac{(1+\tilde{\square})}{2}\tilde{P}_1(1-\tilde{P}_1) \qquad s_{22} = \frac{(1+\tilde{\square})}{2}\tilde{P}_2(1-\tilde{P}_2)$$

$$s_{12} = \frac{(1+\tilde{\square})}{2}\tilde{D}_{12} \tag{9}$$

with

$$\tilde{D}_{12} = \tilde{P}_{11}\tilde{P}_{22} - \tilde{P}_{12}\tilde{P}_{21} ,$$

the observed disequilibrium between the two loci. The computed covariance matrix $\mathbf{S}_i$ is an estimate of the parametric matrix $\mathbf{V}_i = \{v_{ihk}\}$, where

$$v_{ihk} = \frac{(1+\square_i)}{2}P_{ih}(1-P_{ih}) \quad h = k$$

$$v_{ihk} = \frac{(1+\square_i)}{2}D_{ihk} \quad h \neq k . \tag{10}$$

The matrix $\mathbf{S}_W$ is a weighted linear combination of the $\mathbf{S}_i$, and an estimate of the corresponding weighted combination of the $\mathbf{V}_i$. The matrix $\mathbf{V}_W$ is defined as

$$\mathbf{V}_W = \frac{\sum_{i=1}^{I}(J_i - 1)\mathbf{V}_i}{(N-I)} . \tag{11}$$

The matrix $\mathbf{S}_T$ is an estimate of the matrix $\mathbf{V}_T = \{v_{Thk}\}$, where

$$v_{Thk} = \frac{(1+\square_T)}{2}P_h(1-P_h) \quad h = k$$

$$v_{Thk} = \frac{(1+\square_T)}{2}\Delta_{hk} \qquad h \neq k , \tag{12}$$

with P_h the tribal average allele frequency for the h-th locus, and with the covariance measure Δ_{hk} defined by

$$\Delta_{hk} = \frac{(N-I)}{(N-1)} \bar{D}_{hk} + \frac{(I-1)}{(N-1)} \sigma_{hk} \quad . \tag{13}$$

The term $\bar{D}_{hk}$ is the hk-th element of $\mathbf{V}_W$, and σ_{hk} is the corresponding element of $\mathbf{V}_A$, the "among-populations" covariance matrix, estimated by $\mathbf{S}_A$. The diagonal elements σ_h^2 of $\mathbf{V}_A$ are the usual WAHLUND (1928) variances, while the off-diagonal elements are the corresponding covariance terms. The matrices $\mathbf{V}_G$ and $\mathbf{V}_V$ are estimated by $\mathbf{S}_G$ and $\mathbf{S}_V$, respectively, and describe allelic frequency variation and covariation among and within clusters.

For our purposes, the key consideration is that the $\boxed{\ }$ measures can be factored out of $\mathbf{S}_T$, $\mathbf{S}_W$, and the $\mathbf{S}_i$, by computing correlation matrix equivalents. For the i^{th} village, we have

$$\tilde{r}_{hk} = \frac{\tilde{D}_{hk}\,(1+\tilde{\boxminus}_i)/2}{\sqrt{\tilde{P}_h \tilde{Q}_h \tilde{P}_k \tilde{Q}_k\,(1+\tilde{\boxminus}_i)^2/4}} = \frac{\tilde{D}_{hk}}{\sqrt{\tilde{P}_h \tilde{Q}_h \tilde{P}_k \tilde{Q}_k}} \quad , \tag{14}$$

where $Q_h = (1 - P_h)$ and $Q_k = (1 - P_k)$. The $\boxminus$ terms also cancel out of $\mathbf{R}_A$, $\mathbf{R}_G$, and $\mathbf{R}_V$. The correlations of these matrices are simply two-locus allelic frequency correlations among villages, among clusters, and among villages-within clusters, respectively.

Test criteria

ANDERSON (1958, Ch. 9) has given asymptotic χ^2 procedures for testing the independence of sets of multivariate normal variables. Since our scaling device leads asymptotically to normality, we shall also use these test criteria. By doing so, we are adding an additional degree of approximation. If the matrix R is partitioned into sections corresponding to $G < H$ sets of variables

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1G} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2G} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \mathbf{R}_{G1} & \mathbf{R}_{G2} & \cdots & \mathbf{R}_{GG} \end{bmatrix} \quad , \tag{15}$$

then the test of the hypothesis that the off-diagonal submatrices have zero elements is given by

$$-M \log Z \sim \chi^2_f \qquad Z = \frac{|\mathbf{R}|}{\prod_{g=1}^{G} |\mathbf{R}_{gg}|} \tag{16}$$

where H_g is the dimension of R_{gg}, and

$$M = N - 3/2 - \frac{H^3 - \sum_{g=1}^{G} H_g^3}{3(H^2 - \sum_{g=1}^{G} H_g^2)} \qquad f = \tfrac{1}{2}\,[H^2 - \sum_{g=1}^{G} H_g^2)] \,. \tag{17}$$

If $G = H$, each "set of variables" consists of a single character (representing a single locus), and we need a test of the hypothesis that all correlations are zero

$$\begin{gathered} -M' \log |\mathbf{R}| \sim \chi^2_{f'} \\ M' = \left[N - 3/2 - \frac{H+1}{3} \right] \qquad f' = \tfrac{1}{2}\,[H(H-1)] \,. \end{gathered} \tag{18}$$

We shall have occasion to use the test in both forms. If the matrix R is diagonal (all off-diagonal elements are zero), then perfect gametic equilibrium obtains. If (18) is significant, recourse to (16) should allow us to pursue the source of the disequilibrium.

Although the determinants of the **R**-matrices provide an economical summarization of the information about correlations among a set of loci, they are a bit unwieldy for ready interpretation. As an aid to communication, we find it useful to carry the summarization one step further, and therefore define an "effective correlation" (r_e) as follows. For any particular matrix **R** of rank (k), there exists a matrix $\mathbf{R}_e$, with a constant correlation r_e in all off-diagonal positions, such that det $\mathbf{R}$ = det $\mathbf{R}_e$. The effective correlation r_e is therefore that constant value which yields the same determinant as the actual set of correlations encountered and can be used to summarize the whole set. The value of r_e is related to det **R** by the polynominal (see HOHN 1964, p. 71)

$$\det \mathbf{R} = \det \mathbf{R}_e = (1 - r_e)^k + kr_e(1 - r_e)^{k-1} \quad . \tag{19}$$

Although we shall use det **R** for testing, we shall routinely report r_e. Although (19) is a polynomial in r_e, there is a single solution on the [0, 1] interval, and it is this solution we shall report. The Z-value of (17) does not readily admit of such a clear-cut translation and we shall report it as calculated.

RESULTS

We report here the disequilibrium within and among a set of 50 villages, for most of which allelic frequencies have already been published. We have grouped these 50 villages into a set of nine clusters (or groups) defined by WARD and MIGLIAZZA (personal communication). These clusters are based on historical,

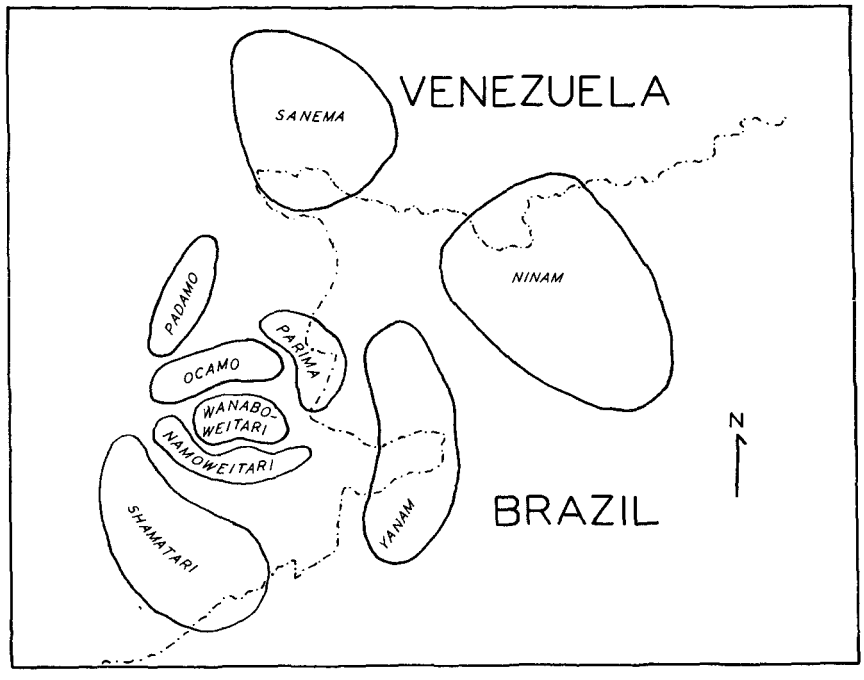

FIGURE 2.—Territorial extent of nine clusters of Yanomama villages.

cultural and linguistic information [see also WARD (1972); SPIELMAN, MIGLIAZZA and NEEL (1974); WARD (1976)], and exhibit considerable geographic cohesiveness as well. We have utilized here only those individuals in each village with a complete set of characterizations for the eight codominant marker systems [*MN*, *Ss*, *RhC*, *RhE*, *Hp*, *Gc*, Serum *Alb*, and PGM_1], so that our sample sizes are sometimes smaller than reported elsewhere for these same villages. The village, cluster, and tribal allelic frequencies are listed, for reference, in Table 3; a map of the clusters is presented in Figure 2.

TABLE 3

Allelic frequencies for 8 loci and 50 Yanomama villages

Cluster & Village	Size	*M*	*S*	*C*	*E*	*Hp-1*	*Gc-1*	*Alb-N*	*PGM-1*
Namoweitari									
03A	45	0.600	0.100	0.889	0.189	0.922	0.922	0.889	0.878
03B	27	0.593	0.019	0.889	0.185	0.944	0.944	0.944	0.907
03C	28	0.589	0.125	0.982	0.179	0.839	1.000	0.893	0.929
08ABC	165	0.518	0.064	0.936	0.258	0.845	0.976	0.906	0.942
08UVW	138	0.558	0.152	0.996	0.080	0.909	0.971	0.938	0.971
11T	31	0.355	0.177	0.919	0.177	0.790	0.952	0.919	1.000
Pooled	434	0.537	0.105	0.949	0.179	0.876	0.967	0.917	0.946
Shamatari									
03D	23	0.826	0.065	0.978	0.196	0.674	0.978	1.000	0.978
03H/11J	71	0.704	0.176	0.958	0.176	0.641	0.746	0.993	0.873
11G	52	0.798	0.106	0.885	0.212	0.625	0.750	1.000	0.942
11HI	124	0.798	0.129	0.956	0.250	0.786	0.903	0.984	0.923
11YZ	100	0.795	0.125	0.980	0.235	0.870	0.900	0.925	0.920
15QR	108	0.870	0.093	0.875	0.213	0.685	0.958	0.870	0.926
Pooled	478	0.801	0.121	0.936	0.221	0.736	0.879	0.950	0.920
Padamo									
08O	30	0.633	0.100	1.000	0.167	0.867	0.717	0.900	0.983
08Q	47	0.649	0.053	0.915	0.309	0.511	0.915	0.947	0.989
08R	56	0.705	0.268	0.866	0.339	0.768	0.857	0.946	0.973
Pooled	133	0.669	0.154	0.914	0.289	0.699	0.846	0.936	0.981
Wanaboweitari									
03E	24	0.583	0.125	0.896	0.250	0.917	0.979	0.958	1.000
03F	14	0.679	0.179	0.821	0.321	0.750	0.929	1.000	0.929
03G	28	0.500	0.107	0.911	0.196	0.893	0.982	0.839	0.964
03I	24	0.667	0.188	0.979	0.104	0.854	1.000	0.958	0.875
08N	34	0.809	0.132	0.971	0.147	0.985	0.956	0.912	0.941
08S	48	0.448	0.094	0.979	0.229	0.823	0.917	0.927	0.990
08T	25	0.620	0.100	0.960	0.220	0.820	0.940	0.880	1.000
Pooled	197	0.599	0.124	0.944	0.203	0.871	0.954	0.919	0.962

TABLE 3—Continued

Cluster & Village	Size	*M*	*S*	*C*	*E*	*Hp-1*	*Gc-1*	*Alb-N*	*PGM-1*
Ocamo									
03J	53	0.877	0.302	0.858	0.208	0.708	0.849	0.962	0.962
08J	26	0.615	0.250	1.000	0.096	0.865	0.846	0.942	0.904
08L/11P	79	0.753	0.253	0.968	0.108	0.911	0.962	0.981	0.956
08K/15M	77	0.487	0.117	0.987	0.071	0.851	0.903	0.974	1.000
11K	44	0.909	0.205	0.943	0.273	0.864	0.977	0.909	1.000
11M	28	0.661	0.268	0.946	0.196	0.929	0.911	1.000	0.964
Pooled	307	0.710	0.221	0.951	0.147	0.852	0.915	0.964	0.971
Sanema									
03U	36	0.514	0.278	0.958	0.042	0.861	0.653	0.681	0.819
08D	30	0.467	0.067	1.000	0.000	0.867	0.683	0.733	0.950
08E	38	0.724	0.171	0.921	0.079	0.934	0.882	0.605	0.974
08F	30	0.583	0.100	0.950	0.050	0.850	0.867	0.600	0.917
08I	59	0.475	0.068	0.975	0.042	0.932	0.864	0.932	0.949
Pooled	193	0.547	0.132	0.961	0.044	0.896	0.801	0.738	0.925
Parima									
08XY	120	0.575	0.288	0.921	0.033	0.912	0.854	0.971	0.988
11ABC	147	0.483	0.207	0.990	0.010	0.796	0.915	0.997	1.000
11D	50	0.380	0.180	1.000	0.000	0.880	0.900	0.740	1.000
11S	39	0.474	0.090	0.987	0.038	0.872	0.667	0.821	1.000
11U	28	0.446	0.161	0.946	0.196	0.875	0.893	0.929	0.964
11V	27	0.574	0.222	0.963	0.148	0.907	0.630	0.778	1.000
11X	43	0.407	0.070	1.000	0.058	0.988	0.733	0.942	1.000
15H	41	0.451	0.183	0.927	0.073	0.854	0.878	0.951	0.927
Pooled	495	0.488	0.199	0.966	0.044	0.871	0.843	0.926	0.989
Yanam									
03KP	71	0.648	0.324	0.859	0.162	0.880	0.894	0.915	1.000
03LMN	58	0.543	0.302	0.759	0.172	0.845	0.793	0.966	1.000
03Q	30	0.733	0.533	0.750	0.217	0.883	0.833	0.967	1.000
03RS	41	0.671	0.220	0.793	0.207	0.780	0.744	0.988	0.988
03T	30	0.750	0.250	0.817	0.217	0.650	0.733	0.983	0.867
Pooled	230	0.650	0.317	0.802	0.187	0.824	0.813	0.957	0.980
Ninam									
03W	64	0.828	0.328	0.695	0.336	0.797	0.922	0.961	0.961
03X	69	0.768	0.087	0.681	0.297	0.804	0.797	0.978	0.942
15L	78	0.622	0.179	0.788	0.244	0.596	0.795	0.917	0.840
15O	23	1.000	0.370	0.717	0.413	0.739	0.457	0.957	0.957
Pooled	234	0.759	0.212	0.724	0.301	0.726	0.797	0.951	0.915
Overall Pooled	2701	0.635	0.171	0.917	0.167	0.822	0.876	0.924	0.954

For purposes of later reference, we should remind the reader that the *C* and *E* markers of the *Rh* system are, for all practical purposes, absolutely linked, as are *MN* and *Ss* of the *MNSs* system. Each of these complexes (*Rh*, *MNSs*) is unlinked to the other loci. These latter are also unlinked, except for the *Gc* and *Alb* loci, which are thought to be about two centimorgans apart (WEITKAMP, RUCKNAGEL and GERSHOWITZ 1966). We know *a priori* that there will be disequilibria within the *Rh* and *MNSs* complexes [this is almost always so], but the unlinked loci are expected to be uncorrelated, except for stochastic factors. The *Gc-Alb* pair is less predictable, and will bear watching as we proceed.

The correlation matrices $\mathbf{R}_T$ and $\mathbf{R}_A$ are listed in Table 4. The former are in the upper triangular portion of the matrix and the latter in the lower triangular portion. The matrix $\mathbf{R}_T$ measures the overall disequilibrium within the tribe, whatever the source. With the exception of the large correlation between *MN* and *Ss* and that between *Rh-C* and *Rh-E*, all of the correlations of $\mathbf{R}_T$ are quite small, including that between *Gc* and *Alb*, suggesting that for the whole tribe, the various loci are quite close to statistical independence. To determine the impact of subdivision on disequilibrium, it is necessary to examine the other **R**-matrices. If villages had been formed by random sampling from a "super-gene-pool," itself in a state of gametic equilibrium, then we should expect no particular correlation in village allelic frequencies across loci, and $\mathbf{R}_A$ should be diagonal. As long as further fusions and migration are random, this situation should persist through time. Most of the off-diagonal elements of $\mathbf{R}_A$ are appreciably different from zero, however, indicating that the villages were neither constructed nor maintained in random fashion. Since fission has occurred along familial lines in dichotomous fashion, and since fusion and migration follow previous sociopolitical alliances, this is entirely to be expected.

A natural expectation from the dichotomous fission-fusion model described above is that the matrix $\mathbf{R}_C$, representing the correlations of cluster allelic frequencies at different loci, should be strongly nondiagonal, while the matrix $\mathbf{R}_V$, measuring the correlations of allelic frequencies among villages-within clusters, should have smaller off-diagonal elements. These latter two matrices are listed in Table 5, $\mathbf{R}_C$ above the diagonal and $\mathbf{R}_V$ below the diagonal. The expected patterns

TABLE 4

Interlocus correlation matrices: (a) total population correlations above the diagonal and (b) among-village correlations below the diagonal

	MN	Ss	Cc	Ee	Hp	Gc	Alb	PGM
MN	1	0.175	—0.070	0.132	—0.053	0.002	0.010	—0.018
Ss	0.197	1	—0.068	0.000	—0.005	0.026	0.053	0.036
Cc	—0.417	—0.387	1	—0.568	0.011	0.024	—0.026	—0.011
Ee	0.594	—0.037	—0.586	1	—0.047	0.005	0.036	0.001
Hp	—0.366	0.063	0.342	—0.450	1	0.035	—0.071	0.101
Gc	—0.025	—0.230	0.264	0.038	0.186	1	0.040	0.018
Alb	0.185	0.173	—0.167	0.234	—0.218	0.120	1	0.007
PGM	—0.271	0.174	0.132	—0.239	0.343	0.145	0.119	1

TABLE 5

Correlation matrices: (a) among-clusters variation above the diagonal and (b) within clusters variation below the diagonal

	MN	Ss	Cc	Ee	Hp	Gc	Alb	PGM
MN	1	0.031	—0.422	0.727	—0.823	—0.125	0.433	—0.582
Ss	0.398	1	—0.529	—0.098	0.005	—0.555	0.332	0.515
Cc	—0.418	—0.207	1	—0.613	0.509	0.511	—0.287	0.268
Ee	0.338	0.048	—0.515	1	—0.775	0.102	0.479	—0.495
Hp	0.140	0.111	0.129	—0.043	1	0.329	—0.443	0.463
Gc	0.074	0.004	—0.028	—0.035	0.088	1	0.211	0.014
Alb	—0.097	0.042	0.001	—0.091	—0.038	0.057	1	0.220
PGM	0.035	—0.079	—0.040	0.053	0.258	0.228	0.045	1

are observed, although the correlations of $\mathbf{R}_V$ are not as small as one might expect. It appears that there is patterned infra-structure even within clusters.

The matrix $\mathbf{R}_W$ is given as the upper triangular portion of Table 6, and is the average pattern within the 50 villages. Unless there are systematic forces operative, all off-diagonal elements should be close to zero. With the exception of the correlation between *MN* and *Ss* and that between *Rh-C* and *Rh-E*, all elements of the matrix are quite close to zero, including that for the *Gc-Alb* pair. The large internal correlations of the *MNSs* (0.172) and *Rh* (—0.566) complexes are entirely expected, and provide a measure of the effects of very tight linkage. There is *no* tendency for these disequilibria to decay. Aside from these two easily explained exceptions, there is essentially no evidence for systematic pressures in the "average village." [The other disequilibria range only from —0.048 (*Hp* and *Gc*) to +0.072 (*Hp* and *PGM*).] The average village is an abstraction, of course, and we should expect individual villages to depart from this ideal for purely stochastic reasons. The individual villages might well be considerably out of equi-

TABLE 6

Interlocus correlation matrices: (a) average within-village correlations above the diagonal, and (b) average absolute within-village correlations below the diagonal

	MN	Ss	Cc	Ee	Hp	Gc	Alb	PGM
MN	1	0.172	—0.000	0.051	0.002	0.007	—0.027	0.016
Ss	0.333	1	—0.018	0.005	—0.015	0.065	0.034	0.021
Cc	0.154	0.151	1	—0.566	—0.048	—0.022	0.004	—0.031
Ee	0.178	0.136	0.535	1	0.016	—0.001	0.000	0.030
Hp	0.130	0.138	0.123	0.117	1	0.009	—0.044	0.072
Gc	0.121	0.130	0.128	0.116	0.143	1	0.024	0.001
Alb	0.107	0.096	0.110	0.107	0.104	0.152	1	—0.009
PGM	0.118	0.087	0.093	0.120	0.111	0.101	0.089	1

librium. The departure of individual villages from this average may be described in various ways, one of which is to define the average absolute r-values

$$\overline{|r_{hk}|} = \frac{\sum_{i=1}^{I} J_i |r_{ihk}|}{N}, \tag{20}$$

which are given as the lower triangular portion of Table 6. The results clearly indicate large departures from equilibrium in any one village.

Another means of assessing the departure of individual villages from gametic equilibrium is to examine the effective correlation values for each village. The r_e values for the (8×8) matrices are listed as r_1 in Table 7. The χ^2 tests were computed according to (18), and those tests exceeding the $\alpha = 0.01$ level are indicated. We have also listed the r_e-values for $\mathbf{R}_A$, $\mathbf{R}_G$, $\mathbf{R}_V$, $\mathbf{R}_W$, and $\mathbf{R}_T$ at the bottom of the table, and those for the pooled within-village matrix separately for each cluster in the body of the table.

The impressions gained from examination of Tables 4, 5, and 6 are borne out by the effective correlations at the bottom of Table 7. The matrix $\mathbf{R}_A$ departs markedly from the independence condition ($r_e = 0.387$), and the anticipated difference between $\mathbf{R}_G$ and $\mathbf{R}_V$ is obtained ($r_e = 1.000$ *vs* $r_e = 0.247$). We should point out that since there are nine clusters and only eight loci, the virtual singularity of $\mathbf{R}_G$ is not a structural feature; rather it is an indication of strong correlations among the cluster allelic frequencies for different loci. The $\mathbf{R}_W$ and $\mathbf{R}_T$ matrices more nearly approach the uncorrelated state ($r_e = 0.154$ and $r_e = 0.162$, respectively). We should mention here that $\mathbf{S}_W$ dominates $\mathbf{S}_T$ (equation 5), $\mathbf{V}_W$ dominates $\mathbf{V}_T$ (equation 13), and thus $\mathbf{R}_W$ dominates $\mathbf{R}_T$.

The r_e-values for individual villages are all larger than that for $\mathbf{R}_W$; many are considerably larger. This finding is consistent with the observation that all of the average absolute values $\overline{|r_{hk}|}$ are quite large. A careful examination of the 50 $\mathbf{R}_i$ matrices would indicate that the disequilibrium between any pair of loci is highly erratic from village to village, but tends to average out overall. The pooled within-village matrices for the various clusters represent averages for a small number of villages, and should yield smaller effective correlations than single villages, but larger values than $\mathbf{R}_W$. That this is so is indicated by the results in Table 7.

The departures of all of these r_1-values from zero are somewhat enhanced by the internal correlations of the *MNSs* and/or *Rh* systems. We may remove the linkage factor from consideration by treating each of these linked pairs as a "set," and using the Z-criterion of (16). We have not translated Z into r_e, and have listed the appropriate values (as computed) in Table 7. All of the conditional test criteria (unlisted) were smaller than the corresponding unconditional test criteria (also unlisted), and this is indicated in the table by the fact that fewer of the former are significant. A comparison of r_e and Z is rather cumbersome, but it is useful to recall that the independence condition is indicated by $r_e = 0$ and

TABLE 7

Standardized measures for all correlation matrices

Source of correlation	Sample size (N)	Correlation measures					
		r_1	Z	r_2	r_3	r_4	r_5
Wanaboweitari							
03E	24	0.471*	0.098*	0.231	0.302	0.483*	0.515*
03F	14	0.532	0.057	0.437	0.528	0.473	0.558
03G	28	0.302	0.390	0.272	0.229	0.206	0.170
03I	24	0.337	0.288	0.243	0.238	0.293	0.377
08N	34	0.235	0.464	0.176	0.230	0.161	0.137
08S	48	0.210	0.598	0.148	0.193	0.119	0.157
08T	25	0.318	0.363	0.233	0.331	0.182	0.268
Pooled	197	0.083	0.866	0.068	0.075	0.072	0.085
Ocamo							
03J	53	0.384*	0.320*	0.270*	0.274*	0.218	0.228
08J	26	0.306	0.380	0.198	0.209	0.280	0.277
08K/15M	77	0.268*	0.343*	0.318*	0.314*	0.299*	0.271*
08L/11P	79	0.199*	0.700	0.143	0.138	0.103	0.127
11K	44	0.340*	0.300*	0.332*	0.251	0.316*	0.258
11M	28	0.284	0.442	0.165	0.199	0.297	0.263
Pooled	307	0.191*	0.754*	0.119*	0.121*	0.116*	0.104*
Sanema							
03U	36	1.000*	0.125*	0.251	0.251	0.320*	0.320*
08D	30	0.308	0.487	0.283	0.283	0.263	0.263
08E	38	1.000*	0.108*	0.395*	0.395*	0.378*	0.378*
08F	30	1.000*	0.102*	0.411*	0.411*	0.282	0.282
08I	59	0.242*	0.513	0.190	0.160	0.217	0.147
Pooled	193	0.336*	0.791	0.121*	0.119*	0.095	0.099
Namoweitari							
03A	45	0.413*	0.207*	0.257	0.283*	0.345*	0.339*
03B	27	0.442*	0.226	0.296	0.321	0.209	0.245
03C	28	0.289	0.472	0.179	0.288	0.106	0.207
08ABC	165	0.215*	0.617*	0.184*	0.203*	0.161*	0.183*
08UVW	138	0.175*	0.608*	0.146*	0.160*	0.154*	0.164*
11T	31	0.397	0.272	0.324	0.308	0.382*	0.331
Pooled	434	0.150*	0.883*	0.086*	0.088*	0.081*	0.083*
Shamatari							
03D	23	0.245	0.492	0.143	0.189	0.250	0.179
03H/11J	71	0.204	0.689	0.158	0.152	0.133	0.125
11G	52	0.382*	0.458*	0.275*	0.254	0.290*	0.278*
11HI	124	0.234*	0.726	0.127	0.114	0.154*	0.143
11YZ	100	0.276*	0.641	0.136	0.134	0.145	0.146
15QR	108	0.293*	0.623*	0.165*	0.186*	0.154	0.176*
Pooled	478	0.186*	0.927	0.059*	0.053*	0.060*	0.054*

TABLE 7—Continued

Source of correlation	Sample size (N)	Correlation measures					
		r_1	Z	r_2	r_3	r_4	r_5
Parima							
08XY	120	0.328*	0.502*	0.208*	0.207*	0.168*	0.166*
11ABC	147	1.000*	0.479*	0.051	0.051	0.066	0.066
11D	50	0.340*	0.751	0.202	0.202	0.152	0.152
11S	39	0.197	0.617	0.144	0.190	0.180	0.166
11U	28	0.377	0.160	0.311	0.330	0.269	0.302
11V	27	0.449*	0.177*	0.341	0.325	0.457*	0.452*
11X	43	0.228	0.624	0.197	0.166	0.208	0.177
15H	41	1.000*	0.285*	0.224	0.224	0.164	0.164
Pooled	495	0.211*	0.902*	0.070*	0.075*	0.073*	0.073*
Yanam							
03KP	71	0.482*	0.625	0.152	0.120	0.209	0.176
03LMN	58	0.411*	0.585	0.235	0.178	0.182	0.134
03Q	30	0.493*	0.513	0.250	0.197	0.293	0.239
03RS	41	1.000*	0.331	0.180	0.180	0.118	0.118
03T	30	0.437*	0.271	0.261	0.231	0.330	0.312
Pooled	230	0.358*	0.865	0.086	0.079	0.101	0.094
Ninam							
03W	64	0.458*	0.613	0.164	0.160	0.166	0.165
03X	69	0.250*	0.494	0.205	0.197	0.139	0.152
15L	78	0.389*	0.540	0.202*	0.202*	0.183	0.186
15O	23	0.428*	0.332	0.347	0.358	0.324	0.334
Pooled	234	0.283*	0.769*	0.114*	0.122*	0.114*	0.127*
Padamo							
08O	30	0.338	0.275	0.368*	0.346	0.253	0.264
08Q	47	0.257	0.450	0.254	0.243	0.237	0.228
08R	56	0.298*	0.319	0.206	0.283*	0.206	0.202
Pooled	133	0.176*	0.694*	0.118	0.157*	0.119	0.129
Among Villages		0.387*	0.191*	0.287*	0.334*	0.286*	0.271*
Among Clusters		1.000*	0.000*	0.652*	0.690*	0.692*	0.735*
Within Clusters		0.247*	0.607*	0.196*	0.168*	0.145*	0.124*
Within Villages (Pooled)		0.154*	0.976*	0.030*	0.030*	0.036*	0.033*
Total Yanomama	2701	0.162*	0.945*	0.046*	0.057*	0.047*	0.045*

* Nominally significant at the $\alpha = 0.01$ level.

$r_1 = r_e$ from (8×8) correlation matrix.

Z = conditional correlation measure.

r_2, r_3, r_4, $r_5 = r_e$ values from reduced (6×6) matrices involving [*MN*, *Cc*], [*MN*, *Ee*], [*Ss*, *Cc*], and [*Ss*, *Ee*], respectively.

$Z = 1$, while maximum correlation (disequilibrium) is indicated by $r_e = 1$ and $Z = 0$.

To gauge the degree of correlation among "unlinked" markers, it is convenient to remove two columns and two rows from the various **R**-matrices. If we were to use *MN* and *Rh-C*, but not *Ss* and *Rh-E*, the internal correlations of the two complexes would be removed from consideration, and we would have a (6×6) matrix of correlations between "unlinked" loci. The determinant of this reduced matrix can be used to construct a test criterion, according to (18), and to derive an r_e-value. It is by no means clear whether we should use (*MN* and *Rh-C*), (*MN* and *Rh-E*), (*Ss* and *Rh-C*), or (*Ss* and *Rh-E*), however, and the results will differ, depending on the choice. We have, therefore, listed the corresponding r_e values for all four of these strategies as r_2, r_3, r_4, and r_5, respectively, in Table 6. Although these values are not statistically independent among themselves, they may be compared with the values listed for r_1.

As a general rule, r_1 is greater than r_2, r_3, r_4, or r_5, indicating inflation of the former by the internal correlations of the *MNSs* and *Rh* complexes. This result was foreshadowed by the *Z*-measures, but is more easily conveyed in this latter vein. After we account for the obvious correlations in these linked complexes, both $\mathbf{R}_W$ and $\mathbf{R}_T$ approach the uncorrelated state. This correction, however, has very little impact on $\mathbf{R}_A$, $\mathbf{R}_C$, and $\mathbf{R}_V$, thus indicating that the decay processes affecting disequilibria within populations do not apply to allelic frequency correlations across loci and among populations.

Within villages 03RS, 03U, 08F, 11ABC, and 15H, the only *Rh* haplotypes recovered are *Ce* and *cE*, so that this system is in a state of maximum disequilibrium. We have set $r_{34} = -0.99$ whenever necessary to test r_1 or Z [$-M'$ log $(0) = \infty$]. No adjustment was necessary for r_2, r_3, r_4, or r_5, since the offending correlation was removed from the matrix. An examination of Table 2 will show that some villages are fixed for one locus (or sometimes more). We have removed all such loci from consideration, and have computed the r_e-values and test-criteria on the reduced matrices.

DISCUSSION

Complex systems

The largest correlations among the eight loci are those internal to the *MNSs* and *Rh* complexes. Most of the purely technical difficulties of the analysis are encountered in the process of trying to "extract" these internal correlations from the various measures of departure from equilibrium. It might appear that this situation is an unfortunate aftermath of our entirely arbitrary decision to treat these complex systems as pairs of "tightly linked" two-allele loci. Although the question of which treatment is used is largely a matter of taste, it does seem appropriate at this juncture to describe how the analysis should proceed if these systems were treated as four-haplotype loci. In a more general vein, it seems advisable to extend the treatment to the multiple-allelic case.

Consider first the three allele-case. We define a pair of Y-variables for such a locus, and assign Y-values to the six genotypes

$$\begin{array}{cccccccc} & A_1A_1 & A_1A_2 & A_1A_3 & A_2A_2 & A_2A_3 & A_3A_3 \\ Y_1 & 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ Y_2 & 0 & \frac{1}{2} & 0 & 1 & \frac{1}{2} & 0 \end{array}, \tag{21}$$

which leads to

$$\begin{aligned} \bar{Y}_1 &= \tilde{f}(A_1) = \tilde{P} & \bar{Y}_2 &= \tilde{f}(A_2) = \tilde{Q} \\ s_{11} &= \frac{(1+\tilde{\exists})}{2}\tilde{P}(1-\tilde{P}) & s_{22} &= \frac{(1+\tilde{\exists})}{2}\tilde{Q}(1-\tilde{Q}) \\ s_{12} &= -\frac{(1+\tilde{\exists})}{2}\tilde{P}\tilde{Q} & r_{12} &= -\sqrt{\frac{\tilde{P}\tilde{Q}}{(1-\tilde{P})(1-\tilde{Q})}}\,. \end{aligned} \tag{22}$$

Correlations of either Y_1 or Y_2 with the Y-values of other loci lead to standard disequilibrium measures, on the pattern of (14). The internal correlations of (22) do not relate to disequilibria.

The extension to four alleles (the present case) is obvious, and leads to

$$\begin{aligned} r_{12} &= -\sqrt{\frac{\tilde{P}\tilde{Q}}{(1-\tilde{P})(1-\tilde{Q})}} & r_{13} &= -\sqrt{\frac{\tilde{P}\tilde{R}}{(1-\tilde{P})(1-\tilde{R})}} \\ & & r_{23} &= -\sqrt{\frac{\tilde{Q}\tilde{R}}{(1-\tilde{Q})(1-\tilde{R})}} \end{aligned} \tag{23}$$

where $\tilde{P} = \tilde{f}(A_1)$, $\tilde{Q} = \tilde{f}(A_2)$, $\tilde{R} = \tilde{f}(A_3)$, and $\tilde{S} = \tilde{f}(A_4) = [1 - \tilde{P} - \tilde{Q} - \tilde{R}]$. Again, correlations with the Y-variables of other loci lead to standard disequilibrium measures, on the pattern of (14).

The test criterion (18) is inflated by the correlations of (22) or (23), which cannot be zero and are not disequilibria. We must have recourse to the conditional test criterion (16), and are no better off than we were with the two-locus treatment. The strategy for obtaining an r_c-value is the same, but instead of $(2 \times 2) = 4$ choices, we have $(4 \times 4) = 16$ choices to consider.

Two other considerations, in addition to the above, led us to opt for the two-locus treatment. The first consideration is one of sampling. Multiple haplotype systems almost always exhibit low frequencies for all but one or two haplotypes, and that is certainly the case for the *MNSs* and *Rh* systems. As a consequence, particular haplotypes are often missing, reducing a potential (3×3) matrix to a (2×2) matrix, no larger than that of the two-locus treatment. The second consideration is that the four-haplotype treatment requires unambiguous resolution

of all genotypes into haplotypes. This is generally difficult for double heterozygotes. A decision on Rh is possible for the Yanomama, because all "double heterozygotes" were typed with anti-f, which yields a positive test for (*CE//ce*) and a negative result for (*Ce//cE*). No such resolution is possible for the (*MS//Ns*) and (*Ms//NS*) genotypes. We have chosen the easiest (two-locus) route, but others may prefer the alternative. The analytical problems are the same in both cases.

Distributional approximations

We find it particularly convenient to use the multivariate normal test criteria given by ANDERSON (1958) and listed above. Because our (1, 1/2, 0) scoring device leads to approximate multivariate normality only in the limit, it is probable that the nominal α-levels are much smaller than the actual values. To check the adequacy of the asymptotic approximation, we have conducted a limited number of Monte-Carlo trials. Using the observed allele frequencies of a particular village ($\bar{\mathbf{Y}}_i = \tilde{\mathbf{P}}_i$), we have generated an array of multiple-locus gametes (assuming gametic equilibrium), and combined them randomly into zygotes. We have then drawn individuals at random (both with and without replacement), and computed the test-criteria.

The results are easily summarized. For samples of size $N = 120$, the actual probability of exceeding the nominal ($\alpha = 0.05$) critical value is seldom more than 0.07; for samples of size $N = 30$, the comparable value may be as high as 0.10. For $N = 120$, the observed probability of exceeding the nominal ($\alpha = 0.01$) critical value is seldom more than 0.02, while the comparable figure for $N = 30$ is sometimes as high as 0.04. The exact figures vary a bit, depending on the mean vector employed, but the above yields a reasonable picture of the overall pattern. We have therefore noted only those test criterial exceeding the nominal ($\alpha = 0.01$) level, and would expect to exceed this level no more than 5% of the time, given that the null hypothesis is correct. In fact (16/50) = 32% of the single village Z-criteria are significant at this level, and we therefore reject the null hypothesis.

Implications

The fission-fusion dynamics inherent in Yanomama village demography lead to considerable gametic disequilibrium within single villages. With the exception of the internal disequilibria of the *MNSs* and *Rh* systems, there is no consistent pattern in the sign (+ or −) of any particular disequilibrium from village to village, suggesting the absence of systematic forces (epistatic selection). This observation follows from a comparison of the symmetric elements of Table 6. The matrix $\mathbf{R}_W$ departs significantly from the independence condition, but the values of r_2, r_3, r_4, and r_5 are all less than 0.04, scarcely an exciting departure from zero. On the other hand, while we see no need to invoke epistatic selection, we certainly cannot exclude it as a possible explanation of these small average correlations.

The r_e-values for $\mathbf{R}_A$, $\mathbf{R}_C$, and $\mathbf{R}_V$ are all large and highly significant, and clearly indicate nonrandomness in the fission-fusion process. The r_e-values for $\mathbf{R}_T$, on the other hand, are fairly close to those of $\mathbf{R}_W$, indicating essential inde-

pendence of different loci for the whole tribe. The cause of this apparent anomaly is that $\mathbf{R}_A$ contributes 49 sets of information to $\mathbf{R}_T$, while $\mathbf{R}_W$ contributes 2651 sets, so that the large correlations of $\mathbf{R}_A$ are diluted. That disequilibrium which does exist is largely hidden by sample lumping, and depends for its elucidation on the ability to subdivide the population properly. We cannot but wonder, along with SINNOCK (1975), how many populations, human and otherwise, have yielded negative evidence of gametic disequilibrium simply because too little attention was directed to the sampling frame.

The results of our analyses indicate that the internal genetic disruption of single villages is both substantial and pervasive. For the tribe as a whole, correlations are small. This is not to say that the Yanomama are in multiple-locus H-W equilibrium. We already know (NEEL and WARD 1972) that $F_{IT} \approx 0.045$. On the whole, however, departure from multiple-locus H-W is more due to the WAHLUND (1928) effect than to gametic disequilibrium.

While the results presented are specific for the Yanomama, we suspect they are reasonably typical of undisturbed tribal populations at this cultural level. In a future paper, we shall examine the disequilibria of more acculturated (and more disrupted) tribal groups, by way of comparison. We will also explore some of the implications of these findings for treatments which measure selective disadvantage in terms of departure from some multi-locus optimal genotype.

The authors would like to express their appreciation to Ms. M. PARK, whose programming efforts reduced this large task to manageable proportions. We also thank DR. P. MOLL for critical comments which have improved the manuscript. The formulation presented here, as well as any errors of omission or commission, remain the responsibility of the authors.

LITERATURE CITED

ANDERSON, T. W., 1958 *An Introduction to Multivariate Statistical Analysis.* Wiley and Sons, New York.

BREWER-CARIAS, C. A., S. LE BLANC and J. V. NEEL, 1976 Genetic structure of a tribal population, the Yanomama Indians. XIII. Dental microdifferentiation. Am. J. Phys. Anthrop. **44:** 5–14.

FALCONER, D. S., 1961 *Introduction to Quantitative Genetics.* Ronald Press, New York.

HOHN, F. E., 1964 *Elementary Matrix Algebra.* 2nd Edition. MacMillan, New York.

NEEL, J. V., 1967 The genetic structure of primitive human populations. Jap. J. Hum. Genet. **12:** 1–16.

NEEL, J. V., F. ROTHHAMMER and J. C. LINGOES, 1974 The genetic structure of a tribal population, the Yanomama Indians. X. Agreement between representations of village distances based on different sets of characteristics. Am. J. Hum. Genet. **26:** 281–303.

NEEL, J. V. and R. H. WARD, 1972 The genetic structure of a tribal population, the Yanomama Indians. VI. Analysis by F-statistics (including a comparison with the Makiritare). Genetics **72:** 639–666.

NEI, M. and W. H. LI, 1973 Linkage disequilibrium in subdivided populations. Genetics **75:** 213–219.

ROTHHAMMER, R., J. V. NEEL, F. DAROCHA and G. Y. SUNDLING, 1973 The genetic structure of a tribal population, the Yanomama Indians. VIII. Dermatoglyphic differences among villages. Am. J. Hum. Genet. **25:** 152–166.

SINNOCK, P., 1975 The Wahlund effect for the two locus model. Am. Naturalist **109**: 565–570.

SPIELMAN, R. S., 1973a Differences among Yanomama Indian villages: Do the patterns of allele frequencies, anthropometrics and map locations correspond? Am. J. Phys. Anthrop. **39**: 461–479. ——, 1973b Do all the natives look alike? Size and shape components of anthropometric differences among Yanomama Indian villages. Am. Naturalist **107**: 694–708.

SPIELMAN, R. S., F. J. DAROCHA, L. R. WEITKAMP, R. H. WARD, J. V. NEEL and N. A. CHAGNON, 1972 The genetic structure of a tribal population, the Yanomama Indians. VII. Anthropometric differences among Yanomama villages. Am. J. Phys. Anthrop. **37**: 345–356.

SPIELMAN, R. S., E. C. MIGLIAZZA and J. V. NEEL, 1974 Regional linguistic and genetic differences among Yanomama Indians. Science **184**: 637–644.

SPIELMAN, R. S., J. V. NEEL and F. H. F. LI, 1977 Inbreeding estimation from population data: Models, procedures and implications. Genetics **85**: 355–371.

WAHLUND, S., 1928 Zusammensetzung von Populationen und Korrelation-serscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. Hereditas **2**: 65–106.

WARD, R. H., 1972 The genetic structure of a tribal population, the Yanomama Indians. V. Comparisons of a series of genetic networks. Ann. Hum. Genet. **36**: 21–43. ——, 1976 Social structure and variability in a tribal population. Am. J. Phys. Anthrop. **44**: 213–214.

WEITKAMP, L. R., D. L. RUCKNAGEL and H. GERSHOWITZ, 1966 Genetic linkage between structural loci for albumin and group specific component (Gc). Am. J. Hum. Genet. **18**: 559–571.

WRIGHT, S., 1943 Isolation by distance. Genetics **28**: 114–138. ——, 1951 The genetical structure of populations. Ann. Eugen. **15**: 323–354. ——, 1965 The interpretation of population structure by F statistics with special regard to systems of mating. Evolution **19**: 395–420.

Corresponding editor: R. W. ALLARD