A COMPARISON OF THE GENETIC INFRASTRUCTURE OF THE YE'CUANA AND THE YANOMAMA: A LIKELIHOOD ANALYSIS OF GENOTYPIC VARIATION AMONG POPULATIONS[1]

PETER E. SMOUSE

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

AND

RICHARD H. WARD

Department of Epidemiology and Center for Child Development and Mental Retardation, University of Washington, Seattle, Washington 98195

Manuscript received October 6, 1977

ABSTRACT

A general procedure is described for measuring and testing population differences in gametic frequencies. The total dispersion among populations is subdivided in hierarchical fashion. The multiple-locus treatment is simply the sum of the single-locus analyses, provided gametic equilibrium obtains among the loci. In the event that gametic equilibrium does not obtain, correlations among loci need to be dealt with.—The analysis is then used to examine the genetic infrastructure of two Indian tribes from South America, the Ye'cuana (Makiritare) and the Yanomama. From historical evidence, we may identify several "clusters" of villages within each tribe. The demographic and cultural practices affecting village formation and the maintenance of peer integrity are rather different in these tribes, however, and lead us to postulate rather different patterns of genetic variation among villages. Analyses of five codominant two-allele loci, four dominant two-allele loci and two complex loci (with four codominant haplotypes each) demonstrate that Yanomama clusters are more disparate than Ye'cuana clusters, as would have been predicted on sociocultural grounds.

THE past decade has witnessed an explosive increase in the number of field studies undertaken to determine the distributions of allelic variants in natural populations. The variety of explanations for and interpretations of the resulting patterns of variation have unfortunately increased at a comparable rate. The interested reader is referred to HEDRICK, GINEVAN and EWING (1976) for an extensive review of the relevant literature. This results partly from a general failure to pose explicit questions and partly from a failure to pose these questions in an appropriate hypothesis-testing framework. This is particularly true of those studies (or analyses) that have attempted to identify and gauge the relative importance of the various factors that are commonly subsumed under the

[1] Supported by National Science Foundation grant BMS-74-11823 and ERDA grant E(11-1)-2828.

rubric "population structure," *i.e.*, population subdivision, interdemic migration, nonrandom mating, etc.

The purposes of this paper are two. First, we shall present a general strategy for the analysis of population differences in allelic frequency, which analysis is constructed on a hypothesis-testing framework. Second, we shall employ this analysis to examine the "population structure" of two tribal Ameridian populations—the Ye'cuana and the Yanomama—which have rather different demographic and cultural histories.

Numerous measures of "genetic distance" have been proposed, each differing from the others in detail, but esentially all derived from population differences in allelic frequencies. While it is generally agreed that population differences shed light on a variety of evolutionary problems, there is certainly no unanimity on procedure. The reader interested in a comparison of methods is referred to GOODMAN (1972), GOWER (1972), and LENNINGTON and FLAKE (1974). These measures convey somewhat different quantitative information (LATTER 1973), but many of them differ essentially by a multiplicative constant (FELSENSTEIN 1973). This suggests that a single metric should suffice for most purposes and that the choice of this metric should be dictated by fairly general considerations.

The genetic distance between two populations is, of course, only one of several population comparisons of interest. What is badly needed is a general analytical framework, within which a variety of problems may be formulated. The strategy will be to develop measures of population dispersion and corresponding χ^2 test criteria simultaneously from a set of likelihood functions describing the variation within various populations. Since our primary interest in this paper is population structure (as it reflects demographic and cultural factors) we shall emphasize the hierarchical partitioning of the total genetic dispersion among populations. The approach is more general than this, however, and also leads naturally to multinomial regression analyses, wherein simple population comparisons may be formulated in terms of linear (SMOUSE and KOJIMA 1972) or loglinear (SMOUSE 1974) models.

THE YE'CUANA AND THE YANOMAMA

Motivation

Our objective is to compare the genetic infrastructure of these two unacculturated tribes of South American Indians. Although these tribes occupy contiguous (partially overlapping) territories in Venezuela and neighboring Brazil, they are remarkably disparate in culture, language and genetic frequencies (cf., Ward *et al.* 1975). From historical and ethnographic evidence, it is possible to identify subtribal clusters of villages within each group, but the manner of formation and maintenance of these clusters is quite different in the two tribes. The manner in which villages are formed and maintain peer integrity differ markedly in the Ye'cuana and Yanomama, and such processes should influence the degree and pattern of microdifferentiation at both the village and cluster levels of "genetic organization."

Formally, we shall pose and answer two questions. (1) How does the total variation among villages compare for the two tribes? (2) How distinct are Ye'cuana clusters, relative to Yanomama clusters? We rather expect that the second question will be particularly interesting, in view of the following observations.

In both tribes, it is possible to define village clusters on the basis of historical events during the last 80–100 years. The Ye'cuana are numerically rather stable, and clusters are based largely on historical affinities rather than geographical, linguistic or cultural differences (WARD and NEEL 1970). The Yanomama, on the other hand, have been expanding numerically and geographically for at least 100 years. Clusters represent groups of villages descended from a single ancestral population (about 80 years ago for the clusters discussed here); these related villages have followed an interrelated migration path ever since (WARD 1972).

Ye'cuana villages are relatively invaariant in size (40–100 persons) and tend to be relatively stable communities. Villages typically form by factions breaking away from an existing village and by accretion (either of individuals or of factions from related villages) until a stable size is reached. The periodic formation and breakup of villages leads to shifting group membership and social organization. Yanomama villages have more often been formed by repeated fission as the population has grown and expanded into new territory. Once fission has occurred, subsequent genetic exchange between the derivative groups is somewhat limited. Thus, the historical relationships among sets of extant villages are best described by a dichotomous network (CHAGNON 1972; WARD 1972).

Genetic exchange among Ye'cuana villages tends to be rather wide-spread and to transcend cluster boundaries. This is so for two reasons. First, Ye'cuana men engage in extensive trading over a far-flung area, and in the course of their travels, marriages may be contracted or more temporary liaisons established. Second, Ye'cuana villigaes react to intrusive stress from the outside (such as intrusions by rubber tappers or Yanomama) by fragmenting into extended families and relocating in "heartland" communities until some adjustment can be made (ARVELLO-JIMENEZ 1971). Once an adjustment has occurred, these genetically disparate communities break up, and the original families return to their previous locations. Inevitably, these periods of retrenchment lead to genetic exchange among more distantly related villages. Genetic exchange between Yanomama villages is largely dictated by the shifting political needs of intratribal warfare (CHAGNON 1968, 1974). While such alliances are in a constant state of flux, and do extend over cluster "boundaries" on occasion, genetic exchange still tends —on the average—to occur more often within than between clusters.

From the foregoing, it seems *a priori* probable that Yanomama clusters should be more disparate genetically than their Ye'cuana counterparts. The relative amounts of variation among villages, irrespective of cluster membership, are a bit harder to predict. We now turn to a brief description of the villages and clusters utilized here.

The Ye'cuana

Between 1966 and 1969, seven Ye'cuana villages were sampled. These villages

[10A, 10BD, 10C, 10F, 10G, 10HI] are of the usual Ye'cuana type and have been described by WARD and NEEL (1970). At the time the remaining Ye'cuana were sampled (1971), members of five additional villages were scattered in six temporary camps, as a consequence of political differences with the Venezuelan government. These camps were typical of the Ye'cuana retrenchment phase, described above. Because of the difficulty of sorting these people, allelic frequencies were routinely reported for the camps (TANIS *et al.* 1974; WARD *et al.* 1975). We have managed (by use of detailed pedigrees) to determine the source village of most of the individuals in these disparate collections and we therefore use the original (reconstituted) villages here. We assign them accession codes [10T, 10U, 10V, 10W, 10X] to avoid confusion with the earlier literature, where the camps are denoted with codes [10L, 10M, 10N, 10P, 10Q, 10S].

The twelve sampled villages [10A–10HI, 10T–10X] can be assigned to four clusters on the basis of historical origin and present-day socio-political relationships. These four clusters are described below, while the component villages are located in Figure 1.

Ashishi cluster: The ancestral village was Ashishi'ña, located in the watershed between the Cuntinamo and Ventuari Rivers in 1920; this village subsequently moved, in response to intrusive pressures. The four villages in our sample are derived from events occurring between 1945 and 1960. Cha'jura'ña (10E) is located on a tributary on the Cuara, having moved into this area in the past 20 years. Curaawa (10W) is located in a tributary of the Ventuari, and is an offshoot of what is now Aquencwa (10X), and acculturated village on the middle Ventuari. Juramato Cana (10V) is further down-river, and represents an earlier split. There has been a considerable shuffling of individuals among these villages since their formation, in spite of the fact that Cha'jura'ña is physically distant from the others.

Tacameña cluster: The four villages in our sample derive from a complex of villages located near a site called Tacameña between 1890 and 1910. The two largest villages, Jööwötö'ña (10BD) and Wasai'ña (10C), occupy the upper Cuara, and have long enjoyed a close relationship. The other two villages, Wajüna'ña (10T) and Tawayu'ña (10U), are located on the upper Ventuari, and have had little contact with the outside world. There is nevertheless a steady flow of migration between the Cuara and Ventuari villages, in terms of shifting residence patterns and marriage.

Wacamu'ña cluster: This cluster is centered predominantly in the Cunucunuma basin, having migrated down from the headwater region of the Cunucunuma and Cuntinamo rivers since about 1920. The two villages in our sample, Wede'ña (10G) and Acana'ña (10HI), are the largest in the group, and are the lowest on the Cunucunuma. Although both have had a fair amount of contact with the outside world, their predominant ties are with the other villages of the Wacamu'ña cluster and to a lesser extent with the Ashishi cluster.

Merevari clusters: This is a rather arbitrary cluster, since the relationship between the two sample villages is ill-defined. Sharama'ña (10F) and Yevarejuri'ña (10A) are thought to have arisen between 1910 and 1920 on a tributary of the

Merevari River, which drains into the upper Cuara. Sharama'ña currently lies on the Padamo River (in Venezuela), while Yevarejuri'ña occupies a site on the Auris River (in Brazil). The latter has had a very checkered history (see CHAGNON *et al.* 1970), and probably has had minimal contact with its antecedents.

The Yanomama

In order to provide a "fair" comparison for the Ye'cuana clusters, defined above, we have chosen four Yanomama clusters that share a common dialect and have developed over the same time span. These are not the most disparate of Yanomama clusters. At the geographic extremes of the tribe, clusters have been separated for a longer time span, and differ by cultural and linguistic features (SPIELMAN, MIGLIAZZI, and NEEL 1974). These more disparate groups might legitimately be considered "incipient" tribes. Since the Ye'cuana show no such pronounced degree of fragmentation, we have restricted attention to a set of four Yanomama clusters of "comparable" cultural diversification and similar time depth. These four clusters are discussed at length in WARD (1972), and we merely list them here. The villages are again indicated in Figure 1.

Shamatari cluster: We have chosen four villages from a larger set, namely Reyabobowei (03H), Ironasi (11G), Mowaraoba (11HI), and Iwahikoroba (11YZ) so as to make this cluster comparable with the *Ashishi.*

Namoweitari cluster: We have again chosen four villages from a larger set to make this group roughly equivalent to the *Tacemeña*: Ora (03A), Kora (03B), Monou (03C), and Patanowa (08ABC).

Ocamo cluster: Here we have chosen a pair of villages, Wabutawa (08K) and Wabarabro (08L), which have migrated down from the headwaters of the Ocamo River. We view this pair as somewhat comparable to the *Wacumu'ña.*

Wanaboweitari cluster: We have chosen two villages, Makorima (08N) and Kashorawa (08S) from this larger cluster. The internal relationships of this cluster and particularly the relationships between these two villages, are somewhat ambiguous (WARD 1972), and the situation may be viewed as analogous to that of the two *Merevari* villages.

We therefore have a set of 12 villages in four clusters within each tribe. By using the same number of villages in each cluster we have not only simplified the comparisons which follow but have removed an confounding effects due to an unbalanced design. Any difference in the relative apportionment of genetic diversity in the two tribes will be a function of their different demographic and socio-cultural patterns.

SINGLE-LOCUS ANALYSIS

Consider a set of I populations, each of which is segregating for a number of loci. The various alleles at the different loci are packaged into multiple-locus gametes, and these gametes are combined into multiple-locus zygotes. Since the zygotic array of a population is almost entirely determined by the gametic array, at least in outcrossing sexual species, a multiple-locus gamete is the real unit of interest. It is nevertheless convenient to begin with the single locus-analysis.

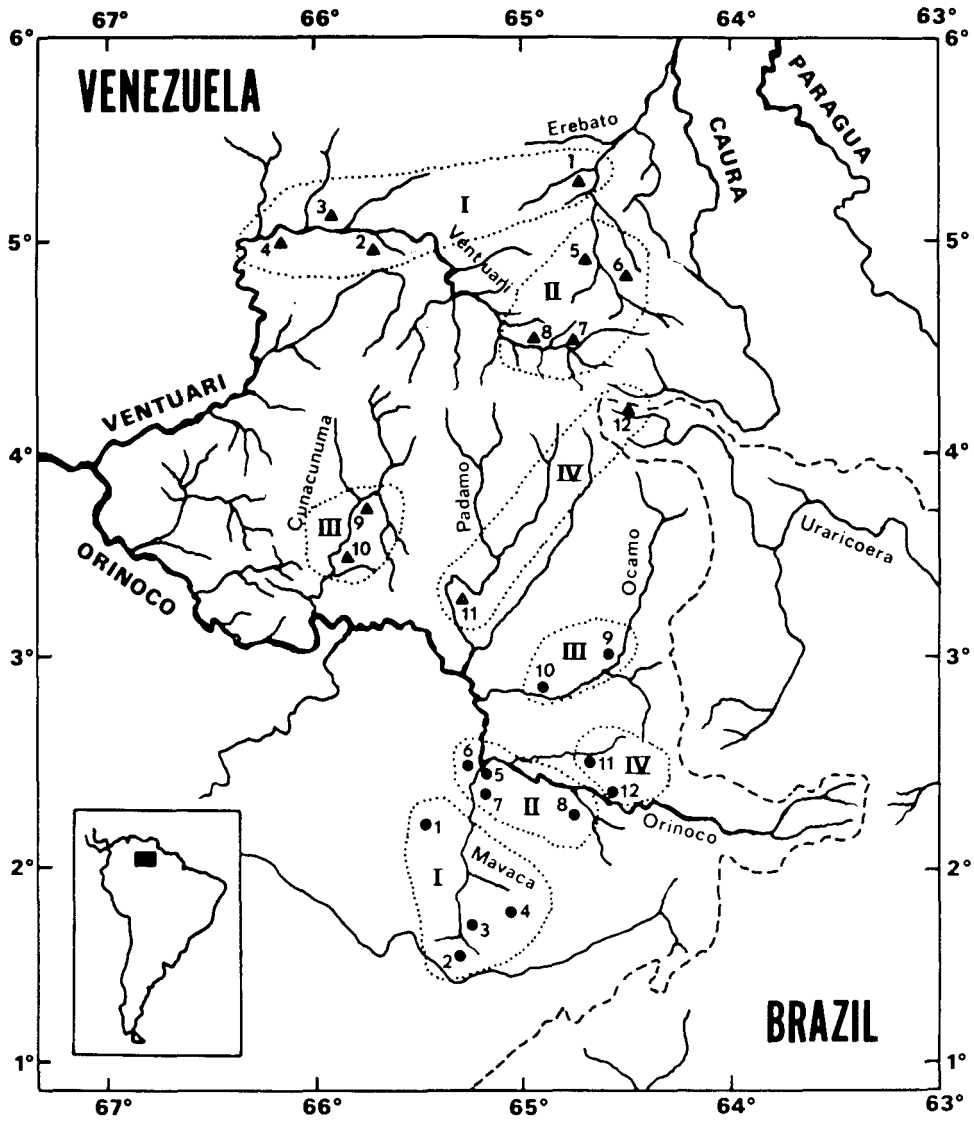

FIGURE 1.—Map showing distribution of the Yanomama and Ye'cuana clusters analyzed in this paper. Ye'cuana villages are denoted by ▲ and Yanomama villagesa by ●. Only villages incorporated into this analysis are indicated. Most of the Yanomama clusters displayed here contain additional villages, while at least 5 other Yanomama clusters can be recognized (WARD *et al.* 1977).

The villages and their cluster affiliations are as follows: *YE'CUANA* I. *Ashishi Cluster*: 1. Cha'jura'ñn; 2. Curaawa; 3. Aguencwa; 4. Juramato Cana. II. *Tacameña Cluster*: 5. Jööwötö'ña; 6. Wasai'ña; 7. Wajüna'ña; 8. Tawayu'ña. III. *Wacamu'ña Cluster*: 9. Wede'ña; 10. Acana'ña. IV. *Merevari Cluster*: 11. Sharama'ña; 12. Yevarejuri'ña. *YANOMAMA* I. *Shamatari Cluster*: 1. Reyabobowei; 2. Ironasi; 3. Mowaraoba; 4. Iwahikaroba. II. *Namoweitari Cluster*: 5. Ora; 6. Koro; 7. Monou; 8. Patanowa. III. *Ocamo Cluster*: 9. Wabutawa; 10. Wabarabro. IV. *Wanoboweitari Cluster*: 11. Makorima; 12. Kashorawa.

Likelihood Analysis of Variation

Given a sample of $(2N_i)$ alleles from the ith population, we wish to draw inference on the population frequencies (P_{ij}) of the alleles from the observed numbers (X_{ij}) of these alleles in the samples. The multinomial likelihood function for all I populations is given by

$$L(P|X) = \prod_{i=1}^{I} L_i \ (P|X) \propto \prod_{i=1}^{I} \prod_{j=1}^{J} P_{ij}^{X_{ij}} \tag{1}$$

The upper index J should here be understood to represent the total number of alleles encountered in the I populations.

The analysis is initiated by constructing a set of test criteria to compare various hypotheses concerning the unknown parameters (P_{ij}). By formulating the population comparisons in this framework, it is possible to extend the sort of questions which may be addressed. The most reasonable null hypothesis Ω_o is that the allelic frequencies are identical in all populations. The most general alternative hypothesis Ω_U states that there are one or more unspecified differences in allelic frequencies among populations.

$$\begin{aligned} &\Omega_o\colon P_{1j} \equiv \ldots \equiv P_{Ij} \equiv P_{.j} \quad j = 1, \ldots, J \ . \\ &\Omega_U\colon P_{ij} \neq P_{.j} \quad i = 1, \ldots, I \quad j = 1, \ldots, J \ . \end{aligned} \tag{2}$$

The generalized alternative Ω_U is an all-embracing departure from the null Ω_o, and is often a confession of ignorance.

The MLE (maximum likelihood estimates) of the P_{ij} for the two hypotheses are given by

$$\begin{aligned} &\Omega_o\colon \bar{P}_{.j} = \sum_{i=1}^{I} X_{ij} \div 2 \sum_{i=1}^{I} N_i \quad j = 1, \ldots, J \\ &\Omega_U\colon \tilde{P}_{ij} = X_{ij} \div 2N_i \quad i = 1, \ldots, I \quad j = 1, \ldots, J \ . \end{aligned} \tag{3}$$

The null hypothesis may be tested against the generalized alternative by recourse to

$$\Lambda_{oU} = 2\left[\sum_{i=1}^{I} \sum_{j=1}^{J} X_{ij}(\operatorname{Log} \tilde{P}_{ij} - \operatorname{Log} \bar{P}_{.j})\right] \tag{4}$$

where all logs are to base (e). The criterion Λ_{oU} is asymptotically distributed as χ^2 with $(I-1)$ $(J-1$ degrees of freedom (see KULLBACK 1968). If Λ_{oU} is not significantly different from zero, the null hypothesis cannot be rejected, and one should normally terminate the analysis. For interesting population problems, Λ_{oU} will usually be significantly different from zero, and we may proceed to more elaborate analyses.

Since $X_{ij} = 2N_i\tilde{P}_{ij}$ and since $\sum_{i=1}^{I} X_{ij} = 2\bar{P}_{.j} \sum_{i=1}^{I} N_i$, equation (4) reduces to the form

$$\Lambda_{OU} = 4N[\sum_{i=1}^{I} f_i \sum_{j=1}^{J} (\tilde{P}_{ij} \operatorname{Log} \tilde{P}_{ij} - \bar{P}_{.j} \operatorname{Log} \bar{P}_{.j})] \ , \tag{5}$$

where $2N = \sum_{i=1}^{I} 2N_i$ and $f_i = (2N_i/2N)$. We could then use $H_{OU} = (\Lambda_{OU}/4N)$ as a sample size independent metric of population dispersion or a measure of "distance" among a set of populations. Λ_{OU} is an increasing function of sample size (N), and should not be used in this fashion; it should always be used for testing. If Λ_{OU} is not significantly different from zero, we would normally terminate the analysis. If Λ_{OU} is significant, or if there are good *a priori* reasons for doing so in any event, we would then proceed to more elaborate analyses.

The apportionment of variability: hierarchical analysis

In the present situation we wish to extend the analysis to determine the relative apportionment of genetic variation attributable to cluster differences and that attributable to differences among villages within clusters. This can be readily formulated in a hypothesis testing framework by extending the above model as follows. For each tribe we have a sample of $I = 12$ villages distributed among four clusters as follows: $I_A = 4$ villages from cluster A; $I_B = 4$ villages from cluster B; $I_C = 2$ villages from cluster C; $I_D = 2$ villages from cluster D. In addition to Ω_O and Ω_u, we may specify an intermediate hypothesis

$$\begin{aligned}
\Omega_1\colon\ & P_{1j} \equiv \ldots \equiv P_{4j} \equiv P_{Aj} \\
& P_{5j} \equiv \ldots \equiv P_{8j} \equiv P_{Bj} \\
& P_{9j} \equiv \ldots \equiv P_{10j} \equiv P_{Cj} \qquad j = 1, \ldots, J \\
& P_{11j} \equiv \ldots \equiv P_{12j} \equiv P_{Dj} \\
& P_{Aj} \not\equiv \ldots \not\equiv P_{Dj} \not\equiv P_{.j} \ .
\end{aligned} \tag{6}$$

As before, the MLE of the P_{ij} are given by (3) for Ω_O and Ω_U. Under Ω_1, the corresponding estimates are seen to be

$$\begin{aligned}
\hat{P}_{Aj} &= \sum_{i=1}^{4} X_{ij} \div 2 \sum_{i=1}^{4} N_i \qquad \hat{P}_{Cj} = \sum_{i=9}^{10} X_{ij} \div 2 \sum_{i=9}^{10} N_i \\
\hat{P}_{Bj} &= \sum_{i=5}^{8} X_{ij} \div 2 \sum_{i=5}^{8} N_i \qquad \hat{P}_{Dj} = \sum_{i=11}^{12} X_{1j} \div 2 \sum_{i=1}^{12} N_i \\
& \qquad j = 1, \ldots, J.
\end{aligned} \tag{7}$$

The criterion Λ_{OU}, given by (5), may be partitioned into a test of the group differences, Λ_{O1}, and a test of the differences within groups, Λ_{1U},

$$\begin{aligned}
\Lambda_{OU} &= 4N\Big[\sum_{i=1}^{I} f_i \sum_{j=1}^{J} (\hat{P}_{ij} \operatorname{Log} \hat{P}_{ij} - \bar{P}_{.j} \operatorname{Log} \bar{P}_{.j})\Big] \\
&\quad + 4N\Big[\sum_{i=1}^{I} f_i \sum_{j=1}^{J} (\tilde{P}_{ij} \operatorname{Log} \tilde{P}_{ij} - \hat{P}_{ij} \operatorname{Log} \hat{P}_{ij})\Big] \\
&= \Lambda_{O1} + \Lambda_{1U}.
\end{aligned} \tag{8}$$

The criterion Λ_{01} is asymptotically χ^2 distributed, and has 3 $(J-1)$ degrees of freedom, while Λ_{1U} is asymptotically χ^2 with $(I_A+I_B+I_C+I_D-4)\ (J-1) = 8(J-1)$ degrees of freedom. This latter may be further partitioned into

$$\begin{aligned}\Lambda_{1U} &= 4\sum_{i=1}^{4} N_i\Big[\sum_{j=1}^{J}(\tilde{P}_{ij}\operatorname{Log}\tilde{P}_{ij}-\hat{P}_{Aj}\operatorname{Log}\hat{P}_{Aj})\Big]\\ &+4\sum_{i=5}^{8} N_i\Big[\sum_{j=1}^{J}(\tilde{P}_{ij}\operatorname{Log}\tilde{P}_{ij}-\hat{P}_{Bj}\operatorname{Log}\hat{P}_{Bj})\Big]\\ &+4\sum_{i=9}^{10} N_i\Big[\sum_{j=1}^{J}(\tilde{P}_{ij}\operatorname{Log}\tilde{P}_{ij}-\hat{P}_{Cj}\operatorname{Log}\hat{P}_{Cj})\Big]\\ &+4\sum_{i=11}^{12} N_i\Big[\sum_{j=1}^{J}(\tilde{P}_{ij}\operatorname{Log}\tilde{P}_{ij}-\hat{P}_{Dj}\operatorname{Log}\hat{P}_{Dj})\Big]\\ &=\Lambda_{1A}+\Lambda_{1B}+\Lambda_{1C}+\Lambda_{1D}\ ,\end{aligned} \tag{9}$$

which have $(I_A-1)(J-1)$, $(I_B-1)(J-1)$, $(I_C-1)(J-1)$ and $(I_D-1)(J-1)$ degrees of freedom, respectively. We may partition H_{OU} in exactly parallel fashion

$$H_{OU}=H_{O1}+H_{1U}=H_{O1}+[H_{1A}+H_{1B}+H_{1C}+H_{1D}]\ . \tag{10}$$

ANALYTICAL COMPLICATIONS

Multiple loci

Consider a two-locus, two-allele system, with gametic genotypes $G_1=A_1B_1$, $G_2=A_1B_2$, $G_3=A_2B_1$, and $G_4=A_2B_2$. This two-locus system may be analyzed as above by simply setting $J=4$. The extension to multiple-locus gametes is obvious. In practice, however, the potential number of classes quickly exceeds the sample size. In extreme cases, one could expect each recovered gamete to be unique and that most potential gametes would not be recovered at all. The χ^2 approximation is useless in this case, and an effort to increase the numbers in each "gametic class" requires some form of judicious "lumping."

If gametic equilibrium exists within each population (SMOUSE 1974), the two-locus analysis degenerates to the sum of separate single-locus analyses, *i.e.*,

$$\begin{aligned}\Lambda_{OU}(A+B) &= 4N\sum_{i=1}^{I} f_i\sum_{k=1}^{2}(\tilde{P}_{ik.}\operatorname{Log}\tilde{P}_{ik.}-\bar{P}_{.k.}\operatorname{Log}\bar{P}_{.k.})\\ &+4N\sum_{i=1}^{I} f_i\sum_{l=1}^{2}(\tilde{P}_{i.l}\operatorname{Log}\tilde{P}_{i.l}-\bar{P}_{..l}\operatorname{Log}\bar{P}_{..l})\\ &=\Lambda_{OU}(A)+\Lambda_{OU}(B)\ ,\end{aligned} \tag{11}$$

where the dot subscripting denotes single-locus marginal frequencies. Each of these test criteria is asymptotically distributed as χ^2 with $(I-1)$ degrees of freedom. The test criteria (8) and (9) may be similarly partitioned, as may the "distance" analogoues H. Extension to multiple alleles and multiple loci is straightforward.

In the event that gametic equilibrium does not obtain within populations, it

is necessary to allow for non-independence of the various loci. Two separate treatments are possible; the choice will depend upon the situation. In the two-locus, two-allele case, one has

$$\Lambda_{OU}(A+B) = \Lambda_{OU}(A) + \Lambda_{OU}(B) + \Lambda_{OU}(AB) \tag{12}$$

where $\Lambda_{OU}(A)$ and $\Lambda_{OU}(B)$ are defined as in (12), and

$$\Lambda_{OU}(AB) = 4N \sum_{i=1}^{I} f_i \sum_{k=1}^{2} \sum_{l=1}^{2} (\tilde{P}_{ijl} \operatorname{Log} \tilde{Q}_{ikl} - \bar{P}_{.kl} \operatorname{Log} \bar{Q}_{.kl}) \ , \tag{13}$$

with

$$\tilde{Q}_{ikl} = \tilde{P}_{ijl} \div (\tilde{P}_{ik.}\tilde{P}_{i.l}) \qquad \bar{Q}_{.kl} = \bar{P}_{.kl} \div (\bar{P}_{.k.}\bar{P}_{..l}) \tag{14}$$

As before, H_{OU} admits of the analogous partition. The values of the interaction terms $\Lambda_{OU}(AB)$ and $H_{OU}(AB)$ may be either positive or negative, and these terms may be viewed as analogous to covariances. The extension to multiple alleles and/or loci is obvious.

The second treatment of the two-locus problem leads to the partition

$$\Lambda_{OU}(A+B) = \Lambda_{OU}(A) + \Lambda_{OU}(B|A) \tag{15}$$

with $\Lambda_{OU}(A)$ as defined in (11) and

$$\Lambda_{OU}(B|A) = \Lambda_{OU}(B) + \Lambda_{OU}(AB)$$

$$= 4N \sum_{i=1}^{I} f_i \sum_{k=1}^{2} \sum_{l=1}^{2} \tilde{P}_{ikl} \operatorname{Log} \frac{\tilde{P}_{ikl}}{\tilde{P}_{ik.}} - \bar{P}_{ikl} \operatorname{Log} \frac{\bar{P}_{.kl}}{\bar{P}_{.k.}} \ . \tag{16}$$

The probability ratios of (16) should be recognized as estimates of the conditional probabilities $\Pr(B_l|A_k)$ for the ith population and total, respectively. The roles of A and B may be reversed, and similar partitions are possible for $H_{OU}(A+B)$. The extension to multiple alleles and/or loci is again obvious. This particular problem (gametic disequilibrium) arises with both the Rh and MNSs complexes, and we shall have more to say about it below.

Lumping of classes

It is often convenient and/or necessary to deal with *classes* of gametes. This situation may arise from ambiguity of assay or from deliberate lumping of rare gametes into sets. To simplify discussion, consider a three-allele locus (A_1, A_2, A_3), and suppose that A_2 and A_3 are not distinguishable. Irrespective of this ambiguity, the likelihood is (ignoring the population subscript)

$$L(P|X) = \frac{(2N)!\, P_1^{X_1} P_2^{X_2} P_3^{X_3}}{(X_1)!\,(X_2)!\,(X_3)!} \tag{17}$$

Since only A_1 and $\bar{A}_1 = A_2 + A_3$ are perceived, we construct

$$L^*(P|X) = \frac{(2N)!\,P_1^{X_1}(1-P_1)^{N-X_1}}{(X_1)!\,(N-X_1)!} = \frac{(2N)!\,P_1^{X_1}(P_2+P_3)^{X_2+X_3}}{(X_1)!\,(X_2+X_3)!}$$

$$= \sum_{X_2=0}^{2N-X_1} L(P|X) \tag{18}$$

The estimates of P_{il} and $(1-P_{il}) = P_{i2}+P_{i3}$ obtained from $L^*(P|X)$ are precisely the same as those obtained from $L(P|X)$, under all hypotheses; we cannot separately estimate P_{i2} and P_{i3}. The impact of the pooling of gametic types on Λ_{OU} is assessed by partitioning as follows

$$\begin{aligned}\Lambda_{OU} = 4 \sum_{i=1}^{I} N_i \,[&\tilde{P}_{i1} \operatorname{Log} \bar{\tilde{P}}_{i1} + (1-\tilde{P}_{i1}) \operatorname{Log} (1-\tilde{P}_{i1}) \\ &- \bar{P}_{.1} \operatorname{Log} \bar{P}_{.1} - (1-\bar{P}_{.1}) \operatorname{Log} (1-\bar{P}_{.1})] \\ + 2 \sum_{i=1}^{I} (2N_i - X_{i1}) \,[&\tilde{Q}_{i1} \operatorname{Log} \bar{\tilde{Q}}_{i1} + (1-\tilde{Q}_{i1}) \operatorname{Log} (1-\tilde{Q}_{i1}) \\ &- \bar{Q}_{.1} \operatorname{Log} \bar{Q}_{.1} - (1-\bar{Q}_{.1}) \operatorname{Log} (1-\bar{Q}_{.1})] \\ = \Lambda_{OU}(P) &+ \Lambda_{OU}(Q) \ ,\end{aligned} \tag{19}$$

where

$$\frac{\tilde{Q}_{i1}}{1-\tilde{Q}_{i1}} = \frac{\tilde{P}_{i2}}{\tilde{P}_{i3}} \text{ and } \frac{\bar{Q}_{.1}}{1-\bar{Q}_{.1}} = \frac{\bar{P}_{.2}}{\bar{P}_{.3}} \ . \tag{20}$$

If $\Lambda_{OU}(P)$ is divided by $4N$ and $\Lambda_{OU}(Q)$ by $2(2N-X_{.1})$, we also have $H_{OU} = H_{OU}(P) + H_{OU}(Q)$. If A_2 and A_3 are lumped, we have only $\Lambda_{OU}(P)$ and $H_{OU}(P)$. The extension to multiple "classes," each a set of gametic types, yields the same result.

Hypergeometric sampling

The appropriateness of the multinomial likelihood function is contingent upon the assumption that the gametes sampled are randomly drawn, with replacement from the gamete pool. The genotypes sampled often represent a considerable fraction of the extant individuals in a population, and are virtually never drawn with replacement. A multi-class hypergeometric is really more appropriate than a multinomial. This fact has some generally unpleasant implications for the estimation and testing procedures described above. (This situation has similarly undesirable implications for most other statistical treatments.) The problem may be circumvented by a change in the choice of the population of inference.

Consider, a single population of gametes, G_1 and G_2 in numbers M_1 and M_2, with $2M = M_1 + M_2$. A sample of size $2N \leq 2M$ is drawn at random, but without replacement. The probability of drawing G_1 and G_2 in numbers X_1 and X_2, with $2N = X_1 + X_2$, given M_1 and M_2, is

$$h(X|M) = \frac{\binom{M_1}{X_1}\binom{M_2}{X_2}}{\binom{2M}{2N}} \quad \begin{matrix} 0 \le X_1 \le M_1 \\ 0 \le X_2 \le M_2 \end{matrix} \qquad (21)$$

$$= 0 \text{ otherwise} .$$

If we view the extant population of gametes as a sample of size $2M$, drawn at random from the potentially infinite gamete pool generated by the previous generation, the probability of obtaining G_1 and G_2 in numbers X_1 and X_2 from this two stage sampling process is

$$f(X) = \sum_{M_1=X_1}^{2M-X_2} g(M) \cdot h(X|M) = \sum_{M_1=X_1}^{2M-X_2} \binom{2M}{M_1} P_1^{M_1}(1-P_1)^{M_2} \frac{\binom{M_1}{X_1}\binom{M_2}{X_2}}{\binom{2M}{2N}}$$

$$= \binom{2N}{X_1} P_1^{X_1}(1-P_1)^{X_2} \sum_{M_1=X_1}^{2M-X_2} \binom{2M-2N}{M_1-X_1} P_1^{M_1-X_1}(1-P_1)^{M_2-X_2}$$

$$= \binom{2N}{X_1} P_1^{X_1}(1-P_1)^{X_2} , \qquad (22)$$

which is compatible with the procedures described above. The extension to multiple alleles and loci is straightforward, and yields the same conclusion.

Zygotic assay

The results above are based on the distribution of gametes. In most situations, we sample zygotes and this imposes some limitations on the generality of the analyses described. For a single codominant locus with random union of gametes, the above treatment is appropriate. Even if the assumption of random union of gametes is not justified, one obtains the same estimates of the P_{ij}. The same is true if the zygotes are not randomly sampled. The Λ and H criteria are thus reasonable *descriptions* of the variation among populations in any event. Statistical testing would require more elaborate treatment in such situations.

The situation with dominance is more complicated, but if we assume Hardy-Weinberg equilibrium, the analysis is straightforward. For a two-allele locus, the results on "lumping" can be used to obtain (ignoring the population subscript)

$$L^*(P|X) \propto \{P^2 + 2P(1-P)\}^{(X_{AA}+X_{Aa})} \{(1-P)^2\}^{(X_{aa})} , \qquad (23)$$

which yields estimates

$$\tilde{P} = 1 - \sqrt{\frac{(X_{aa})}{(2N)}} . \qquad (24)$$

The extension to multiple alleles and complex dominance relations is straightforward. If we cannot assume H-W equilibrium, it is difficult to estimate allelic frequencies for the dominance case.

RESULTS

The data

The frequencies of eleven genetic markers are listed in Tables 1 and 2 for the twelve Ye'cuana and twelve Yanomama villages. We report here only those individuals with a complete set of typings. The frequencies are listed to four decimal places for the reader's computational convenience; two significant digits are about all that are warranted for these sample sizes.

The serum albumin (*Alb*), group specific component (*Gc*), haptoglobin (*Hp*), phosphoglucomutase-1 (*PGM*), and acid phosphatase (*ACP*) loci are codominant systems. Since allelic frequencies are directly obtainable by counting for codominant loci [see (3) and (7)], we have listed same for these five loci in Table 1.

We have treated the Duffy (Fy), Lewis (Le), Kidd (Jk), and Diego (Di) blood groups as if they were all dominant loci for these tribes, because only a single antiserum was systematically employed in each case. We have listed the frequencies of the recessive phenotypes in Table 1. We know that the Hardy-Weinberg assumption is a good first approximation within a single village (NEEL and WARD 1972), and we may therefore use (24) to estimate village allelic frequencies

The Rh and MNSs complexes have been tabled as four-haplotype systems in Table 2. This is easily accomplished for the Rh system, where the "double heterozygotes" (*CDE//cDe*) and (*CDe//cDE*) may be unambiguously separated with anti-f, because all of these Indians are (*DD*-Rh positive). The haplotype constitution of the (MNSs) phenotype is unavoidably ambiguous, however, and we have estimated the haplotype frequencies in another fashion. One may partition the double heterozygotes (MNSs) into coupling (MS//Ns) and repulsion (Ms//NS) phases in such a fashion as to maximize the "two-locus" zygotic likelihood function (under the assumption of random union of two-locus gametes). Prior experience indicates that the resulting estimates are quite accurate whenever the "disequilibrium" between the two loci is large, as is known to be the case here (SMOUSE, unpublished). To simplify the analyses which follow, we shall henceforth treat these estimated haplotype frequencies as "observed," although this will involve a slight approximation.

Analysis

The analysis is done in four separate stages, reflecting the type of phenotypic data available.

Codominant loci: The problem is best handled by a nested form of analysis: villages within clusters and clusters within tribes. The hypothesis set described by (2) and (6) and the corresponding test criteria (8) and (9) are appropriate. the Λ-criteria are presented in Table 3 for each of the codominant loci. The informational measures (H) are obtained by dividing the Λ-criteria by $4N$ ($= 2788$ for the Ye'cuana, $= 3384$ for the Yanomama). (This will result in

TABLE 1

Allelic or zygotic frequencies for nine genetic loci in four clusters each of the Ye'cuana and Yanomama

Tribe, cluster, village		Allelic sample size ($2N$)	Allelic frequencies for codominant loci					Homozygous recessive frequencies for dominant loci			
			Alb-N	*Bc-1*	*Hp-1*	*PGM¹*	*ACP^b*	*Fy(bb)*	*lele*	*Jk(bb)*	*Di(bb)*
YE'CUANA											
Ashishi	10E	152	1.0000	0.9013	0.5461	0.8618	0.9474	0.0658	0.2763	0.5395	0.5921
	10V	58	1.0000	0.9310	0.7759	0.7414	1.0000	0.0345	0.3103	0.3448	0.5172
	10W	46	1.0000	0.7826	0.1957	0.7391	0.8696	0.2174	0.3043	0.6087	0.6957
	10X	30	1.0000	0.8000	0.3667	0.9667	0.8667	0.2000	0.3333	0.4667	0.5333
Tacameña	10BD	312	1.0000	0.8045	0.4872	0.8814	0.9551	0.0513	0.2372	0.4615	0.4551
	10C	136	0.9926	0.8603	0.5147	0.8603	0.8824	0.0588	0.5735	0.2794	0.2353
	10T	72	1.0000	0.7222	0.3056	0.7083	0.8611	0.0278	0.2222	0.3889	0.7500
	10U	82	1.0000	0.9878	0.6951	0.8537	0.9878	0.0732	0.2683	0.6341	0.6341
Wacamu'ña	10G	142	1.0000	0.8873	0.2394	0.7183	0.9437	0.0423	0.2958	0.4930	0.8028
	10HI	148	1.0000	0.8649	0.2905	0.7703	0.9459	0.0541	0.4595	0.6216	0.7162
Merevari	10A	128	1.0000	0.8125	0.2656	0.9766	0.9453	0.1094	0.2500	0.3750	0.9219
	10F	88	1.0000	0.6364	0.6477	0.9091	1.0000	0.0909	0.3636	0.3636	0.9545
YANOMAMA											
Namoweitari	03A	88	0.8977	0.9205	0.9205	0.8750	1.0000	0.1818	0.1364	0.3636	1.0000
	03B	54	0.9444	0.9444	0.9444	0.9074	1.0000	0.1111	0.1111	0.3333	1.0000
	03C	56	0.8929	1.0000	0.8393	0.9286	1.0000	0.0357	0.3214	0.1429	1.0000
	08ABC	330	0.9061	0.9758	0.8455	0.9424	1.0000	0.1333	0.1758	0.2061	1.0000
Shamatari	03H	142	0.9930	0.7465	0.6408	0.8732	1.0000	0.0986	0.3944	0.1408	1.0000
	11G	104	1.0000	0.7500	0.6250	0.9423	1.0000	0.0000	0.3654	0.2115	1.0000
	11HI	248	0.9839	0.9032	0.7863	0.9234	0.9960	0.0403	0.5000	0.3064	1.0000
	11YZ	200	0.9250	0.9000	0.8700	0.9200	0.9800	0.2800	0.2700	0.4200	1.0000
Wanaboweitari	08N	68	0.9118	0.9559	0.9853	0.9412	1.0000	0.0882	0.1471	0.1765	1.0000
	08S	94	0.9255	0.9149	0.8191	0.9894	1.0000	0.2128	0.4043	0.1702	1.0000
Ocamo	08K	152	0.9737	0.9079	0.8487	1.0000	0.9803	0.3590	0.2949	0.1538	1.0000
	08L	156	0.9808	0.9615	0.9103	0.9551	0.9679	0.3289	0.3553	0.1974	1.0000

TABLE 2

Haplotype frequencies for the Rh and MNSs complexes in four clusters each of the Ye'cuana and Yanomama

Tribe, cluster, village		Haplotype sample size ($2N$)	Rh — complex				*MNSs* — complex			
			R^0 *cDe*	R^1 *CDe*	R^2 *cDE*	R^z *CDE*	L^{MS} *MS*	L^{Ms} *Ms*	L^{NS} *NS*	L^{Ns} *Ns*
YE'CUANA										
Ashishi	10E	152	0.0197	0.3553	0.6250	0.0000	0.2500	0.4408	0.0724	0.2368
	10V	58	0.0862	0.2414	0.6034	0.0690	0.4310	0.3103	0.0345	0.2242
	10W	46	0.1087	0.4130	0.4783	0.0000	0.0000	0.6522	0.0435	0.3043
	10X	30	0.0333	0.4000	0.5667	0.0000	0.1333	0.6333	0.0667	0.1667
Tacameña	10BD	312	0.0192	0.4263	0.5545	0.0000	0.3814	0.4071	0.1538	0.0577
	10C	136	0.0441	0.4265	0.5294	0.0000	0.6691	0.1985	0.0809	0.0515
	10T	72	0.0000	0.2639	0.7222	0.0139	0.1389	0.4722	0.2778	0.1111
	10U	82	0.0000	0.3171	0.6829	0.0000	0.4024	0.4756	0.1220	0.0000
Wacamu'ña	10G	142	0.0282	0.4366	0.5211	0.0141	0.2465	0.3944	0.1197	0.2394
	10HI	148	0.0676	0.4189	0.4662	0.0473	0.3446	0.3108	0.1419	0.2027
Merevari	10A	128	0.0391	0.4375	0.5234	0.0000	0.1406	0.5235	0.2734	0.0625
	10F	88	0.0909	0.2841	0.5341	0.0909	0.1932	0.4773	0.2159	0.1136
YANOMAMA										
Namoweitari	03A	88	0.0000	0.8182	0.1023	0.0795	0.0114	0.5909	0.0909	0.3608
	03B	54	0.0000	0.8148	0.1111	0.0741	0.0000	0.5926	0.0185	0.3889
	03C	56	0.0000	0.8214	0.0179	0.1607	0.0000	0.5893	0.1250	0.2857
	08ABC	330	0.0000	0.7424	0.0637	0.1939	0.0152	0.5030	0.0485	0.4333
Shamatari	03H	142	0.0000	0.8239	0.0423	0.1338	0.0563	0.6479	0.1197	0.1761
	11G	104	0.0000	0.7885	0.1154	0.0961	0.0481	0.7500	0.0577	0.1442
	11HI	248	0.0000	0.7500	0.0444	0.2056	0.0282	0.7702	0.1008	0.1008
	11YZ	200	0.0000	0.7650	0.0200	0.2150	0.0050	0.7900	0.1200	0.0850
Wanaboweitari	08N	68	0.0294	0.8235	0.0000	0.1471	0.0588	0.7500	0.0735	0.1177
	08S	94	0.0213	0.7553	0.0000	0.2234	0.0957	0.3511	0.0000	0.5532
Ocamo	08K	152	0.0000	0.9276	0.0132	0.0592	0.1184	0.3750	0.0000	0.5066
	08L	156	0.0000	0.8910	0.0321	0.0769	0.2308	0.5192	0.0256	0.2244

pairwise genetic distances for the Wacamu'ña, Marevari, Wanaboweitari and Ocamo clusters, as each of these contrasts involves a single pair of villages.

Three of the given loci (*Gc*, *Hp*, PGM_1) display similar degrees of variation within both tribes. The (*Alb*) locus exhibits considerable variation within the Yanomama, but virtually none within the Ye'cuana. A single heterozygote in Ye'cuana village 10C represents rare exchange with the neighboring Yanomama. The *ACP* locus, on the other hand, is somewhat variable within the Ye'cuana, but almost fixed within the Yanomama. The amount of the among-village variation which is attributable to cluster differences varies among loci, and shows no locus-by-locus consistency across tribes.

Dominant loci: As stated above, we used (24) to estimate allelic frequencies for dominant loci within a single village. However, as the H-W assumption is seriously violated for any population unit larger than a single village, due to a considerable WAHLUND (1928) effect (NEEL and WARD 1972), the allelic frequencies for clusters are better estimated as weighted averages of single village frequencies. Given these "estimated" cluster and tribal frequencies, we have proceeded as with the codominant analyses, giving these results also in Table 3.

Complex loci: The essential feature of the two complex systems, Rh and MNS, is that the internal correlations of the set lead to considerable redundancy of information; the trick is to extract this information in the form most useful for the problem under consideration.

Inspection of Table 2 reveals that the repulsion haplotypes $R^1(CDe)$ and $R^2(cDE)$ represent the bulk of the gene pool in both tribes, but the pattern of variation is different for the two tribes. Using partitions such as those of (15)

TABLE 3

Components of variation for five codominant and four dominant loci within the Ye'cuana and Yanomama

Source of variation	Degrees of freedom	Codominant loci					Dominant loci			
		Alb	*Gc*	*Hp*	*PGM*	*ACP*	*Fy*	*Le*	*Jk*	*Di*
YE'CUANA										
Among clusters	3	1.68	20.15	53.67	42.95	4.28	11.06	4.14	13.63	116.34
Within clusters	8	2.98	46.38	97.06	29.90	37.86	16.96	40.35	26.80	45.05
Ashishi	3	0.00	7.33	40.08	12.00	12.53	13.82	0.39	5.67	2.18
Tacameña	3	2.98	30.40	24.33	12.00	17.83	2.62	32.22	18.25	40.78
Wacamu'ña	1	0.00	0.33	0.97	1.03	0.01	0.31	5.50	2.86	1.60
Merevari	1	0.00	8.32	31.68	4.87	7.49	0.21	2.24	0.02	0.49
Total	11	4.66	66.53	150.73	72.85	42.14	28.02	44.49	40.43	161.39
YANOMAMA										
Among clusters	3	32.55	57.58	39.87	20.57	18.99	80.61	51.59	14.42	0.00
Within clusters	8	23.40	42.27	59.67	21.48	7.52	146.22	35.32	39.22	0.00
Namoweitari	3	1.25	9.65	7.61	4.47	0.00	9.62	7.52	10.65	0.00
Shamatari	3	21.88	27.78	35.55	4.21	7.05	131.86	16.75	27.70	0.00
Wanaboweitari	1	0.10	1.10	13.73	3.12	0.00	4.53	10.17	0.01	0.00
Ocamo	1	0.17	3.74	2.78	9.68	0.47	0.21	0.88	0.86	0.00
Total	11	55.95	99.85	99.54	42.05	26.51	226.82	86.91	53.64	0.00

and (16), it is possible to write $\Lambda[\text{Rh}] = \Lambda[(R^0 + R^z)\ vs.\ (R^1 + R^2)] + \Lambda[(R^1\ vs.\ R^2)|(R^1 + R^2)] + \Lambda[(R^0\ vs.\ R^z)|(R^0 + R^z)]$. These three contrasts are listed in columns (1), (2) and (3) of Table 4. Within the Ye'cuana, the first contrast accounts for about half the total variation, with the second and third dividing the residual more or less evenly. For all three contrasts, $\Lambda_{01} < \Lambda_{1U}$. Within the Yanomama, the first contrast represents about half the variation, but in this case the second contrast accounts for most of the residual. In contrast to the Ye'cuana result, here $\Lambda_{01} > \Lambda_{1U}$.

An examination of the MNSs frequencies in Table 2 fails to turn up any *a priori* reason to partition in any particular fashion. [Indeed, we have computed several different partitions, none of which is outstandingly informative.] For illustrative purposes, we have partitioned along single-locus lines, as in (12), *i.e.*, $\Lambda[MNSs] = \Lambda[MN] + \Lambda[Ss] + \Lambda[MN \times Ss]$, and these three contrasts are listed in columns (4), (5) and (6) of Table 4. The major point of interest is the fact that $\Lambda[MN \times Ss]$ is generally quite substantial, and one must explicitly deal with this fact, a finding which was expected in view of our earlier comments about disequilibrium between these two markers loci. In these small populations, it is better to view haplotypes as indivisible units, because the life span of the deme is orders of magnitude less than the time required for gametic equilibrium to obtain. Overall, $\Lambda_{01}[MNSs] > \Lambda_{1U}[MNSs]$ for both the Ye'cuana and the Yanomama.

TABLE 4

Components of variation for the Rh and MNSs *complexes within the Ye'cunana and Yanomama*

Source of variation	Degrees of freedom	Rh-contrasts*			*MNSs*-contrasts*		
		(1)	(2)	(3)	(4)	(5)	(6)
YE'CUANA							
Among clusters	3	25.93	5.31	5.95	29.19	64.54	68.33
Within clusters	8	38.42	16.14	19.60	23.34	61.03	44.60
Ashishi	3	14.45	3.32	6.70	1.69	28.60	11.83
Tacameña	3	6.42	10.54	7.05	21.57	28.13	29.23
Wacamu'ña	1	5.45	0.08	0.12	0.07	4.29	0.16
Merevari	1	12.10	2.20	5.73	0.01	0.01	3.38
Total	11	64.35	21.45	25.55	52.53	125.57	112.93
YANOMAMA							
Among clusters	3	28.43	26.86	16.55	79.42	26.38	176.83
Within clusters	8	23.36	16.59	0.47	52.94	19.91	43.22
Namoweitari	3	11.29	4.89	0.00	3.06	6.62	7.01
Shamatari	3	10.59	10.36	0.00	5.38	2.96	14.45
Wanaboweitari	1	1.10	0.00	0.47	22.61	0.53	21.51
Ocamo	1	0.38	1.34	0.00	21.89	9.80	0.25
Total	11	51.79	43.45	17.02	132.36	46.29	220.05

* Contrasts
(1) $[(R^0+R^z)$ vs. $(R^1+R^2)]$.
(2) $[(R^1$ vs. $R^2)|(R^1+R^2)]$.
(3) $[(R^0$ vs. $R^z)|(R^0+R^z)]$.
(4) $[(MS+Ms)$ vs. $(NS+Ns)]$.
(5) $[(MS+NS)$ vs. $(Ms+Ns)]$.
(6) $[(MN \times Ss)$ Interaction].

Summary comparison

Now that we have illustrated various aspects of the analysis, we turn to a consideration of the question posed at the outset. How distinct are Ye'cuana clusters, compared to their Yanomama "counterparts"? We are not particularly interested at this stage in the pattern exhibited by any particular locus and will combine the eleven genetic systems. This is accomplished by summing across the columns of Tables 3 and 4. The resulting Λ criteria and the derivative measures are shown in Table 5.

The essential features of the analysis are evident from the sample size independent H-measures. The amount of village-to-village variation encountered within the two tribes is quite comparable (0.3636 for the Ye'cuana, 0.3550 for the Yanomama). The distribution of this variation, however, is somewhat different in the two tribes. Although the clusters account for an appreciable fraction of the variation within both tribes, the Yanomama clusters are rather more distinct. This is made more evident by comparing the standardized $M\Lambda$-values ($\Lambda \div df$). The within-cluster $M\Lambda$ values are almost identical for both tribes, indicating that the within-cluster dispersion is comparable. However, the among-cluster $M\Lambda$ value is almost 50% greater for the Yanomama than the comparable value for the Ye'cuana. The conclusion that the among-cluster variability is greater for the Yanomama than the Ye'cuana can be demonstrated in yet another way, by considering the ratio of the $M\Lambda$ (among clusters): $M\Lambda$ (within clusters), which is approximately distributed as an F-ratio. This is given in the last column of Table 5. The Yanomama ratio (3.36) is somewhat larger than the Ye'cuana

TABLE 5

Components of variation, pooled over eleven genetic systems (13 genetic loci), standardized $M\Lambda$-measures, informational (H)-measures, and approximate F-ratios of among-clusters to within-clusters variation for the Ye'cuana and Yanomama

Source of variation	Degrees of freedom	Likelihood test criterion (Λ)	Information measure (H)	Standardized $M\Lambda$-measure	Approximate F-ratio
YE'CUANA ($4N$=2788)					
Among clusters	45	467.15	0.1676	10.38	2.28
Within clusters	120	546.47	0.1960	4.55	
Ashishi	45	160.59	0.0576	3.57	
Tacameña	45	284.35	0.1020	6.32	
Wacamu'ña	15	22.78	0.0082	1.52	
Merevari	15	78.75	0.0282	5.25	
Total	165	1013.62	0.3636	6.14	
YANOMAMA ($4N$=3384)					
Among clusters	45	670.65	0.1982	14.90	3.36
Within clusters	120	531.59	0.1571	4.43	
Namoweitari	45	83.64	0.0247	1.86	
Shamatari	45	316.52	0.0935	7.03	
Wanaboweitari	15	78.98	0.0233	5.27	
Ocamo	15	52.45	0.0155	3.50	
Total	165	1202.24	0.3553	7.29	

ratio (2.28), conclusive evidence that differences in demographic and sociocultural factors do indeed affect the distribution of genetic variation within tribal populations. The Yanomama, who have undergone a period of rapid population growth and village fissioning, have a greater proportion of their total genetic variation attributable to cluster differences than is the case for the Ye'cuana, who have experienced repeated disruption and periodic long range migration. In this connection, it is worth recalling that the Yanomama clusters chosen for our comparison are *not* the most distinct of those available. A comparison of the Ye'cuana with these other clusters would only exacerbate the differences.

DISCUSSION

As a general proposition, the hypotheses under consideration more often concern the populations themselves than any particular genetic loci. Since some loci create analytical difficulties, it would seem that careful attention to the choice of genetic markers will expedite the analysis. Generally, two-allele codominant marker loci are preferred. In practice, of course, this may bias the outcome of the analysis, and within certain limits, it is probably best to look at as many loci as possible, dealing with ambiguities as they arise.

The sampling procedure has to be considered in evaluating the test criteria and χ^2 approximations. We have already commented on the fact that individuals are sometimes not sampled independently and may be assayed in family units, particularly for humans. The resulting inflation of the test criteria calls the nominal significance levels into dispute. In such situations, the Λ-criteria should be viewed with a cautious eye. The "F-ratios" we have presented are obviously more conservative tests than the Λ-criteria, and we prefer them on that account. Whether the same reservation concerning the sampling frame applies to other organisms is unknown, but we suspect the problem is fairly general.

The situation ie exacerbated for multiple-locus zygotic samples, even in the absence of dominance. The ambiguities encountered with multiply heterozygous classes are well known. HILL (1974) has given estimation procedures for two-locus gametic frequencies for the case of random union of two-locus gametes under a variety of dominance-codominance situations. If this assumption is relaxed or if more loci are included, the estimation problems become formidable. It may be difficult in practice to determine whether loci are in linkage-equilibrium within a population, and one may wish to assume equilibrium, in the face of ignorance. The work of SINNOCK and SING (1972), CHARLESWORTH and CHARLESWORTH (1973) and LANGLEY, TOBARI and KOJIMA (1974) suggests that this strategy is reasonable for all but closely linked loci, at least in sexually outcrossing species.

We should say a few words about analytic alternatives. Most genetic distance methods currently in use are inherently "pairwise" and are simply inadequate to deal with the range of problems discussed here. It should be possible to construct alternative formulations of a generalized hypothesis-testing sort, using extensions of certain Euclidean distance measures now in use, and it is not our

intent to claim that ours is the only possible treatment. The important point is that one should use an analytical framework sufficiently flexible to permit testing of a variety of hypotheses.

There is one very important extension of the likelihood method that we have not yet mentioned. This likelihood analysis is closely related to the clinal (regression) analyses described by SMOUSE and KOJIMA (1972) and SMOUSE (1974) and utilized by KOJIMA *et al.* (1972), TANIS *et al.* (1974) and WARD and NEEL (1976). In fact, it can readily be shown that the hierarchical analysis outlined above is a special case of these more general regression formulations. One may even extend the analysis to a consideration of the degree to which village-to-village variation is attributable to cluster formation, geographic location or admixture (WARD *et al.* in preparation). The opportunity to apply a mixed analytical strategy is only one of several advantages which may accrue from the application of these procedures to data drawn from natural populations.

We should particularly like to thank DR. NELLY ARVELLO-JIMENEZ for her extremely helpful information concerning the history of the Ye'cuana.

LITERATURE CITED

ARVELO-JIMENEZ, N., 1971 *Political relations in a tribal society: a study of the Ye'cuana Indians of Venezuela.* Unpubl. Ph.D. Dissertation, Dept. Anthrop., Cornell University. Univ. Micro. 71-18: 901.

CHAGNON, N. A., 1968 *Yanomama: The Fierce People.* Holt, Rinehart & Winston, New York. ——, 1972 Tribal social organization and genetic microdifferentiation. pp. 252–282. In: *The Structure of Human Populations.* Edited by G. A. HARRISON and A. J. BOYCE. Oxford Press, Oxford. ——, 1974 *Studying the Yanomama.* Holt, Rinehart & Winston, New York.

CHAGNON, N. A., J. V. NEEL, L. R. WEITKAMP, H. GERSHOWITZ and M. AYRES, 1970 The influence of cultural factors on the demography and pattern of gene flow from the Makiritare to the Yanomama Indians. Amer. J. Phys. Anthrop. **32:** 339–350.

CHARLESWORTH, B. and D. CHARLESWORTH. A study of linkage disequilibrium in populations of *Drosophila melanogaster.* Genetics **73:** 351–359.

FELSENSTEIN, J., 1973 Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. **25:** 471–492.

GOODMAN, M. M., 1972 Distance analysis in biology. Syst. Zool. **21:** 174–186.

GOWER, J. C., 1972 Measures of taxonomic distance and their analysis. pp. 1–24. In: *The Assessment of Population Affinities in Man.* Edited by J. S. WEINER and J. HUIZINGA. Clarendon Press, Oxford.

HEDRICK, P. W., M. E. GINEVAN and E. P. EWING, 1976 Genetic polymorphism in heterogeneous environments. Ann. Rev. Ecol. Syst. **7:** 1–32.

HILL, W. G., 1974 Estimation of linkage disequilibrium in randomly mating populations. Heredity **33:** 229–239.

KOJIMA, K., P. SMOUSE, S. YANG, P. S. NAIR and D. BRNCIC, 1972 Isozyme frequency patterns in *Drosophila pavani* associated with geographical and seasonal variables. Genetics **72:** 721–731.

KULLBACK, S., 1968 *Information Theory and Statistics.* Dover, New York.

LANGLEY, C. H., Y. N. TOBARI and K. KOJIMA, 1974 Linkage disequilibrium in natural populations of *Drosophila melanogaster*. Genetics **78:** 921–936.

LATTER, B. D. H., 1973 The estimation of genetic divergence between populations based on gene frequency data. Am. J. Hum. Genet. **25:** 247–261.

LENNINGTON, R. K. and R. H. FLAKE, 1974 Cluster analysis for biological population data. Tech. Rep. No. RHF-1 Biomed. Engin. Res. Lab., Univ. Texas.

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.

SINNOCK, P. and C. F. SING, 1972 Analysis of multi-locus genetic systems in Tecumseh, Michigan. II. Consideration of the correlation between non-alleles in gametes. Am. J. Hum. Genet. **24:** 393–415.

SMOUSE, P. E., 1974 Likelihood analysis of geographic variation in allelic frequencies. II. The logit model and an extension to multiple loci. Theoret. Appl. Genet. **45:** 52–58.

SMOUSE, P. E. and K. KOJIMA, 1972 Maximum likelihood analysis of population differences in allelic frequencies. Genetics **72:** 709–719.

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

TANIS, R. E., R. E. FERRELL, J. V. NEEL and M. MORROW, 1974 Albumin Yanomama-2, a "private" polymorphism of serum albumin. Ann. Hum. Genet. **38:** 179–190.

WAHLUND, S., 1928 Zusammensetzung von populationen und Korrelationserchernungen 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.

WARD, R. H., H. GERSHOWITZ, M. LAYRISSE and J. V. NEEL, 1975 The genetic structure of a tribal population, the Yanomama Indians. XI. Gene frequencies for 10 blood groups and the ABH-Le secretor traits in the Yanomama and their neighbors; the uniqueness of the tribe. Am. J. Hum. Genet. **27:** 1–30.

WARD, R. H. and J. V. NEEL, 1970 Gene frequencies and microdifferentiation among the Makiritare Indians. IV. A comparison of a genetic network with ethnohistory and migration matrices; a new index of genetic isolation. Am. J. Hum. Genet. **22:** 538–561. ——, 1976 The genetic structure of a tribal population, the Yanomama Indians. XIV. Clines and their interpretation. Genetics **82:** 103–121.

Corresponding editor: R. W. ALLARD