Patterns of Differentiation Between Human Local Populations

N. Jardine

Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click **here**

[1]

PATTERNS OF DIFFERENTIATION BETWEEN HUMAN LOCAL POPULATIONS

By N. JARDINE

King's College, Cambridge

CONTENTS

Quantitative methods for describing infraspecific differentiation are introduced and applied to three questions about the nature of human differentiation. The first question is whether or not human local populations are differentiated consistently by different attributes; the second is whether or not human local populations are differentiated into geographical races; the third is whether or not human local populations show clinal differentiation. The anthropometric data used were published by Mahalanobis, Majumdar & Rao (1949), and Hiernaux (1956).

INTRODUCTION

One of the main goals of the study of the physical diversity, both of human local populations and of local populations of other species, is the discovery of ways in which interaction of genetic, historical, ecological, and geographical factors has determined the observed patterns of differentiation. A prerequisite for the formulation and testing of such explanations of human differentiation is precise and unambiguous description of what is to be explained. Biologists and physical anthropologists cannot hope to explain differentiation of human populations until they know how to describe it. This paper is concerned with quantitative methods for the description of patterns of differentiation. Such methods simplify complex and intractable data in ways which may suggest explanatory hypotheses and which may facilitate the testing of such hypotheses.

The study of human differentiation has long been overshadowed by the pursuit of racial classifications as ends in themselves, and by confusion between concepts of race founded on morphological and genetic differentiation and concepts of race founded on difference in habit and custom. Recently many authors (for example, Hogben 1960; Livingstone 1962; Ehrlich & Holm 1969; Hiernaux 1969, 1971 a) have pointed out that differentiation into discrete or partially discrete geographical races and subspecies is but one of many patterns of infraspecific differentiation, and that the differentiation of human populations may not be of a kind which can well be expressed by any hierarchic classification.

In this paper I introduce and apply methods of attacking three questions about the nature of human differentiation. The first question is whether or not human populations are differentiated consistently by different attributes. The second question is whether or not the differentiation of human local populations is usefully represented by classification into geographical races. The final question is whether or not human local populations show clinal differentiation.

The plan of the paper is as follows. Section I deals with the various patterns of differentiation which may occur within the human species. All the methods offered in this paper for analysing patterns of differentiation depend upon measurement of dissimilarity between local populations with respect to sets of attributes; measurement of dissimilarity is discussed in §II. In §III I introduce and apply methods for investigating the first question, whether or not different attributes and sets of attributes differentiate local populations consistently. The methods for investigating the remaining questions, whether or not racial classification of human local populations is valid, and whether or not human populations show clinal differentiation, use as data dissimilarity measures on populations. Section IV lays down certain general constraints on methods which operate on dissimilarity measures. In §V I describe and apply methods for investigating racial classifiability of populations, and §VI deals with methods for detecting clinal differentiation amongst populations.

The ideal data for these investigations would cover as exhaustive as possible a selection of populations from a geographical area in which there was substantial human diversity. Large numbers of attributes would be recorded for adequate samples from each local population. Some of the most complete and reliable published data are due to Hiernaux (1956). He personally recorded thirty-two morphological measurements for large samples from each of fifteen local populations in Ruanda-Urundi and Kivu. In addition ABO, MN and sickle-cell anaemia allele-frequencies were recorded for several of the populations. The other data used in this study were published by Mahalanobis, Majumdar & Rao (1949), who recorded twelve morphological measurements for large samples from twenty-two tribal and caste groups, mainly from the region of India which is now called Uttar Pradesh. In both areas there is diversity of populations of a kind which has prompted their assignment to discrete races. The populations studied by Mahalanobis *et al.* have been partitioned between an 'Indo-Aryan' race and an allegedly primitive 'Dravidian' race sometimes thought to be related to Australian aboriginals. Hiernaux (1969) stated of the Congolese populations that '...in terms of classical anthropology, the group includes populations so different as to call the one "Hamiticized Bantu" and the other "pygmoid"'. If racial classification of human populations is appropriate it should be possible to analyse the information about these populations in such a way as to reveal their classifiability. I shall show that the populations are not racially classifiable, but that in both cases there is a clinal pattern of differentiation.

I. Patterns of infraspecific differentiation

In this section I give informal accounts of some of the many patterns of infraspecific differentiation. Only patterns of differentiation which are likely to occur amongst human populations are discussed. Thus such complex patterns as are found in plant populations which belong to polyploid and aneuploid series are omitted.

A single terminological preliminary is needed. Descriptors of individuals such as *eyes blue*, and *overall height 1.7 m* are called *attribute states*. Sets of attribute states such as *overall height in metres* are called *attributes*.

(i) *Populations and attributes*

The data on which descriptions of patterns of differentiation are based are records of the attribute states of individual members of samples from local populations. The first stage in any study of infraspecific differentiation is the selection of such populations. Ideally, the local populations selected should be Mendelian populations within each of which free gene-flow occurs and between which gene-flow is restricted. In practice gene-flow is usually inferred indirectly from geographical proximity, overall morphological resemblance, and with human populations, verbal reports, so that the demarcation of local populations may be somewhat arbitrary. This arbitrariness need not carry over to descriptions of patterns of differentiation provided that significance is attached only to differences between populations which are larger than occur between subpopulations of any local population. In practice, therefore, it suffices to select populations which are relatively homogeneous with respect to the attributes studied. Hiernaux (1956, 1966a) presented evidence that the Congolese populations which he studied are reasonably close approximations to Mendelian populations. Several of the populations considered by Mahalanobis *et al.* were drawn from members of a single caste from widely scattered localities (see figure 24). Hiernaux (1971a) emphasized the hazards of such pooling of local populations. Description of patterns of differentiation under circumstances where relatively homogeneous local populations cannot be selected is a problem outside the scope of this paper.

Once local populations have been selected, attribute states of individuals sampled from each population are recorded. In most studies of infraspecific differentiation the attributes encountered vary within local populations. The first stage in all the quantitative methods which I shall consider is to calculate a dissimilarity coefficient on all pairs of populations with respect to each attribute. These may be combined to obtain dissimilarity coefficients on all pairs of populations with respect to any selected set of attributes. Detailed discussion of ways of measuring dissimilarity is deferred until the next section.

(ii) *Discordance in differentiation*

The differentiation of a set of populations is said to be *discordant* when the relative magnitudes of values of dissimilarity coefficients based on different single attributes or selections of attributes differ widely. Clear examples of discordant differentiation of populations within animal species have been given by Gillham (1956), Hagmeier (1958), Highton (1962), and Sokal & Rinkel (1963). Gillham showed how chaos has arisen in the infraspecific taxonomy of certain butterflies from attempts to represent discordant patterns of differentiation by allocation to discrete subspecies.

High discordance between dissimilarity coefficients based on different single attributes is likely to occur in all infraspecific studies, and such discordance does not necessarily render meaningless dissimilarity coefficients based on many attributes. What is important is to find out whether dissimilarity coefficients based on different large sets of attributes show high discordances. If they do then dissimilarity coefficients based on many attributes may be of little significance, however large the number of attributes used, because they are unstable: that is, the relative dissimilarity values between pairs of populations are likely to be changed substantially when further attributes are considered. Likewise any classification or scaling derived from an unstable dissimilarity coefficient is likely to be substantially altered when further attributes are considered. It is only when discordances between dissimilarity coefficients based on different large selections of attributes are low that it makes sense to write of the relative dissimilarities of populations without specifying the particular list of attributes considered.

If different large sets of attributes give widely discordant dissimilarity coefficients it is of interest to find out why this is so. It may turn out that the attributes used have been poorly selected so that the list of attributes is redundant. For example, if many sizes of parts are included, all the size attributes may differentiate the populations in a way which is characteristically different from the way in which they are differentiated by other attributes, and this may reflect dependence of all the sizes of parts on a single overall size factor. Alternatively, it may turn out that different kinds of attributes differentiate the populations in characteristically different ways. For example, Briggs (1957; cited in Hiernaux 1969) suggested that certain Saharan populations are differentiated in characteristically different ways by blood-group allele frequencies, and by morphological attributes.

This rather imprecise discussion will be rendered more precise in §III where measures of discordance are suggested together with methods for describing the pattern of discordances between dissimilarity coefficients based on sets of attributes of different sizes and kinds.

(iii) *Geographical clustering*

Sometimes the relative dissimilarities of populations within a species allow partition of the populations into discrete clusters. The notion of a 'cluster' will be made more precise in §V; here it suffices to regard as clusters groups of populations which are, in terms of the values of a dissimilarity coefficient, relatively isolated from other populations. When such clusters of populations occupy distinct geographical ranges, subspecies or races may be recognized. Commonly subspecies or races are recognized also when a certain amount of intergradation takes place at the common boundaries of their ranges. In such cases clusters may not be found by methods of hierarchic clustering, but can be found by the various methods of non-hierarchic clustering described in §V which allow limited overlap between clusters. The extent to which intergradation between geographical subspecies or races can occur without invalidating their recognition is a matter of controversy. Where substantial intergradation occurs the pattern of differentiation is often better regarded as clinal (see below).

A crucial point in the discussion of human differentiation is that even when populations show consistent differentiation, and the amount of differentiation is well-correlated with the extent of geographical separation or some other barrier to gene-flow, there may be no grounds for recognition of distinct subspecies or races. This is illustrated by an artificial example. Suppose that local populations are uniformly distributed throughout an area and that the dissimilarity between each pair of populations is proportional to their geographical separation. This is a

perfect case of the pattern of differentiation described below as topoclinal. In this case no discrete clusters would be found by analysis of the dissimilarity values. Nevertheless, however the area is divided up into contiguous regions the populations from each region will be on average less dissimilar from one another than are populations from different regions. Discriminant tests between samples of inhabitants of each region, based on morphological measurements or allele-frequencies, will generally yield positive results, so that the illusion that the inhabitants of each region constitute distinct races is fostered. A concept of geographical race based on such discriminant tests is quite arbitrary since it depends on the way in which the area is subdivided. The fact that racial classifications must be based on measures of difference between popula-tions, rather than on discriminant tests between supposed races, was emphasized by Mahal-anobis (1930) and Fisher (1936). It appears that the various racial classifications of humans which have been offered by physical anthropologists likewise depend to a large extent on more or less arbitrary subdivision of the world. Examples of such classifications are those of Deniker (1900), Biasutti (1953–7), Coon (1962) and Garn (1968). Such taxonomies of mankind may result from failure to realize that geographical differentiation is not necessarily differentia-tion into races or subspecies.

(iv) *Clinal differentiation*

Differentiation is said to be clinal when the dissimilarity between populations with respect to some attribute or set of attributes tends to be larger the more widely separated are the popula-tions geographically (*topoclinal* differentiation) or by some ecological factor (*ecoclinal* differen-tiation). The term 'cline' was introduced by Huxley (1938). Clear-cut examples of ecoclinal differentiation were described by Barber & Jackson (1957) and Gregor & Watson (1961), and of topoclinal differentiation by Petersen (1947) and Mayr & Vaurie (1948). Ecoclinal differen-tiation of populations is often discordant in different attributes. Uneven sampling from a topo-cline or ecocline may mislead by suggesting that populations within a species show differentia-tion into discrete races or subspecies.

I find it convenient to extend slightly the usual definitions of clinal differentiation to include the pattern of differentiation which arises when the dissimilarities between populations tend to decrease the more extensive has been their interbreeding and consequent gene-exchange. The quantitative methods described in §VI for the investigation of ecoclinal differentiation are appropriate also for investigating this kind of differentiation, which may be called *genoclinal* differentiation.

II. Dissimilarity between populations

A dissimilarity coefficient (DC) on a set of populations P is a function d from pairs of popula-tions to the non-negative real numbers such that

$$d(a, b) \geqslant 0 \qquad \text{for all} \quad a, b \text{ in } P,$$

$$d(a, a) = 0 \qquad \text{for all} \quad a \text{ in } P,$$

$$d(a, b) = d(b, a) \quad \text{for all} \quad a, b \text{ in } P.$$

If a dissimilarity coefficient satisfies in addition the triangle inequality

$$d(a, c) \leqslant d(a, b) + d(b, c) \quad \text{for all} \quad a, b, c \text{ in } P$$

it is said to be *metric*.

Some of the various dissimilarity measures which have been used in physical anthropology were reviewed by Hiernaux (1964), and Jardine & Sibson (1971). The raw data from which a

DC is calculated are records of the attribute states possessed by each individual in a sample from each population. Two radically different approaches to the calculation of DCs have been suggested in the literature.

One approach is due to Mahalanobis (1930, 1936). The observations are used to estimate a DC on the joint distributions of each population over a set of attributes. Mahalanobis's D^2 statistic is defined only for the case where the joint distributions for each population are multi-variate normal with constant dispersion.

$$D^2(a, b) = \sum_{i,j} (x_i^a - x_i^b)(x_j^a - x_j^b) \frac{|R_{ji}|}{|R|} (-1)^{i+j},$$

where x_i^p is the mean of the ith attribute in population p ($p = a, b$), R is the covariance matrix within a population, and R_{ji} is the matrix obtained from R by deleting the jth row and ith column, so that $(|R_{ji}|/|R|)(-1)^{i+j}$ is the (i,j)th element of R^{-1}.

D is metric. It is related to the probability of correct reassignment of an individual to one of the populations, given its attribute states; and its value is bounded above by $+\infty$ in the case of complete discrimination of two populations. Generalizations of D^2 to the case of inconstant dispersion have been suggested by Ali & Silvey (1966), Chaddha & Marcus (1968), and Sibson (1969). An informal account of the use of D^2 in physical anthropology was given by Bronowski & Long (1952).

The alternative approach uses the observations to estimate DCs on the marginal distributions of the populations. If n attributes are considered n DCs are calculated and these are then combined to yield an overall DC with respect to the entire set of attributes. Many of the coefficients of this kind which have been suggested are valid only when the distributions are of some special kind. Thus the measure

$$C^2(a, b) = \sum_i \frac{(x_i^a - x_i^b)^2}{\sigma_i^2},$$

the square-root of which is often called *taxonomic distance*, is appropriate for normal distributions with the same variance in each population. It has been used widely in physical anthropolgy and is equivalent to Mahalanobis's D^2 in the case when the marginal distributions are un-correlated. Hiernaux (1965) suggested for use in physical anthropology a coefficient in which the difference between population means in each attribute is normalized by division by the observed range of its values on the set of populations studied. The K-dissimilarity coefficient introduced by Jardine & Sibson (1971) is designed to deal with the case where the distributions over attribute states shown by populations may be of arbitrary form. K-dissimilarity with respect to a single attribute between populations for which the distributions over attribute states are μ_1 and μ_2 is given by

$$K(\mu_1, \mu_2) = \tfrac{1}{2} \inf_\nu \{I(\mu_1|\nu) + I(\mu_2|\nu)\}.$$

In the above expression $I(\mu|\nu)$ is the information gain of order 1 given by

$$\int_X \log_2 \frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\, \mu(\mathrm{d}x),$$

where $(X, \mathcal{X}) = X$ is a measurable space, and μ and ν are probability measures on X with μ absolutely continuous with respect to ν. When each attribute is defined for all individuals the

K- dissimilarity value for two populations with respect to a set of attributes is obtained by adding the K-dissimilarity values for each attribute. The mathematical background to K-dissimilarity was given in Sibson (1969) and computational details for its estimation were given in Jardine & Sibson (1971).

It is an open question whether coefficients of dissimilarity which operate on joint distributions, or coefficients which are obtained by combining measures of dissimilarity on marginal distributions, should be used in describing infraspecific differentiation. Coefficients of the first kind are at present unavailable for cases in which the distributions over states of each attribute depart from normality. Further, even when the distributions are normal the correlations between attributes may not be the same in different populations so that D^2 cannot be used. In both the sets of populations studied correlation of attributes within populations is variable. Mahalanobis suggested pooling of correlations in the calculation of D^2 when dispersion is inconstant, but if this is done the justification of D^2 by its relation to probability of misidentification is lost. Such coefficients as Mahalanobis's D^2, which are related to the extent to which populations are discriminated by joint distributions, achieve an upper bound in case any attribute or combination of attributes completely discriminates a pair of populations. They are therefore inappropriate for general taxonomic use. Men are completely discriminated both from rhesus monkeys and from gorillas, but in any study of the differentiation of primates we require that a DC based on many attributes should achieve a higher value between men and rhesus monkeys than between men and gorillas.

In the study of infraspecific differentiation, where populations are often incompletely discriminated by any set of attributes, Mahalanobis's D^2 has been recommended on the grounds that it takes account of the statistical correlations of attributes, and hence eliminates redundancy amongst the attributes. This point deserves careful consideration. Two kinds of correlation between attributes are relevant to the selection of attributes as a basis for study of population differentiation. One is the statistical correlation of attributes within populations. The numerous variants of R-type factor analysis provide useful techniques for investigating the statistical correlations of attributes, and elegant examples of the use of these methods in physical anthropology are given by Howells (1951, 1957). The other, which may be called *taxonomic correlation*, arises when the relative extents to which populations are discriminated by different attributes are similar. The taxonomic correlation between two attributes over a set of populations may be measured by the discordance between the K-dissimilarity coefficients based on each.

These two sorts of correlation yield very different kinds of information about attributes. When attributes are highly correlated both statistically and taxonomically there may be reason to suspect redundancy. For example, if several slightly different versions of the cephalic index were included in a list of attributes they would be both statistically and taxonomically correlated. On the other hand, high statistical correlation accompanied by low taxonomic correlation does not imply redundancy in the list of attributes, but may imply that the selected attributes do not convey the available information as efficiently as possible. Suppose that the planar shape of some part discriminates well the selected populations. If length and breadth of the part were used as attributes they could show high statistical correlation within populations and low taxonomic correlation, and would give poor discrimination of the populations. In the extreme case where shape was constant within populations the two attributes would show perfect statistical correlation within populations, but would discriminate populations

discordantly. Replacement of the two attributes by their ratio would give better discrimination. These considerations suggest that investigation of both statistical and taxonomic correlations of attributes is a prerequisite for selection of non-redundant lists of attributes each of which efficiently discriminates the populations studied. Transformation of a set of statistically correlated
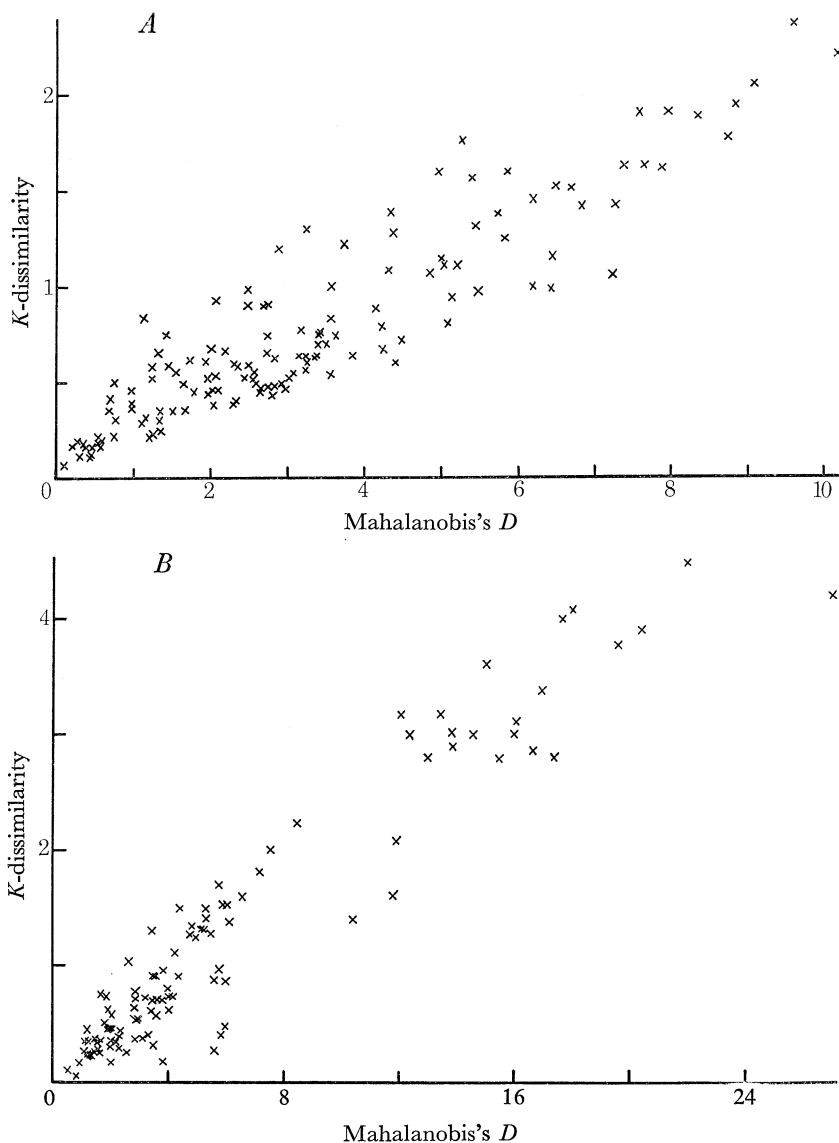

FIGURE 1. A: comparison of values of K-dissimilarity and Mahalanobis's D on twenty-three Indian populations. B: comparison of values of K-dissimilarity and Mahalanobis's D on fifteen Congolese populations. In each case D^2 was calculated using pooled covariances.

attributes into a new set of uncorrelated attributes which is the basis for calculation of Mahalanobis's D^2 is not, however, the same thing as elimination of redundancy amongst attributes.

   In the analysis of the data collected by Mahalanobis et al. and Hiernaux the K-dissimilarity coefficient has been used. Although in both cases the statistical correlations between some of the attributes are high, the relative values obtained by Mahalonobis's D and K-dissimilarity are

quite similar—as is shown in figure 1. Huizinga (1962) reported high correlations between values of C^2 and D^2 based on the same data, and Hiernaux (1964) showed that C^2 and D^2 have similar relative values on the fifteen Congolese populations described in Hiernaux (1956).

III. Discordance in the differentiation of populations

To investigate the consistency of the ways in which different attributes or sets of attributes differentiate human populations, a measure of discordance between DCs is needed. Since it is the relative rather than the absolute values of DCs which are to be compared the measure of discordance should be scale-free. In the following investigations the discordance between dissimilarity coefficients d_1 and d_2 is measured by

$$\Delta(d_1, d_2) = \Sigma \left| \frac{d_1(a, b)}{\Sigma d_1(a, b)} - \frac{d_2(a, b)}{\Sigma d_2(a, b)} \right|,$$

summation being over pairs of elements of the set of populations P. The discordance measure is a normalized modular difference bounded below by the value 0 in the case of perfect fit, and above by the value 2. The discordance between two DCs can be regarded as a measure of the taxonomic correlation of the attributes or sets of attributes on which each is based. Several other measures of discordance between DCs have been suggested; see Cormack (1971) for a survey. For example, Sokal & Rohlf (1962) proposed a measure related to the product-moment correlation coefficient. Choice of a measure is arbitrary in the absence of an appropriate statistical model for comparison of DCs.

Using a discordance measure on dissimilarity coefficients it is possible to find out whether certain groups of attributes differentiate populations in characteristically different ways: in other words, whether there are clusters of taxonomically correlated attributes. It is possible also to find out whether a DC based on a given set of attributes is unstable in the sense that it is likely to be altered substantially by consideration of further attributes.

To find out whether there are clusters of taxonomically correlated attributes K-dissimilarity coefficients on the set of populations are calculated for each attribute. The discordances between all pairs in this set of DCs are then calculated. If there are n attributes the discordance values form a half-matrix with $\frac{1}{2}n(n-1)$ entries. A cluster method is then applied to the half-matrix of discordance values to find out whether particular groups of attributes form taxonomically correlated clusters.

Discordances were calculated between K-dissimilarity coefficients on the fifteen populations studied by Hiernaux based on each of the thirty-two morphological attributes which he recorded. In figure 2 the clustering obtained by the single-link cluster method (described in detail in §V) is shown.

The hierarchic clustering reveals some well-marked clusters. The well-isolated cluster {4, 7, 11, 2, 3, 5, 6, 26, 23, 24} consists largely of lengths of parts which may be dependent upon an underlying size factor. Hiernaux (1956) calculated the average statistical correlations between attributes 2, 3, 5, 28, 32, and these are compared with their discordances in figure 3A. The fact that for these five attributes low discordance is accompanied by high statistical correlation and vice versa suggests that the list of attributes may be partially redundant. Mahalanobis et al. calculated the statistical correlations within each population between all pairs of the twelve attributes which he recorded. The average statistical correlations of these attributes

A

1. Weight
2. Stature
3. Height of antero-posterior iliac spine
4. Sitting-height
5. Total arm length
6. Humeral length
7. Radial length
8. Biacromial length
9. Transverse diameter of chest
10. Sagittal diameter of chest
11. Separation of distal crests of clavicles

12. Chest girth
13. Upper arm girth (extended)
14. Upper arm girth (flexed)
15. Thigh girth
16. Calf girth
17. Maximum glabello-occipital length
18. Maximum biparietal length
19. Bizygomatic length
20. Bigonial length
21. Internal bipalpebral length
22. External bipalpebral length
23. Nose height

24. Nose breadth
25. Nose depth
26. Face height
27. Naso-buccal height
28. Lip height
29. Mouth width
30. Ear height
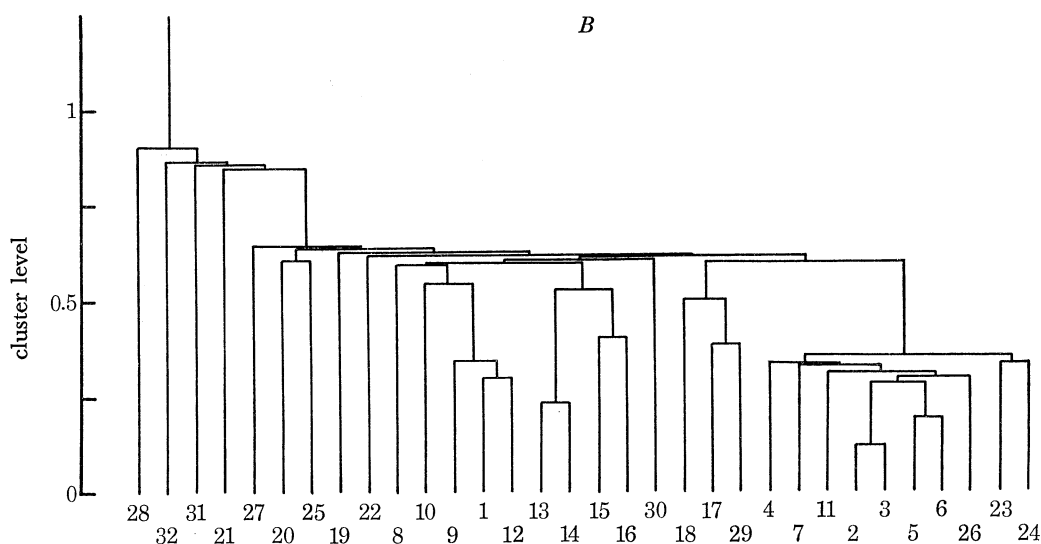31. Ear width
32. Head height

FIGURE 2. *A*: key to the thirty-two morphological attributes recorded by Hiernaux (1956). *B*: hierarchic clustering of attributes obtained by application of the single-link cluster method to the discordances between all pairs of the DCs based on each attribute. The clusters represent groups of attributes which differentiate the fifteen Congolese populations in relatively similar ways.
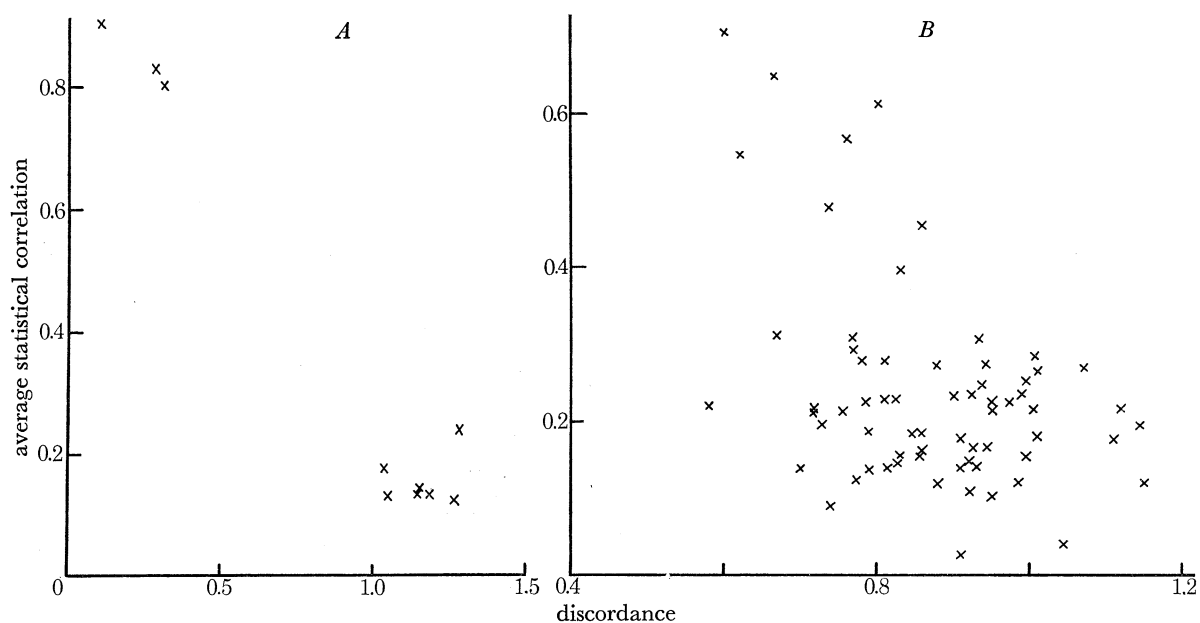

FIGURE 3. *A*: comparison between the average statistical correlations of pairs of five of the attributes recorded by Hiernaux (1956) and their discordances. *B*: comparison between the average statistical correlations of pairs of the attributes recorded by Mahalanobis *et al.* (1949) and their discordances.

within populations are compared with their discordances in figure 3*B*. In this case statistical correlation and discordance of attributes are not significantly correlated.

Hiernaux estimated ABO, MN, and Sickle-cell allele-frequencies for nine populations. *K*-dissimilarities on the nine populations were calculated for each allele-frequency and each of the morphological attributes. The discordances between DCs based on the allele-frequencies are all high (1.11 between ABO and MN, 0.69 between ABO and Sickle-cell and 1.10 between MN and Sickle-cell). It should be noted that allele-frequencies in many of the pairs of populations do not differ significantly (Hiernaux 1956, Tables X-XIII), and the corresponding *K*-dissimilarity values are close to zero.

It has been suggested that allele-frequencies provide a more objective basis for racial classification than do morphological characteristics (see, for example, Boyd 1950, 1963). However, the information given in Mourant (1956) shows that the differentiation of populations throughout the world by frequencies of different alleles is highly discordant. This suggests that it is unwise to use allele-frequencies as a basis for racial classification without first investigating the discordance of differentiation with respect to different sets of alleles. The idea that allele-frequencies provide a more objective basis for racial classification may have arisen from the fact that once geographical races have been delimited, however arbitrary the grounds, they provide more convenient diagnostic criteria than do morphological measurements.

It has been suggested that blood-group allele-frequencies may be better indicators of evolutionary relationships than are other attributes, because they are unlikely to be subject to natural selection. Cavalli-Sforza & Edwards (1967) have used dissimilarity measures on human populations based on blood-group allele-frequencies in a heuristic method for estimating the most likely evolutionary branching sequence, on the hypothesis that the observed differences are the result of genetic drift rather than selection. An evolutionary branching sequence may not be the most suitable model for population differentiation within the human species, and there is disputed evidence that at least the ABO alleles may be subject to selection (see Vogel 1970; and Wiener 1970 for details of the dispute).

Investigation of the stability of a dissimilarity coefficient is a more complex matter. First it must be noted that whether or not the values of a DC on a set of populations settle down to approximately fixed ratios as progressively more attributes are considered is entirely an empirical question. There is no justification for the view that dissimilarity between populations is a parameter which can be estimated with increasing accuracy as more attributes are considered. This view would be correct only if the attributes selected in the study of population differentiation formed random samples from some population of attributes. In practice taxonomists often select first the more easily recorded attributes which are good discriminators and then those which are less good discriminators or are less easily recorded.

Suppose that a DC whose stability is to be investigated is based on *n* attributes. Pairs of subsets of size $m \leqslant \frac{1}{2}n$ are selected at random from the set of *n* attributes. Discordances between the DCs based on these pairs of subsets are then calculated. The behaviour of the distribution of discordances as the subset size is increased gives information about the stability of the DC based on the *n* attributes.

Figure 4 shows the means and standard deviations of discordances between DCs derived from pairs of successively larger random subsets of the thirty-two morphological attributes recorded in Hiernaux (1956). The results indicate that relatively little improvement in the stability of a DC on the Congolese populations is to be expected by considering more than fifteen

2-2

morphological attributes. DCs based on different selections of fifteen out of the thirty-two attributes have mean discordance 0.228. Jardine, Davies & van Rijsbergen (unpublished) have shown that even when differentiation between well-demarcated species is considered, a mean discordance of about 0.25 between DCs based on different random selections of fifteen from a set of morphological attributes is usual. It is, however, possible that a selection of
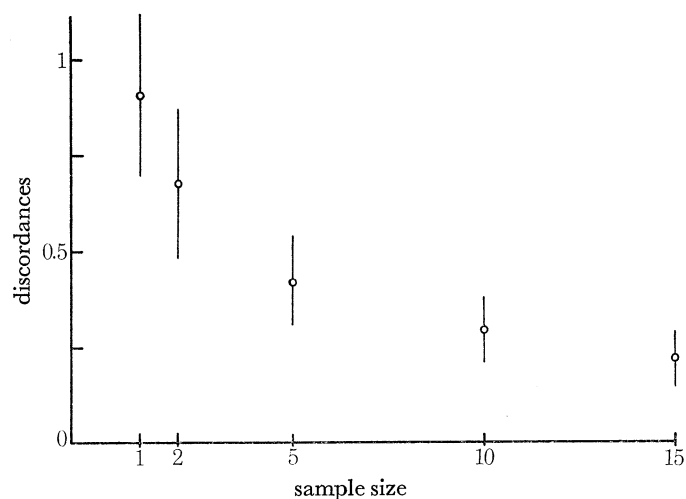

FIGURE 4. Mean discordances between K-dissimilarity coefficients based on pairs of randomly selected subsets of the thirty-two morphological attributes recorded by Hiernaux (1956). The standard deviations for pairs of selections of each size are indicated.
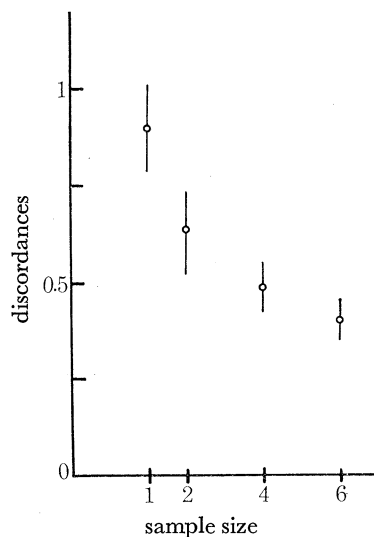

FIGURE 5. Mean discordances between K-dissimilarity coefficients based on pairs of randomly selected subsets of the twelve morphological attributes recorded by Mahalanobis *et al.* (1949). The standard deviations of the discordances for pairs of selections of each size are indicated.

attributes which included more shape factors and less size factors might embody less redundancy, and achieve greater stability.

Figure 5 shows the means and standard deviations of discordances derived from pairs of successively larger random subsets of the twelve morphological attributes studied by Mahalanobis. The mean discordance of DCs based on random selections of six attributes is 0.415, but the behaviour of the mean discordance as the attribute sample size is increased suggests that

greater stability could be achieved by selection of further attributes. The K-dissimilarity co-efficient based on the twelve attributes is substantially less stable than the K-dissimilarity coefficient based on the thirty-two attributes recorded in Hiernaux (1956). All conclusions about the pattern of differentiation of populations based on the K-dissimilarity coefficient on the twenty-three Indian populations are correspondingly less secure.

IV. Analysis of structure in a dissimilarity coefficient

A DC on a set of populations based on many attributes is a simplification of the raw observations on which it is based. As shown in the previous section, investigation of the stability of a DC is needed to determine whether or not it is a valid simplification of the information on which it is based. However, except when the number of populations is very small, it is relatively rarely that mere inspection of values of a DC suffices to determine whether any of the populations form significant clusters or whether clinal differentiation is indicated. Further simplification is usually needed to uncover such patterns of differentiation. The significance which is attached to a DC on a set of populations constrains the choice of methods for its simplification. The numerical values of a DC are never significant in isolation. Sometimes their ratios may be significant. Hence methods used for analysing DCs should be independent of scale factors. Generally it is doubtful whether much reliance can be placed even on the ratios of values of a DC, because no justification can be stated for using the DC rather than, say, its square root or its logarithm. Under these circumstances it may still be reasonable to suppose that the ordering of values of a DC is significant. Methods which operate on DCs with only *ordinal* significance must make use only of such order operations as taking maxima and minima of dissimilarity values, and must not make use of such algebraic operations as multiplication and addition (cf. Lerman 1970). Under certain circumstances even the assumption of ordinal significance may be too strong, and only the local ordering of dissimilarity values may be significant. When a DC has *local ordinal* structure it is meaningful to compare values defined on pairs of populations with a population in common. Thus $d(a, b) < d(b, c)$ is significant but $d(a, b) < d(c, d)$ may not be.

None of the methods for simplification of a DC used in this paper assumes strict numerical significance of DC values. The clustering methods and the scaling method described assume ordinal significance and the sequencing method assumes local ordinal significance.

V. Hierarchic and non-hierarchic classification of populations

The quantitative methods whereby a DC on a set of populations is transformed into a classification are generally called *cluster methods*. There are two very different kinds of cluster methods. *Simple cluster methods* (sometimes called clumping methods) seek a simple covering or partition of a set of populations. *Stratified cluster methods* seek system of clusters in which each cluster has a level, the clusters at each level except the highest being nested within clusters at a higher level. The kinds of classification obtained by simple and stratified cluster methods are shown in diagrammatic form in figure 6. Simple cluster methods have been described by Ball (1966), Jackson & Spärck Jones (1967) and Lerman (1970). Stratified cluster methods have been reviewed by Sokal & Sneath (1963), Gower (1967), Lance & Williams (1967), Cormack (1971) and Jardine & Sibson (1971).

There are reasons for using stratified rather than simple cluster methods in the study of
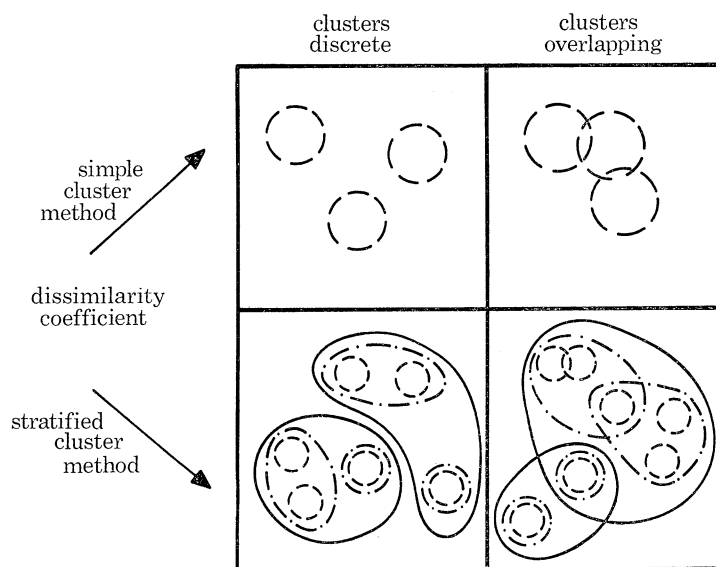
FIGURE 6. Diagrammatic representation of the results produced by
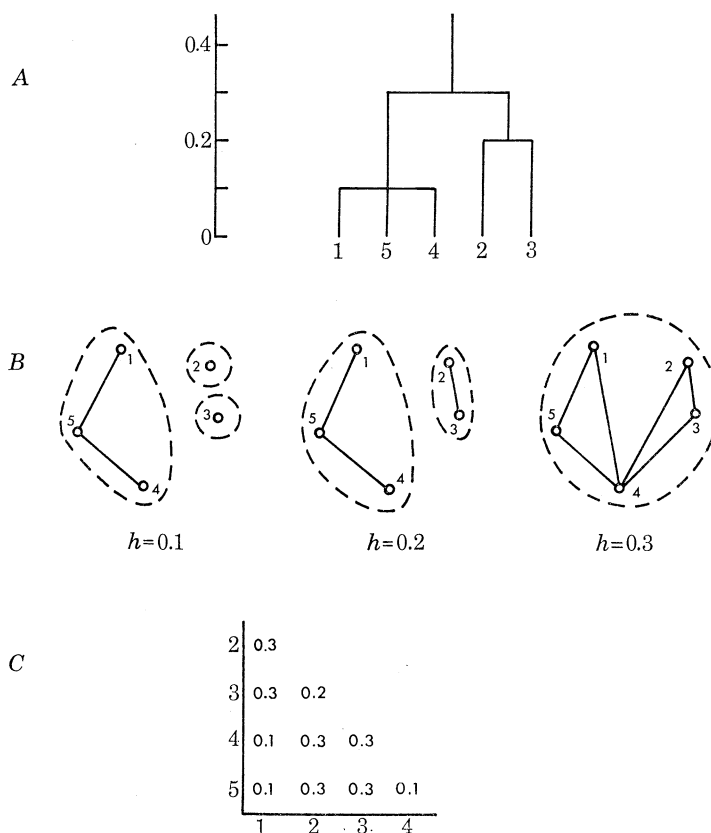different kinds of cluster methods.

FIGURE 7. Representations of a numerically stratified hierarchic clustering on five populations. A: a geometrical
representation (dendrogram). Only the levels at which populations are amalgamated into clusters are signi-
ficant; the ordering along the base-line is arbitrary. B: a graphical representation. C: a numerical representa-
tion.

infraspecific differentiation. The clusterings obtained by stratified cluster methods are more readily interpreted than are simple partitions or coverings because both the levels of clusters and their inclusion relations give information about their coherence and their isolation from other clusters. Algorithms are available for stratified clustering which produce a classification which represents a dissimilarity coefficient as accurately as possible, subject to certain constraints; criteria of accuracy of representation are discussed below. Although several criteria of goodness-of-fit of simple clusterings to a DC have been suggested in the literature, no algorithms have been developed to optimize these criteria. The available simple clustering algorithms start with an arbitrary initial partition or covering and attempt to improve the goodness of-fit to the data by successive reallocations of populations.

Hierarchic stratified cluster methods produce systems of clusters in which the clusters at each level are discrete. I shall be concerned only with hierarchic stratified clusterings in which the levels are numerical rather than, as in the taxonomic hierarchies used by biologists, merely ranked. Such clusterings may be represented in various ways. Figure 7A shows a geometrical representation, commonly called a *dendrogram*. Figure 7B shows a graphical representation. A graph for each level of the clustering is drawn with vertices representing the populations. The clusters at each level are indicated on the corresponding graphs. Some idea of the internal structure of the clusters can be conveyed by drawing in edges between just those pairs of vertices which represent populations with dissimilarity values less than or equal to the cluster level. Figure 7C shows a numerical representation of a hierarchic stratified clustering obtained by listing for each pair of populations the level h at which they cluster together.

The levels h at which populations cluster together in a numerically stratified hierarchic clustering satisfy the three conditions which define a DC:

$$h(a, b) \geqslant 0 \qquad \text{for all} \quad a, b \text{ in } P,$$
$$h(a, a) = 0 \qquad \text{for all} \quad a \text{ in } P,$$
$$h(a, b) = h(b, a) \quad \text{for all} \quad a, b \text{ in } P;$$

and also the ultrametric inequality

$$h(a, c) \leqslant \max\{h(a, b), h(b, c)\} \quad \text{for all} \quad a, b, c \text{ in } P.$$

Such clusterings are therefore characterized by *ultrametrics*. Hierarchic stratified cluster methods can be regarded as methods whereby a DC is transformed into an ultrametric DC: this characterization of hierarchic cluster methods was shown in Jardine, Jardine & Sibson (1967); Jardine & Sibson (1968a, b) to provide the key to the analysis of their properties as methods of data simplification. The main results are sketched informally below.

It is desirable that the transformation of a DC by a stratified cluster method should satisfy at least the following conditions. Conditions (a)–(c) are standard invariance requirements. Conditions (d)–(f) are special to stratified clustering methods and it is probable that for certain applications of these methods other requirements may be more appropriate.

(a) The transformation should leave unchanged data which already represents a classification of the kind sought.

(b) The transformation should be independent of any prior labelling of the populations.

(c) The transformation should be independent of scale factors.

(d) The transformation should preserve clusters which are evident in the DC. This may be imposed by requiring that a set of populations which is completely linked at a given value of

the DC should not be distributed amongst two or more clusters at the same level in the resultant clustering.

(e) Subject to conditions (a)–(d) the clustering should represent the DC as accurately as possible. The following family of measures of the accuracy with which a clustering obtained by a transformation D represents a DC d was proposed in Jardine et al. (1967):

$$\Delta_\mu(d, D(d)) = \{\Sigma|d(a, b) - D(d)(a, b)|^\mu\}^{1/\mu},$$

and it seems reasonable to interpret optimality subject to conditions (a) to (d) as minimization of the Δ_μ.
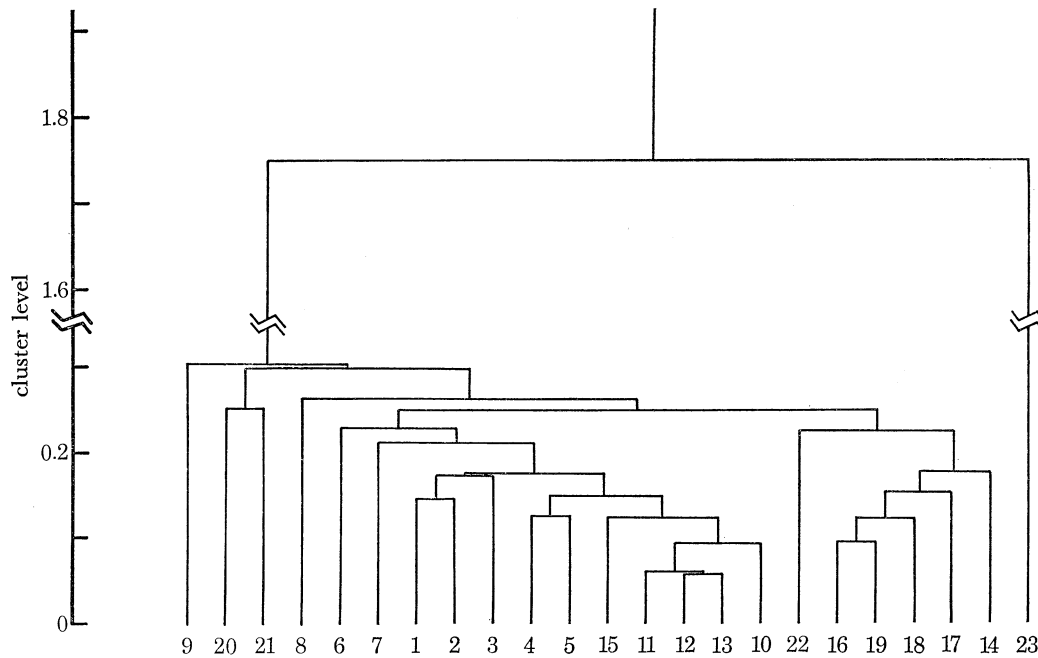

FIGURE 8. Hierarchic clustering of twenty-three Indian populations obtained by application of the single-link cluster method to DCM. For a key to the populations see table 1.

(f) The transformation should be stable in the sense that small changes in the input DC should produce commensurately small changes in the resultant clustering. This may be imposed by stipulating that the transformation from input DC to output ultrametric be a continuous transformation. Stability is crucial. If a method is unstable, sampling errors may have drastic effects and computational rounding errors may cause an algorithm to yield very different results when applied to the same data on different occasions. Further, meaningful comparison of the results obtained by applying a given method to different DCs on the same set of populations is impossible if an unstable method is used, because it is impossible to discriminate between differences in the clusterings which represent genuine differences in the structure of the data and differences which are a by-product of the instability of the cluster method.

Conditions (a)–(e) uniquely determine as a hierarchic cluster method the single-link (nearest-neighbour) cluster method originally proposed by Florek et al. (1951), which also satisfies condition (f). The method was independently suggested by Sneath (1957), who applied it to problems of bacterial classification. The single-link cluster method commutes with monotone transformations of a DC and hence can be used when a DC has only ordinal significance. The operation of the single-link method can be given a simple description in terms of graphs as

TABLE 1. *A*: KEY TO THE TWENTY-TWENTY-THREE INDIAN LOCAL POPULATIONS AND CASTE GROUPS STUDIED BY MAHALANOBIS *ET AL.* (1949); *B*: *K*-DISSIMILARITY COEFFICIENT ON THE TWENTY-THREE POPULATIONS BASED ON NINE MORPHOLOGICAL ATTRIBUTES. THIS DISSIMILARITY COEFFICIENT IS DENOTED DCM

1. Basti Brahmin
2. Other Brahmin
3. Agharia
4. Chattri
5. Muslim
6. Bhatu
7. Habru
8. Bhil
9. Dom
10. Ahir
11. Kurmi
12. Other artisan
13. Kahar
14. Tharu
15. Chamar
16. Chero
17. Majhi
18. Panika
19. Kharwar
20. Oraon
21. Rajwar
22. Korwa
23. Tharu females

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
2	0.147																					
3	0.465	0.171																				
4	0.196	0.189	0.432																			
5	0.287	0.173	0.348	0.131																		
6	0.297	0.251	0.483	0.276	0.230																	
7	0.449	0.214	0.301	0.364	0.218	0.237																
8	1.076	0.686	0.653	1.010	0.654	0.756	0.477															
9	0.360	0.370	0.584	0.576	0.404	0.306	0.411	0.765														
10	0.678	0.342	0.391	0.502	0.257	0.423	0.266	0.326	0.690													
11	0.488	0.216	0.305	0.360	0.151	0.366	0.220	0.386	0.533	0.093												
12	0.665	0.349	0.367	0.496	0.209	0.462	0.228	0.291	0.604	0.101	0.063											
13	0.793	0.445	0.446	0.576	0.265	0.628	0.321	0.331	0.777	0.153	0.117	0.062										
14	1.104	0.793	0.648	0.995	0.596	0.882	0.568	0.585	0.913	0.589	0.505	0.395	0.441									
15	1.061	0.619	0.519	0.830	0.492	0.832	0.390	0.262	0.959	0.223	0.211	0.129	0.142	0.409								
16	1.290	0.916	0.814	1.155	0.680	1.099	0.660	0.425	0.937	0.565	0.490	0.317	0.320	0.225	0.302							
17	1.619	1.186	0.947	1.512	1.020	1.455	0.952	0.518	1.269	0.772	0.749	0.518	0.487	0.403	0.430	0.160						
18	1.592	1.153	0.982	1.373	0.872	1.303	0.821	0.419	1.280	0.540	0.552	0.342	0.315	0.352	0.251	0.147	0.183					
19	1.309	0.923	0.762	1.183	0.710	1.006	0.669	0.392	0.976	0.506	0.454	0.305	0.353	0.176	0.284	0.097	0.213	0.129				
20	1.608	1.218	0.944	1.456	1.009	1.480	1.019	0.827	1.479	0.776	0.784	0.603	0.594	0.356	0.529	0.416	0.386	0.297	0.324			
21	1.663	1.207	0.924	1.357	0.954	1.258	0.880	0.806	1.536	0.575	0.655	0.500	0.556	0.499	0.419	0.579	0.714	0.392	0.382	0.255		
22	2.373	1.914	1.650	2.140	1.524	2.070	1.514	0.996	1.974	1.131	1.196	0.906	0.836	0.657	0.760	0.447	0.346	0.228	0.386	0.371	0.588	
23	4.301	3.787	3.361	3.952	3.536	3.824	3.098	2.492	4.050	2.812	3.016	2.632	2.497	2.727	2.281	2.599	2.368	1.888	2.404	1.880	1.803	1.770

3

follows. A graph is drawn for each value of the DC with vertices representing populations and with edges joining just those pairs of vertices which represent populations with dissimilarity less than or equal to the current value. The single-link method picks out as clusters the connected components of each of the graphs (see figure 7B). Several authors have objected to the way in which the single-link method may pick out straggly clusters (see, for example, Lance & Williams 1967). The various average-link and centroid methods attempt to avoid this defect, but are themselves unsatisfactory because they are unstable.

Table 1 shows the K-dissimilarity coefficient based on the nine listed morphological attributes on the twenty-three populations studied by Mahalanobis. Three of the twelve attributes which he recorded were omitted because the observations are incomplete for certain populations. This dissimilarity coefficient is denoted DCM. The dissimilarity coefficient in which population twenty-three, consisting entirely of females, is omitted is denoted DCM′. Figure 8 shows the hierarchic clustering obtained from DCM by the single-link method.

Table 2 shows the K-dissimilarity coefficient based on thirty-two morphological attributes on the fifteen populations studied by Hiernaux. This dissimilarity coefficient is denoted DCH. Figure 9 shows the hierarchic clustering obtained from DCH by the single-link method.

The way in which Mahalanobis interpreted his data is summarized in figure 10 and in the following passage from Mahalanobis et al. (1949).

> 'The D^2 values thus supply a general picture of the following kind. There are three well-demarcated clusters, the Brahmins (B-cluster) at the top of the Hindu social hierarchy; the Artisans (A-cluster) in the middle; and the tribal groups (T-cluster) at the bottom. Among the tribals, a small subcluster is formed by Cheros (16), Majhis (17), Panikars (18) and Kharwars (19); Oraons (20) and Rajwars (21) are somewhat close while Korwas (22) occupy the very bottom place furthest away from the Brahmins.
>
> 'Chattris (4) and Muslims (5) form a subcluster and together with Agharias (3) occupy a position between the Brahmins (B-cluster) and the Artisans (A-cluster) but somewhat closer to the latter.
>
> 'The two criminal tribes Bhatus (6) and Habrus (7), and Doms (9), and Bhils (8) have highly individual features of their own, and stand apart from the other groups which form a main sequence. Each of these four groups has in fact a distinct position of its own, and cannot be fitted into the general scheme determined by the three basic clusters (B), (A), and (T). Tharu (14) has also a place of its own somewhere on the border line of the T-cluster but quite distinct from the tribal groups. Finally, Chamar (15) is also a border line case separating the higher castes and tribals but nearer to the Artisans. [My reference numbers.]

The hierarchic clustering obtained from DCM appears to vindicate some of these conclusions. Bhatus (6), Habrus (7), Bhils (8), and Doms (9) form a cluster. Mahalanobis's A and B clusters are confirmed, and his suggested subcluster {16, 17, 18, 19} of his T-cluster is confirmed. Likewise, Chattris (4) and Muslims (5) cluster together. However, none of these clusters is well-isolated. Indeed, the investigation of the stability of DCM carried out in §III reveals that any of the clusters might disappear if a slightly different selection of attributes had been made. Population (23) which consists of female Tharus is very well-isolated from all the populations which consist of adult males only.

Hiernaux (1956) suggested that a group of populations forms a cluster only when all pairs of populations in the cluster are linked at some dissimilarity value and no population in the cluster
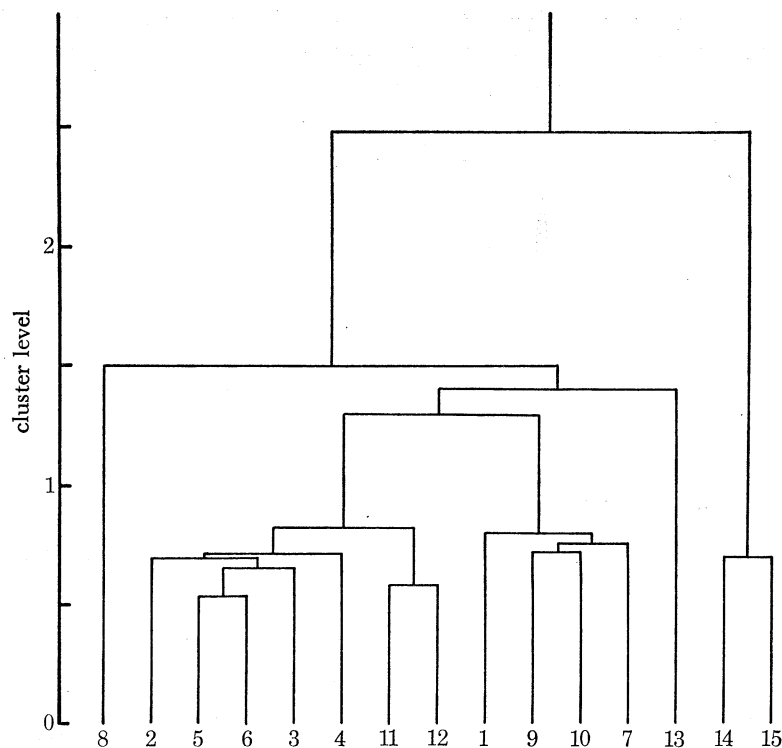
FIGURE 9. Hierarchic clustering of fifteen Congolese populations obtained by application of the single-link cluster method to DCH. For a key to the populations see table 2.

TABLE 2. A: KEY TO FIFTEEN CONGOLESE LOCAL POPULATIONS STUDIED BY HIERNAUX (1956); B: K-DISSIMILARITY COEFFICIENT ON THE FIFTEEN POPULATIONS BASED ON THIRTY-TWO MORPHOLOGICAL ATTRIBUTES. THIS DISSIMILARITY COEFFICIENT IS DENOTED DCH

1. Fulero	9. Nyanga	
2. Shi	10. Tembo	
3. Havu	11. Hutu of Ruanda	
4. Hunde	12. Hutu of Urundi	
5. Swaga	13. Twa of Ruanda-Urundi	
6. Shu	14. Tutsi of Ruanda	
7. Mbuba	15. Tutsi of Urundi	
8. Rega		

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	2.566													
3	2.717	0.966												
4	3.931	1.093	0.705											
5	2.634	0.740	0.822	0.915										
6	2.301	0.695	0.654	0.894	0.523									
7	1.299	2.295	2.028	2.815	1.936	1.982								
8	4.383	1.717	1.932	1.502	1.586	2.245	3.143							
9	1.200	1.740	1.494	2.044	1.370	1.301	0.755	2.639						
10	0.799	2.335	2.005	3.070	2.137	2.054	1.136	3.562	0.711					
11	5.727	1.700	2.886	2.197	2.534	2.497	5.065	2.400	4.688	5.457				
12	4.087	0.831	1.801	1.623	1.492	1.393	3.532	1.999	3.239	3.782	0.589			
13	1.406	3.151	3.743	4.815	3.516	3.226	1.606	5.002	2.025	1.951	5.731	4.206		
14	7.620	4.185	5.985	5.811	5.894	5.132	7.840	6.795	7.494	7.667	2.990	2.946	7.239	
15	7.052	3.758	5.475	5.554	5.468	4.573	7.562	6.519	7.040	7.063	2.910	2.481	7.044	0.699

is linked at that value to any population not in the cluster. Hiernaux's criterion selects the most homogeneous single-link clusters. Only pairs of populations, for example, 14 and 15 (Tutsi of Ruanda and Tutsi of Urundi) and 11 and 12 (Hutu of Ruanda and Hutu of Urundi) form clusters which satisfy Hiernaux's criterion. By the weaker single-link criterion several larger clusters are admitted. Again, none of the clusters is well isolated.
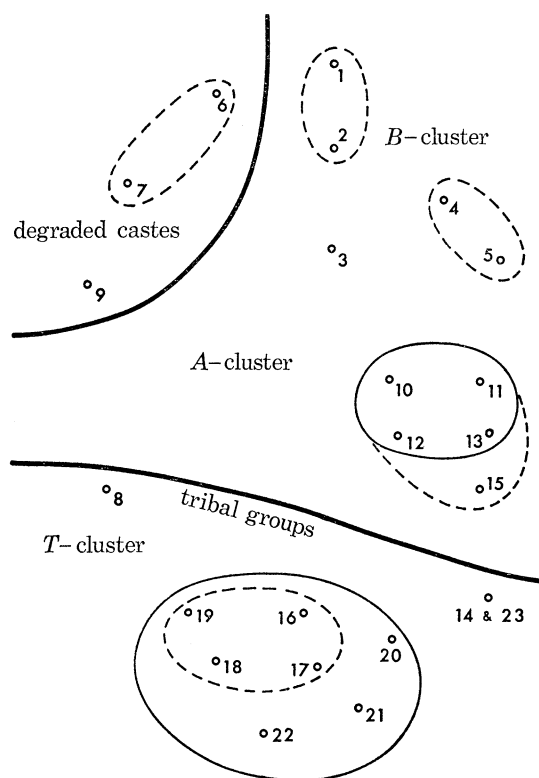

FIGURE 10. Mahalanobis's interpretation of the differentiation of twenty-three tribal and caste groups. Redrawn from Mahalanobis *et al.* (1949, Chart 8 B).

There are reasons why these results should be interpreted with extreme caution. It has been mentioned previously that subject to certain conditions the single-link method minimizes a family of distortion measures Δ_μ. In the following account

$$\hat{\Delta}_1 = \frac{\Sigma|d(a,b) - D(d)\ (a,b)|}{\Sigma d(a,b)}$$

is used as a measure of hierarchic classifiability. It is bounded below by the value 0 in case of perfect hierarchic classifiability and above by the value 1. Related measures of hierarchic classifiability have been suggested by Lerman (1969). The values of $\hat{\Delta}_1$ for DCM, DCM′ and DCH are 0.560, 0.634 and 0.532, respectively, which indicate very poor hierarchic classifiability. When distortion values are high and clusters are ill-isolated the suspicion arises that there may be no significant hierarchic structure in the DC. To test this it is necessary to estimate the probability of obtaining a distortion as least as low as the observed distortion on an appropriate null-hypothesis.

A null-hypothesis, H, which enables us to test for this kind of hierarchic structure, is that the DC is obtained by allocating the observed values at random to pairs of populations. Analytic

estimation of the probability, on this null-hypothesis, of obtaining a distortion at least as low as an observed distortion is intractable, so that a Monte Carlo method has to be used. If the observed DC is defined on p populations, new DCs can be generated by allocation of its values at random to pairs in a set of p populations. Values of the distortion imposed by single-link clustering on each of the new DCs are then calculated to obtain a sampling distribution for $\hat{\Delta}_1$. Sampling histograms for $\hat{\Delta}_1$ obtained on DCs generated from DCM′ and DCH are shown in figures 11 and 12.

The sampling distributions show that, despite the poor hierarchic classifiability, on the null-hypothesis H the probabilities of obtaining distortions as low as those observed are extremely small. Calinski & Harabasz (1971) applied a very different test for hierarchic cluster structure to the values of Mahalanobis's D^2 on the Indian populations, and concluded that there is no significant clustering of the populations.

That hierarchic clustering induces high distortion on DCM, DCM′ and DCH, suggests that some alternative method of data simplification should be used. One alternative is to use a non-hierarchic stratified cluster method. Non-hierarchic methods yield systems of more homogeneous clusters which, at the cost of increased complexity, represent a dissimilarity coefficient more accurately than is possible with a hierarchic clustering. Jardine & Sibson (1968 b) defined a sequence of stratified cluster methods B_k each of which satisfies the criteria of adequacy given above, and which requires only ordinal significance of a DC. B_1 is the single-link method; B_2 leads to a system in which clusters at a given level may have one object in common. Subsequent members of the sequence allow increasing overlap between clusters and represent the data with progressively more accuracy, until finally, in case arbitrary overlap between clusters is allowed, a perfect representation of the data is obtained. The computer programs for implementing the sequence of cluster methods B_k and for listing the clusters obtained were written by Dr J. K. M. Moody and are described in Jardine & Sibson (1971, Appendices 3 and 5). Programs for implementing B_k were described independently by Cole & Wishart (1970).

Non-hierarchic stratified clusterings cannot conveniently be represented as geometrical diagrams like the dendrograms used to represent hierarchic clusterings. They can, however, be represented by drawing the clusters at each level on the graph which represents the corresponding value of the dissimilarity coefficient. Figures 13 to 16 show the clusters at selected levels in the clusterings obtained by application of B_2 and B_3 to DCM. Figures 17 to 20 show the clusters at selected levels in the clusterings obtained by application of B_2 and B_3 to DCH. In both cases the non-hierarchic clusterings $B_2(d)$ and $B_3(d)$ achieve only a slight reduction in the distortion imposed on the DC. The distortions induced by B_1, B_2 and B_3 on DCM are 0.560, 0.504 and 0.474, respectively, and the distortions induced by B_1, B_2 and B_3 on DCH are 0.532, 0.409 and 0.330, respectively. This is in marked contrast to the situation where non-hierarchic clustering reveals well-marked cluster structure in a DC which cannot be revealed by hierarchic clustering because a small number of populations are intermediate between otherwise homogeneous and isolated clusters of populations. This situation arises typically when a DC is calculated on a set of populations which includes both populations from well-marked species and interspecific hybrid populations. It arises also when populations are chosen from well-marked geographical races or subspecies together with a few intermediate populations from localities where their ranges overlap. Non-hierarchic cluster analysis was successfully applied to such data in Jardine & Sibson (1968 a).

A striking feature of the non-hierarchic clusterings deserves notice. In both cases chains

of overlapping clusters are produced. Thus in the clustering of the Indian populations by B_2 at level $h = 0.300$ there occurs the chain $\{3, 2\} \{2, 1, 6, 4, 7, 12, 11, 5, 8, 10, 13, 15\} \{15, 14, 16, 19, 17, 18\} \{18, 20\} \{20, 21\}$. In the clustering of the Congolese populations by B_3 at level
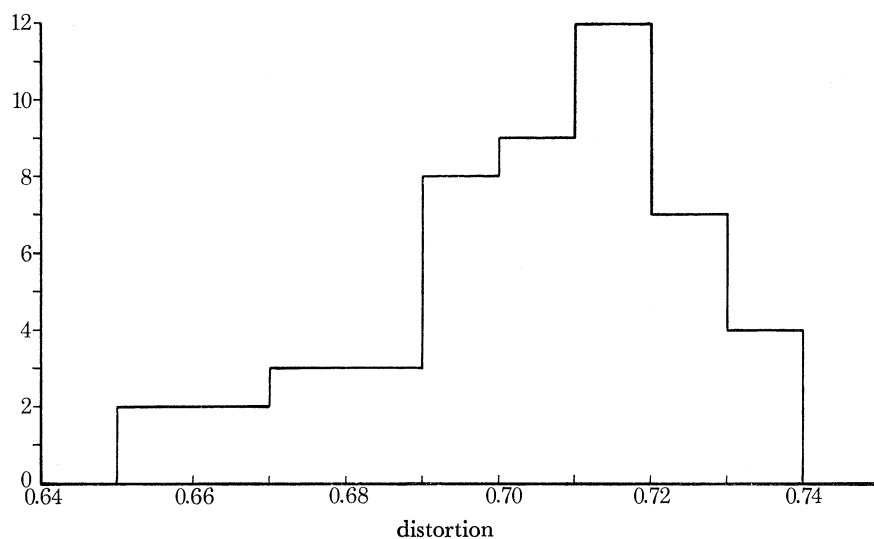

FIGURE 11. Distribution of values of $\hat{\Delta}_1$ on fifty dissimilarity coefficients generated by random allocation of the values of DCM' to pairs in a set of twenty-two populations.
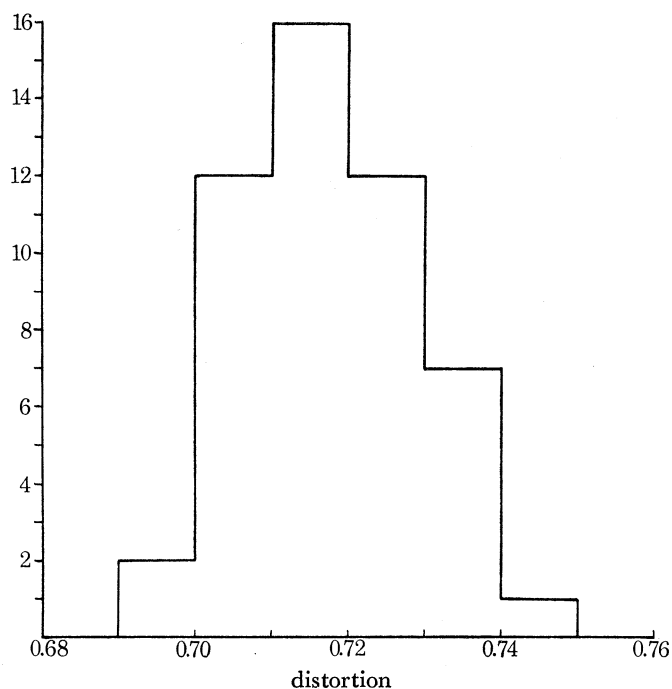

FIGURE 12. Distribution of values of $\hat{\Delta}_1$ on fifty dissimilarity coefficients generated by random allocation of the values of DCH to pairs in a set of fifteen populations.

$h = 1.70$ there occurs the chain $\{13, 1, 7\} \{1, 7, 10, 9\} \{9, 3, 6, 5, 2, 4, 12\}$. Production of chains of clusters by non-hierarchic clustering suggests that the populations show clinal differentiation, and that the apparent clustering is in fact the result of uneven sampling along the clinal

$k=2$
$h=0.160$
13

$k=2$
$h=0.300$
14

$k=3$
$h=0.160$
15

$k=3$
$h=0.300$
16
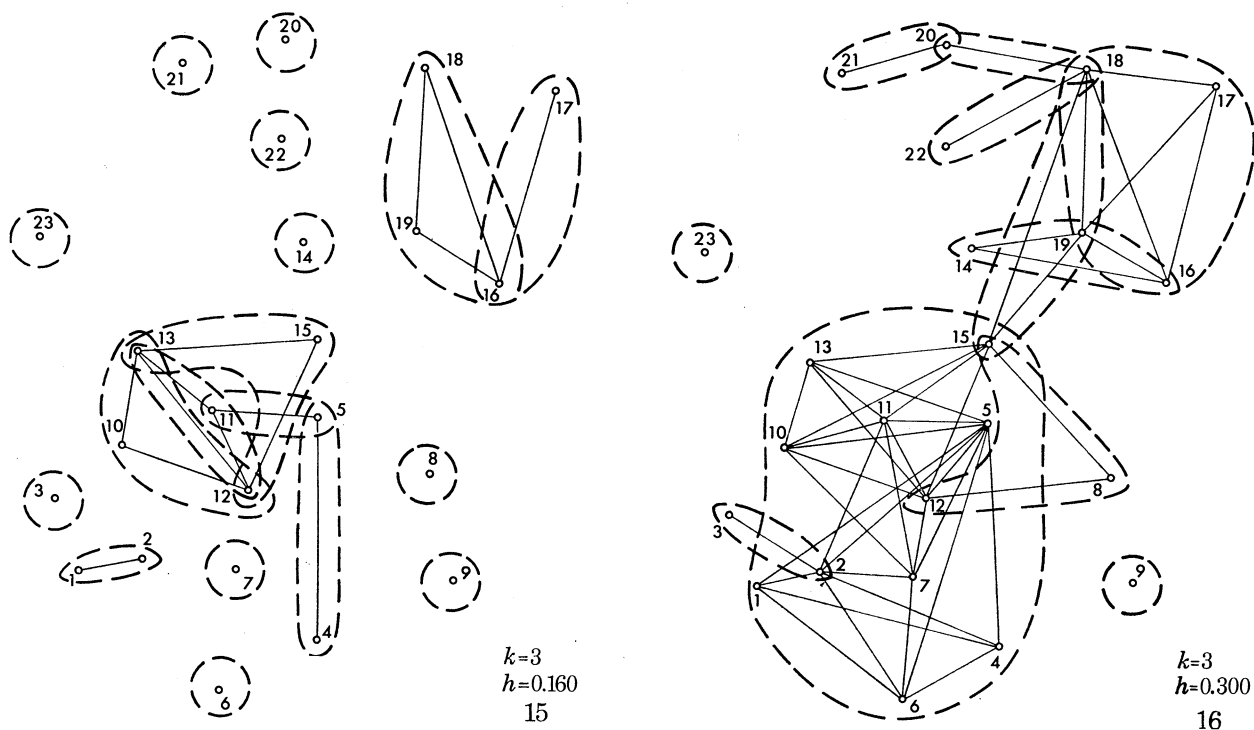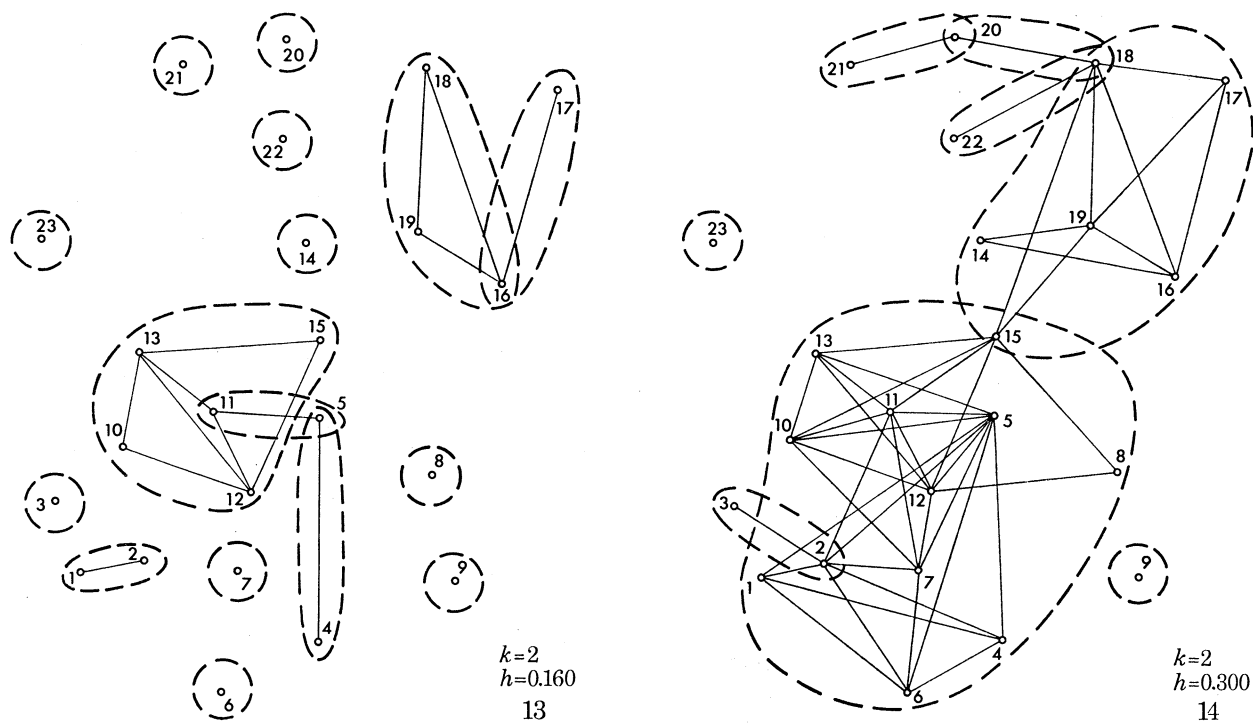
FIGURES 13 TO 15. Clusters at selected levels in the overlapping clusterings obtained by application of B_2 and B_3 (the second and third members of the sequence of cluster methods B_k) to DCM. The level, h, is indicated and edges join just those pairs of populations with dissimilarity value less than or equal to h.
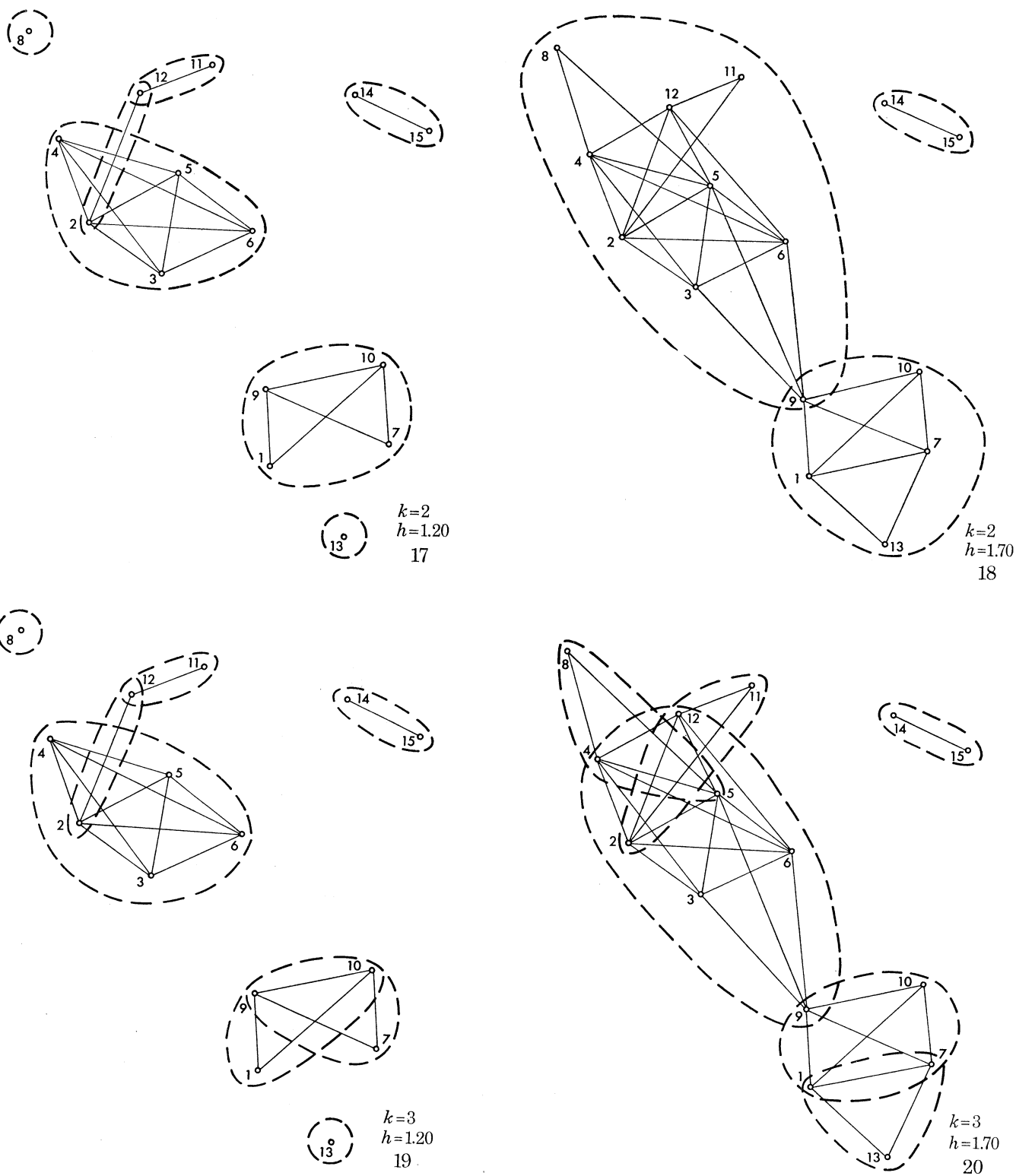
FIGURES 17 TO 20. Clusters at selected levels in the overlapping clusterings obtained by
application of B_2 and B_3 to DCH.

gradient. This possibility is consistent with the high distortion induced by all the clusterings and is investigated in detail in the next section.

VI. Sequencing and scaling of populations

At the outset it is necessary to make more precise the nature of clinal differentiation. It is useful to consider first clinal differentiation in a single attribute, say, overall height, determined by a single environmental factor, say, altitude. If the populations considered show normal distributions in overall height with approximately constant variance, the differences
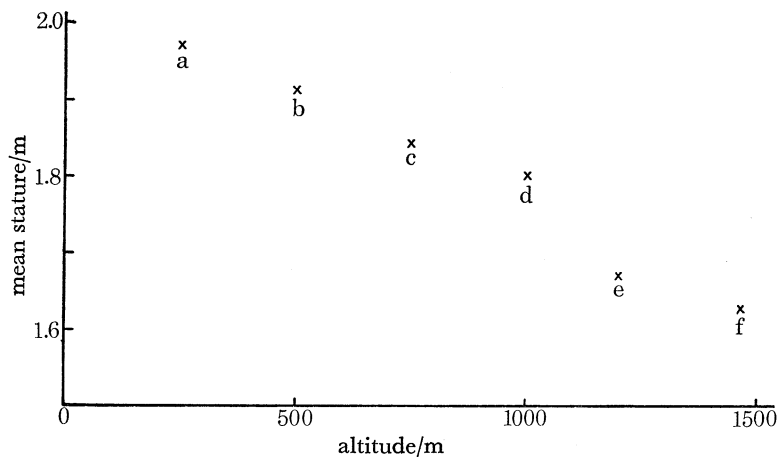
Figure 21. An imaginary set-up in which clinal differentiation of local populations in stature is determined by altitude.

between means can be used as a dissimilarity measure. The mean overall heights at different altitudes shown in figure 21 constitute a cline.

It is important to note that existence of a cline does not imply a globally monotone relation between dissimilarity values and the factor which determines the cline. Thus, in the example given, populations d and e differ in mean height more than populations b and c but their localities differ less in altitude. Postulating a cline involves only the weaker supposition that if two populations which differ by θ lie at values x and y of the clinal factor, then any pair of populations which lie at values in the open interval (x, y) should differ by less than θ.

Sequencing methods are appropriate for the investigation of one-dimensional clinal differentiation. Given a DC on a set of populations a *sequencing* method seeks that sequence of populations which minimizes some measure of distortion defined on sequences and DCs. It appears that no sequencing methods have been developed specifically for the study of clinal differentiation. Hole & Shaw (1967) gave a full account of the various sequencing methods which have been used in the closely analogous problem of chronological seriation in archaeology. None of the methods they discuss appears well-suited to the study of clinal differentiation, or to problems of chronological seriation in biology, because all use distortion measures which involve global comparison of dissimilarity values. Sibson (1971) has suggested the following measure of the distortion imposed by a sequence $A = a_1, ..., a_n$ on a DC d. λ is the proportion of inequalities

$$d(a_i, a_j) \leqslant d(a_i, a_k), \quad d(a_j, a_k) \leqslant d(a_i, a_k),$$

with $i \leqslant j \leqslant k$, which are violated by the sequence. This measure involves only comparisons of dissimilarity values which can meaningfully be made on the clinal postulate given above.

The algorithm devised by Sibson (1971) for minimizing λ starts with an arbitrary initial sequence and relocates the first object to the position which gives the least value of λ; the second object is then relocated to the position which gives the least value of λ and so on. When object n is reached we start again with the first object and repeat the cycle. The process is continued until there is no object whose relocation reduces λ. Unlike algorithms which attempt to minimize a distortion measure by running through all permutations of an initial sequence, this algorithm is computationally feasible for quite large numbers of populations (up to *ca.* 100). It cannot be guaranteed that the sequence finally obtained is at a global minimum of λ, but recovery of the same final sequence from a variety of different initial sequences provides valuable corroboration.

TABLE 3. A: SEQUENCE OF TWENTY-TWO INDIAN POPULATIONS OBTAINED FROM DCM′ BY THE ALGORITHM FOR MINIMIZING THE LOCAL DISTORTION MEASURE λ. B: SEQUENCE OF FIFTEEN CONGOLESE POPULATIONS OBTAINED FROM DCH BY THE ALGORITHM FOR MINIMIZING THE DISTORTION MEASURE λ. λ IS DEFINED ABOVE

A	1–9–4–6–2–3–5–7–11–12–10–13–15–8–14–16–19–18–21–17–20–22	$\lambda = 0.081$
B	14–15–11–12–8–4–2–6–5–3–9–7–10–1–13	$\lambda = 0.092$

The sequences obtained by Sibson's algorithm from DCM′ and DCH are shown in table 3. Both sequences are almost certainly global optima, and the low values of λ indicate that DCM′ and DCH show well-marked sequential structure.

Topoclinal differentiation is a stronger postulate than the various one-dimensional forms of clinal differentiation. The dissimilarity values of populations which show topoclinal differentiation should be approximately monotone with their geographical separations. The methods appropriate for the investigation of topoclinal differentiation are scaling methods which seek that disposition of points representing populations in two-dimensional euclidean space in which the interpoint distances minimize some measure of departure from monotonicity with values of a DC.

The non-metric multidimensional scaling algorithm devised by Kruskal (1964a, b) is appropriate for this purpose. This algorithm seeks that disposition of points in a given number of euclidean dimensions which minimizes the stress measure:

$$S = \sqrt{\frac{\Sigma\{d'(a, b) - \hat{d}(a, b)\}^2}{\Sigma\{d'(a, b)\}^2}},$$

where $d'(a, b)$ are the interpoint distances, and $\hat{d}(a, b)$ are numbers which minimize S subject to the condition that they be monotone with the values of the DC d. The minimization process is described fully in Kruskal (1964b). The algorithm cannot be used to find the one-dimensional disposition of points which minimizes the distortion measure, because in this case it becomes trapped in local minima. It does not, therefore, provide a direct alternative to sequencing methods for investigating one-dimensional clinal differentiation. If topoclinal differentiation over a large portion of the Earth's surface is to be investigated the method would have to be modified to seek a disposition of points on a part of the surface of a sphere.

Other quantitative methods for the study of topoclinal variation which might profitably be applied to human populations have been proposed by Marcus & Vandermeer (1966) and Gabriel & Sokal (1969).

The two-dimensional disposition of points representing populations found when Kruskal's algorithm was applied to DCM′ is shown in figure 22. The two-dimensional representation found when the algorithm was applied to DCH is shown in figure 23. The values of S for these dispositions are very low. The two-dimensional representation of DCH is closely similar to the scatter diagram for the fifteen populations given by Hiernaux (1966 a, figure 3) which is based on
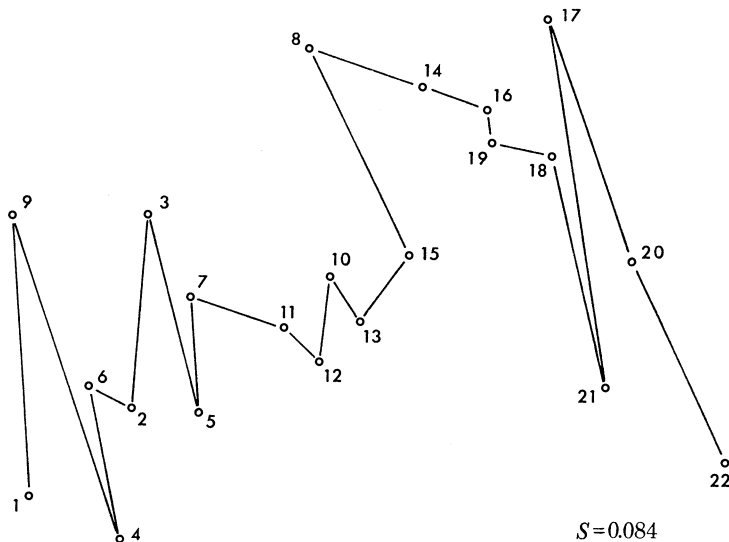

$S = 0.084$

FIGURE 22. Two-dimensional disposition of points representing twenty-two Indian populations obtained by applying Kruskal's non-metric multidimensional scaling algorithm to DCM′. S is the distortion imposed on DCM′ by the interpoint distances. Populations are linked in the sequence found by Sibson's scaling algorithm (see table 3 A).

$S = 0.041$
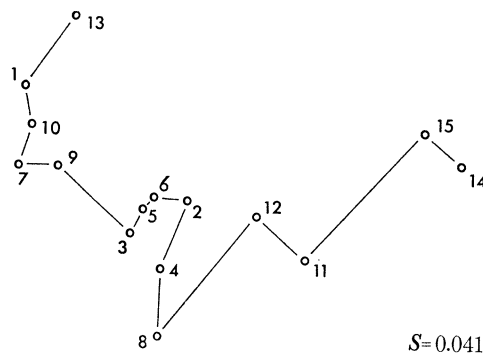
FIGURE 23. Two-dimensional disposition of points representing fifteen Congolese local populations obtained by applying Kruskal's non-metric multidimensional scaling algorithm to DCH. S is the distortion imposed on DCH by the interpoint distances. Populations are linked in the sequence found by Sibson's scaling algorithm (see table 3 B).

inspection of the values of the D^2 statistic based on nine ratios of morphological measurements. A good fit of the two-dimensional representations could have been predicted from the fact that sequences were found which impose low distortion on the DCs. Conversely, well-fitting sequences could have been predicted from the fact that each of the two-dimensional representations shows a pronounced linear tendency.

The results strongly suggest that in both cases the pattern of differentiation is clinal rather than clustered. The interpretation of this, and the factors which may determine the clinal differentiation are discussed in the final section.

VII. Interpretation of patterns of differentiation

Hiernaux (1956, 1966 a) provided a model of close reasoning and scientific caution in relating the results he obtained using Mahalanobis's D^2 on the fifteen Congolese populations to environmental and historical factors, and no brief account can do justice to his treatment. His summary of his main conclusions is translated in full:

'Having completed our analysis of the D^2 table we present the conclusions which we have drawn:

(1) A consistency in the rankings of values of D^2 between [each of] ten populations and the remaining five [i.e. clinal differentiation];

(2) A correlation of D^2 values with the number of genetic barriers;

(3) A mean of D^2 values for populations from different habitats greater than each of the means for populations from the same habitat.

With the help of sociological, historical and linguistic data, and, on the third point, with the help of the gene frequencies of the ABO and MN systems, we have been led to explain these three findings in evolutionary terms:

(1) There is gene flow in both directions along the gradient indicated; it is possible to specify in which direction the flow has predominated for certain of the stages;

(2) Gene flow from each population decreases in strength with each genetic barrier crossed;

(3) Community of origin and recent separation, indicated by historical and lexicostatistical data, combine to reduce values of D^2;

(4) Adaptation of morphology to environment brings together populations from the same habitat and separates those from different habitats.'

[Hiernaux 1956, pp. 93, 94]

My analyses of Hiernaux's data corroborate certain of his conclusions. The conclusions drawn by Mahalanobis *et al.* (1949), quoted on p. 18, are not confirmed and an alternative interpretation of this data is suggested.

It has already been pointed out that extreme caution should be exercised in interpreting the results of clustering, sequencing, and scaling methods. First, it is advisable always to check that a DC based on many attributes is stable in the sense defined in §III. A DC which is unstable is likely to be altered substantially if a slightly different selection of attributes is made, and correspondingly substantial changes in the results obtained by clustering are to be expected. Secondly, it is advisable to use some measure of the distortion imposed by a clustering on a DC. When the distortion imposed is high, or when the individual clusters are poorly isolated, it may be advisable to use some such significance test as was used in this study. Finally, it should be noted that clusters obtained by hierarchic clustering should not automatically be interpreted as indicative of racial or subspecific differentiation. Alternative explanations for cluster formation include uneven sampling from topoclinal, ecoclinal or genoclinal gradients; bias or incompleteness in the selection of local populations; and unwarranted pooling of local populations.

Any interpretation of the hierarchic clusterings obtained from DCM and DCH is open to serious doubt because they impose high distortion on the DCs and because the majority of the clusters are poorly isolated. Hiernaux (1956) suggested on linguistic and historical grounds that the Shi (2), Havu (3), Hunde (4), Swaga (5), and Shu (6) formerly inhabited a common territory in the west of Uganda. These populations form a poorly isolated cluster in the clustering obtained by the single-link method, and are grouped together in the two-dimensional array

obtained by non-metric multidimensional scaling, and it is possible that this is indicative of their common ancestry. However, as Hiernaux pointed out, these populations, unlike the other populations from Kivu, live in high savannah so that an alternative explanation in terms of parallel adaptation is available.

The clustering of the Indian populations appears at first sight to vindicate recognition of Chero (16), Kharwar (19), Panika (18), Majhi (17), Tharu (14), and Korwa (22) as a distinct racial group. They form a poorly isolated and inhomogeneous cluster. These populations had
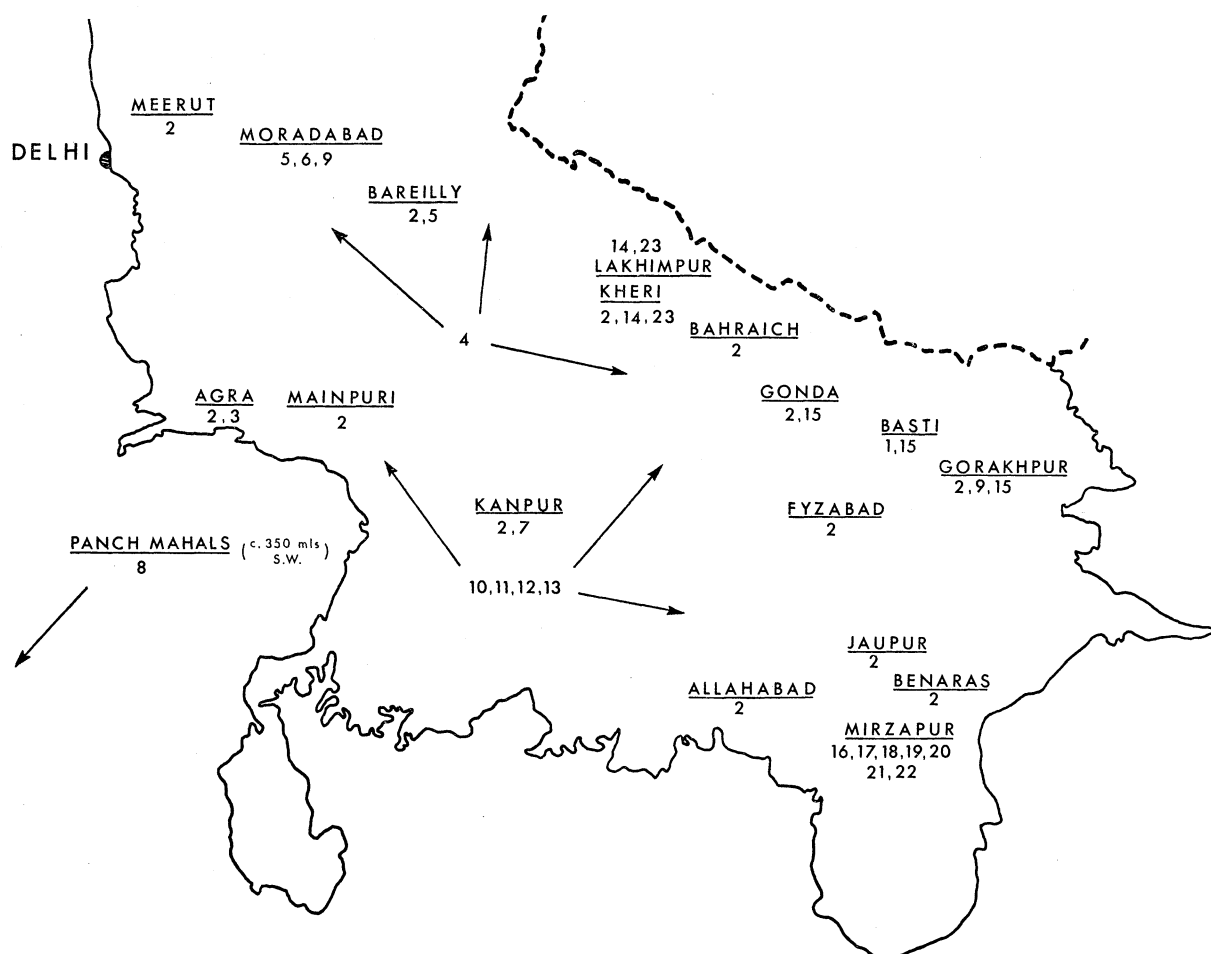
FIGURE 24. Geographical disposition of the twenty-three Indian caste and tribal groups
studied by Mahalanobis *et al.* (1949).

been considered by previous physical anthropologists cited by Mahalanobis *et al.* (1949) as representatives of a primitive 'Dravidian' stock related in ancestry to tribes in Southern India and Ceylon. Mahalanobis *et al.* cite evidence for their linguistic affinity and sharing of certain customs. A quite different interpretation of this clustering is offered below.

Application of sequencing and scaling methods to DCH and DCM′ shows that in both cases there is definite clinal structure. Whether or not this is topoclinal differentiation can be determined by comparing the two-dimensional dispositions of populations with their geographical distributions. Comparison of the two-dimensional representation of DCM′ (figure 22) with the geographical disposition of the populations, shown in figure 24, is complicated by the fact that several of the populations are caste-groups sampled from several localities. If these populations

are omitted, the comparison suggests topoclinal differentiation. The grouping together of the populations 16–22 in the single-link clustering is readily explained by their being sampled from a single locality.

Comparison of the two-dimensional representation of the DC on the Congolese populations (figure 23) with the geographical disposition of the populations, shown in figure 25, indicates this is not a case of topoclinal differentiation. The sequence obtained by Sibson's sequencing method appears not to be related directly to any of the environmental factors discussed by
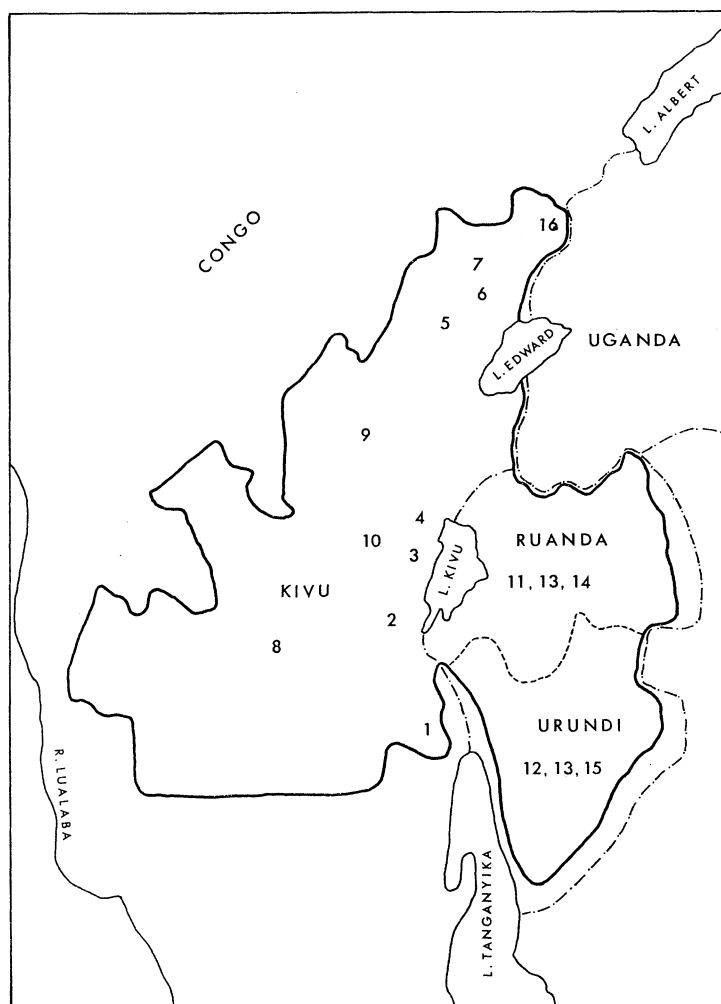
FIGURE 25. Geographical disposition of the fifteen Congolese populations.
Redrawn from Hiernaux (1956, figure 1).

Hiernaux (1956, 1966a), although there is some correlation with humidity of the environment. This does not rule out an ecoclinal gradient, because, as Mayr (1956) emphasized, clines may result from natural selection determined by the interaction of several environmental factors.

Hiernaux (1971 b) showed a two-dimensional representation of a DC on thirty-two populations which include the fifteen populations. The DC is based on ten morphological attributes, and the two-dimensional representation was obtained by Gower's (1966) method of latent vector analysis which is related to principal component analysis. Professor Hiernaux kindly made these data available to me. Non-metric multidimensional scaling of this DC yields a

two-dimensional representation similar to that obtained by Gower's method. In this representation the disposition of the fifteen populations remains similar to that obtained from DCH by multidimensional scaling.

Some further information about the clinal structure of DCH is obtained by sequencing of the DCs on certain selections from the fifteen populations. The DC on populations 2, 11, 12, 14 and 15 almost exactly fits the sequence 14–15–11–12–2. The DC on populations 1, 7, 9, 10, 13 almost exactly fits the sequence 13–1–10–7–9. The DC on the remaining populations yields no such well-fitting sequence. These results suggest that the populations in the tails of the disposition shown in figure 23 are clinally differentiated, but that the populations in the central part of the disposition may not be.

Hiernaux (1956) postulated the sequence, Tutsi (14 and 15)–Hutu (11 and 12)–Shi (2) as a genocline and this is confirmed. He further suggested that there might be more extensive clinal differentiation, but that this was not evident from examination of the D^2 values. The sequence Twa (13)–Fulero (1)–Tembo (10)–Mbuba (7)–Nyanga (9) may be interpreted in several ways. It may be a genocline showing decreasing influence of gene-flow from 'Pygmoid' populations which inhabit the same areas as these populations. Or it may be interpreted as showing progressive parallel adaptation to a humid forest environment. Or both processes may be involved. Hiernaux (1962, 1966 b, 1971 b) found that the ABO allele-frequencies of the Twa conflict with the hypothesis of gene-flow from the Mbuti pygmies, which suggests that the cline may represent progressive parallel adaptation rather than gene-flow.

I draw the following conclusions about the nature of the differentiation of the Indian and Congolese local populations:

(1) Differentiation with respect to different small sets of attributes is highly discordant. Differentiation with respect to different larger sets of attributes is less discordant so that in each case the pattern of differentiation can reasonably be described by a dissimilarity coefficient based on all the available attributes.

(2) In neither case is the pattern or differentiation well represented by any hierarchic racial classification.

(3) In both cases a clinal pattern of differentiation predominates.

Any extrapolation from the results obtained on these limited selections of local human populations must be tentative. Both the Indian and the Congolese populations show substantial diversity and each has in the past been partitioned into discrete races. The fact that in both cases racial classification is unwarranted casts doubt on the validity of racial classification of humans. The extent to which clinal differentiation is prevalent amongst human populations, and the factors which determine human clinal differentiation, are questions which deserve further investigation.

REFERENCES

Ali, S. M. & Silvey, S. D. 1966 A general theory of coefficients of divergence of one distribution from another. *Jl R. Statist. Soc.* B, **27**, 131–142.

Ball, G. H. 1966 Data analysis in the social sciences: what about the details? *Proc. Fall Joint Comput. Conf.* **27**, 533–559.

Barber, H. N. & Jackson, W. D. 1957 Natural selection in action in *Eucalyptus. Nature, Lond.* **179**, 1267–1269.

Biasutti, R. 1953–7 *Razze e popoli della terra*, 2nd ed. Turin: Unione Tipographico.

Boyd, W. C. 1950 *Genetics and the races of man.* Boston, D.C.: Heath.

Boyd, W. C. 1963 Genetics and the human race. *Science, N.Y.* **140**, 1057–1064.

Briggs, L. C. 1957 A review of the physical anthropology of the Sahara and its prehistoric implications. *Man* **57**, 20–23.

Bronowski, J. & Long, W. M. 1952 Statistics of discrimination in physical anthropology. *Am. J. phys. Anthrop.* **10**, 385–393.

Calinski, T. & Harabasz, J. 1971 A dendrite method for cluster analysis. *Biometrics.* (In the press.)

Cavalli-Sforza, L. L. & Edwards, A. W. F. 1967 Phylogenetic analysis: models and estimation procedures. *Evolution* **21**, 555–570.

Chaddha, R. L. & Marcus, L. F. 1968 An empirical comparison of distance statistics for populations with unequal covariance matrices. *Biometrics* **24**, 683–694.

Cole, A. J. & Wishart, D. 1970 An improved algorithm for the Jardine–Sibson method of generating overlapping clusters. *Comput. J.* **13**, 156–163.

Coon, C. S. 1962 *The origin of races.* New York: Knopf.

Cormack, R. M. 1971 A review of classification. *Jl R. statist. Soc.* B. (In the press.)

Deniker, J. 1900 *The races of man.* London: Walter Scott.

Ehrlich, P. R. & Holm, R. W. 1969 A biological view of race. In Montagu, A. (ed.), *The concept of race*, pp. 153–179. London: Collier-Macmillan.

Fisher, R. A. 1936 'The coefficient of racial likeness' and the future of craniometry. *Jl R. anthrop. Inst.* **66**, 57–63.

Florek, K., Łukaszewicz, J., Perkal, H., Steinhaus, H. & Zubrzycki, S. 1951 Sur la liaison et la division des points d'un ensemble fini. *Colloquium math.* **2**, 282–285.

Gabriel, K. R. & Sokal, R. R. 1969 A new statistical approach to geographic variation analysis. *Syst. Zool.* **18**, 259–278.

Garn, S. M. 1968 *Human races*, 2nd ed. Springfield, Illinois: Thomas.

Gillham, N. W. 1956 Geographic variation and the subspecies concept in butterflies. *Syst. Zool.* **5**, 110–120.

Gower, J. C. 1966 Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **53**, 325–338.

Gower, J. C. 1967 A comparison of some methods of cluster analysis. *Biometrics* **23**, 623–637.

Gregor, J. W. & Watson, P. J. 1961 Ecotypic differentiation: observations and reflections. *Evolution* **15**, 166–173.

Hagmeier, E. M. 1958 The inapplicability of the subspecies concept in the North American marten. *Syst. Zool.* **7**, 1–7.

Hiernaux, J. 1956 Analyse de la variation des caractères physiques humains en une région de l'Afrique centrale: Ruanda-Urundi et Kivu. *Annls. Mus. r. Congo belge (Sciences de l'homme)* **3**, 7–131.

Hiernaux, J. 1962 Données génétiques sur six populations de la République du Congo. *Annls Soc. belge Méd. trop.* **42**, 145–174.

Hiernaux, J. 1964 La mesure de la différence morphologique entre populations pour un ensemble de variables. *Anthropologie* **68**, 559–67.

Hiernaux, J. 1965 Une nouvelle mesure de distance anthropologique entre populations, utilisant simultanément des fréquences géniques, des pourcentages des traits descriptifs et des moyens métriques. *C. r. hebd. Séanc. Acad. Sci., Paris* **260**, 1748–1750.

Hiernaux, J. 1966 a Human biological diversity in central Africa. *Man* **1**, 287–306.

Hiernaux, J. 1966 b Les Bushong et les Cwa du royaume Kuba (Congo-Kinshaha): Pygmées, pygmöides et pygméisation; anthropologie, linguistique et expansion bantoue. *Bull. Mém. Soc. Anthrop., Paris* **9**, 299–336.

Hiernaux, J. 1968 *La diversité humaine en Afrique subsaharienne.* Editions de l'Institut de Sociologie de l'Université Libre de Bruxelles.

Hiernaux, J. 1969 The concept of race and the taxonomy of mankind. In Montagu, A. (ed.). *The concept of race*, pp. 29–45. London: Collier-Macmillan.

Hiernaux, J. 1971 a The analysis of multivariate biological distance between human populations: principles, and applications to Subsaharan Africa. In Weiner, J. S., and Huizinga, J. (eds.), *The assessment of population affinities in man.* Oxford: Clarendon Press.

Hiernaux, J. 1971 b Numerical taxonomy of man; an application to a set of 32 African populations.

Highton, R. 1962 Revision of North American salamanders of the genus *Plethodon. Bull. Fla St. Mus. biol. Sci.* **61**, 235–367.

Hogben, L. 1960 The race concept. In Mason, P. (ed.), *Man, race and Darwin*, pp. 11–23. Oxford University Press.

Hole, F. & Shaw, M. 1967 Computer analysis of chronological seriation. *Rice Univ. Stud.* **53**, 1–166.

Howells, W. W. 1951 Factors of human physique. *Am. J. phys. Anthrop.* **9**, 159–192.

Howells, W. W. 1957 The cranial vault. Factors of size and shape. *Am. J. phys. Anthrop.* **15**, 19–48.

Huizinga, J. 1962 From DD to D^2 and back. The quantitative expression of resemblance. *Proc. K. ned. Akad. Wet.* C **65**, 1–12.

Huxley, J. S. 1938 Clines: an auxiliary taxonomic principle. *Nature, Lond.* **142**, 219–220.

Jackson, D. M. & Spärck-Jones, K. 1967 Current approaches to classification and clump finding at the Cambridge Language Research Unit. *Comput. J.* **10**, 29–37.

Jardine, C. J., Jardine, N. & Sibson, R. 1967 The structure and construction of taxonomic hierarchies. *Math. Biosci.* **1**, 173–179.

Jardine, N. & Sibson, R. 1968*a* The construction of hierarchic and non-hierarchic classifications. *Comput. J.* **11**, 177–184.

Jardine, N. & Sibson, R. 1968*b* A model for taxonomy. *Math. Biosci.* **2**, 465–482.

Jardine, N. & Sibson, R. 1971 *Mathematical taxonomy*. New York: Wiley.

Kruskal, J. B. 1964*a* Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika* **29**, 1–27.

Kruskal, J. B. 1964*b* Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129.

Lance, G. N. & Williams, W. T. 1967 A general theory of classificatory sorting strategies. I. Hierarchical systems. *Comput. J.* **9**, 373–380.

Lerman, I. C. 1969 H-classificabilité. In *L'emploi des calculateurs en archéologie, problèmes sémiologiques et mathématiques*. Colloque international du Centre National de la Recherche Scientifique, Marseille.

Lerman, I. C. 1970 *Les bases de la classification automatique*. Paris: Gauthier-Villars.

Livingstone, F. B. 1962 On the nonexistence of human races. *Curr. Anthrop.* **3**, 279–281.

Mahalanobis, P. C. 1930 On tests and measures of group divergence *J. asiat. Soc. Beng.* **26**, 541–588.

Mahalanobis, P. C. 1936 On the generalised distance in statistics. *Proc. natn. Inst. Sci. India* **2**, 49–55.

Mahalanobis, P. C., Majumdar, D. N. & Rao, C. R. 1949 Anthropometric survey of the United Provinces. *Sankhyā* **9**, 89–324.

Marcus, L. F. & Vandermeer, J. H. 1966 Regional trends in geographic variation. *Syst. Zool.* **15**, 1–13

Mayr, E. & Vaurie, C. 1948 Evolution in the family Dicruridae (birds). *Evolution* **2**, 238–265.

Mayr, E. 1956 Geographical character gradients and climatic adaptation. *Evolution* **10**, 105–108.

Montagu, A. 1969 (ed.) *The concept of race*. London: Collier-Macmillan.

Mourant, A. E. 1956 *The distribution of blood groups in animals and humans*. Springfield, Mass.: Thomas.

Petersen, B. 1947 Die geographische Variation einiger Fennoskandischer Lepidopteren: *Zool. Bidr. Upps.* **26**, 329–531.

Sibson, R. 1969 Information radius. *Z. Wahr.* **14**, 149–160.

Sibson, R. 1971 Some thoughts on sequencing methods. *Proceedings of the Anglo-Romanian Conference on Mathematics in the Historical and Archaeological Sciences, Mamaia, Romania, 1970.*

Sneath, P. H. A. 1957 The application of computers to taxonomy. *J. gen. Microbiol.* **17**, 201–226.

Sokal, R. R. & Rinkel, R. C. 1963 Geographical variation of alate *Pemphigus populi-transversus* in eastern North America. *Kans. Univ. Sci. Bull.* **44**, 467–507.

Sokal, R. R. & Rohlf, F. J. 1962 The comparison of dendrograms by objective methods. *Taxon* **11**, 33–40.

Sokal, R. R. & Sneath, P. H. A. 1963 *Principles of numerical taxonomy*. San Francisco and London: Freeman.

Vogel, F. 1970 ABO blood groups and disease. *Am. J. hum. Genet.* **22**, 464–475.

Wiener, A. S. 1970 Blood groups and disease. *Am. J. hum. Genet.* **22**, 476–483.