A QUANTITATIVE APPROACH TO A PROBLEM IN CLASSIFICATION [1]

Charles D. Michener and Robert R. Sokal [2]

Department of Entomology, University of Kansas, Lawrence

Introduction

The purpose of the study reported in this paper was to determine whether certain statistical procedures might aid persons interested in the relationships among organisms. The objectives of our study were to investigate numerous characters simultaneously in a considerable group of species; to quantify the relations shown among the species, using objective methods; and to indicate these relationships. The organisms selected as an example for use in this study are solitary bees in the family Megachilidae. This choice was made because one of us (C. D. M.) has made recent systematic studies of these insects, so that conclusions as to the relationships obtained by usual systematic procedures could be compared

with the results of the methods described below.

A major concern of the systematist, when he devises a classification, is to make one in accordance with probable phylogeny, but first he must assess the characters and determine relationships in a static or nonhistorical sense. Later he may decide on the most likely lines of descent. The greater part of this paper concerns the first step, determination of static relationships, although ideas as to phylogeny are, of course, included.

In the absence of data from fossils, and sometimes even when these are available, systematists often differ widely as to relationships among organisms. This is true even if problems raised by convergence and varying rates of evolution can be settled or temporarily ignored. It is easier to appreciate differences than similarities. Therefore, in most groups, the species are differentiated long before serious studies of relationships begin. A systematist who is interested in phylogeny, but lacks paleontological data, utilizes character correlations observed in the course of his study in determining relationships and delimiting the levels (genera, families, etc.) in the hierarchy of classification.

Weighting of Characters

Each species has an enormous number of characters. The systematist normally tries to use as many as he can find, but is hampered in his study of character correlations by the limitations of the human mind. Also, he is necessarily subjective in his choice and weighting of characters.

If he notes that certain characters vary erratically with respect to one another

among various species, the systematist will regard the numerous combinations of such characters as valuable in differentiating species, but will not use these characters in erecting higher categories, such as genera. On the other hand, when he finds several characters, which are not related by function or other obvious causal mechanisms, consistently associated with one another among several species, and a quite different combination of characters associated in another group of species, he feels that he has found evidence of two species clusters and probably of two phyletic lines. The more numerous are the differentiating characters associated within each group, the more confident is he of the conclusion. Even though they may be minute and inconspicuous, the systematist will weight such characters heavily in comparison with those which vary from species to species with no obvious relation to one another. His judgments as to relationships, and hence higher categories, will therefore be largely based upon a few characters which he finds to be perfectly associated. Especially if the number of such characters is small, the discovery of additional character associations may alter judgment about relationships. Sometimes, as a result of parallel evolution, two or more sets of associated characters may suggest different relationships. In this case the systematist must decide, on the basis of experienced judgment or mere intuition, which set of characters indicates phyletic relationships. In brief, he will give greater weight to the characters that he believes most significant.

The problem of weighting characters is important, but it is not obvious how the weighting should be done. No doubt the so-called conservative characters should be heavily weighted, but how are they to be recognized with certainty? A systematist can often give sound reasons for his judgments on this matter; as indicated above, they usually have to do with character associations. On the other hand, his judgments are often quite sub-

jective so that the development of a classification may be more of an art than a science.

Weighting might be according to the number of genes influencing a character, but no one knows how many genes are involved in the formation of most characters of even the genetically best known organisms; therefore, this method of weighting is useless. Even if we knew the number of genes affecting a character, this would not give us a basis for evaluating its importance in judging relationships.

It is often said that "purely adaptive" characters are unlikely to be conservative and hence should receive less weight in studies of relationships than characters having no obvious adaptive value. Thus, among a group of species, several may have independently invaded a particular habitat and independently evolved the same adaptive character. This situation becomes obvious if conservative characters are found which show the true phylogeny. On the other hand, it is often evident that obviously adaptive features do characterize large groups of species and hence do indicate phyletic relationships. It is also probable that every character, if not itself adaptive, is at least determined by genes which have adaptive value through some other influence. For these reasons we cannot use adaptiveness, even if there were a way to measure it, as a criterion for weighting.

We are left with the conclusion that, in the absence of fossils, constant association of several characters which are not related causally is the best evidence of the conservatism of these characters. If only a few characters are used, studies of relationships require weighting of certain conservative characters in the mental processes of the systematist. By using a large number of characters and species, however, we believe that weighting becomes unnecessary because the magnitude of correlation coefficients calculated between species would be little affected except by extreme weighting.

Kinds of Characters

Almost all of the characters used in our study are morphological or color features of the adult insects. One of us (Michener, 1953) has emphasized the value of characters of other stages (e.g., larvae, pupae), of behavioral characters, etc., but also has stressed that these are merely additional characters, of neither more nor less importance than other adult morphological characters that might be found. We have used external adult characters because they are the only ones known for most of the species concerned. Also, since a very large number of them were considered, we believe that the use of additional characters might not have materially altered our results, because additional characters have a decreasing effect on the index of relationship which we have used.

We believe that our morphological characters represent, in a genetic sense, a random sample of all the characters, and thus a random sample of genetic variation. There are no special groups of morphological genes. Many genes influence morphological as well as behavioral or physiological characters.

Quantification of Relationships

The methods described below make possible a certain degree of quantification of similarities, using many more characters than are ordinarily employed in a systematic study. This means that we can give some measure of the amount of evolutionary change that has occurred since phyletic lines diverged. It is our basic assumption that degree of difference can give an estimate of evolutionary divergence (E). E is clearly the product of $r \times t$, where r is the rate of evolution (dE/dt) and t is time. From an estimate of E, there is no way to obtain r or t; it is, for example, quite probable that one pair of species diverged quite recently compared to another equally different pair.

Possible parallelisms present a difficult problem. The amount of evolution, after divergence from a common stem, of two forms that have evolved in a parallel manner, would be underestimated by our method, as well as by customary systematic methods. In the majority of cases in animals, however, parallelisms probably involve only a few characters. Thus in a study of parallelisms in saturniid moths (Michener, 1949a) many structures were found to appear or disappear independently in various genera but in no instance did a large group of associated characters behave in this fashion. Among the bees used in the present study there is evidence of acquisition of clear wings, red integumental pigment and abundant pale pubescence by various unrelated species in desert areas. These tendencies could affect a maximum of only five of the 122 characters studied. We believe that much of the difficulty due to parallelisms is eliminated in our study by the large number of characters used. Parallelism or convergence affecting only a few characters would not greatly influence our measure of relationship.

It is impossible to show all details of relationships by any hierarchy of systematic categories unless the latter are excessively numerous. Therefore, it is common for a systematist to state that a certain genus is closely related to another and that these two are more distantly related to others within a tribe or subfamily. He does not try to indicate all such details of relationships by the usual classificatory system. The systematist himself will be quite aware of this situation and may explain it, but to the uninitiated user of his work, the hierarchy may constitute something of a strait jacket for thoughts on relationships.

To indicate relationships more precisely than is possible by means of systematic categories, various diagrammatic schemes have been adopted. The commonest such scheme is a phylogenetic tree. Such trees are most justifiable if abundant and well distributed fossil evidence supports them. In the absence of such evidence we prefer to call such trees diagrams of relation-

ships. The dangers of such diagrams have been explained by many authors, but diagrams of relationships continue to be useful, and as yet no better method of indicating relationships has been devised for general use. We have therefore presented many of our results as diagrams of relationships. The details of the meanings of such diagrams will be discussed later.

The roles of the two authors may be described as follows: The basic data were gathered from previous publications and from specimens by C. D. M. After considerable discussion of methods, the statistical work (described in detail by Sokal and Michener, 1957) was carried out by R. R. S. and assistants without reference to C. D. M.'s systematic knowledge of the group. Thus, the relationships determined from statistical procedures could not have been biased by prior views of relationships based on ordinary systematic methods.

It may be suggested that bias has been introduced by our use of the same characters that were used in previous systematic studies. As other characters are not known, we were, of course, forced to use these. Moreover, it is our belief that any large group of characters would indicate similar relationships since, as we have said, we believe these characters to represent a random sample of the genetic variation.

MATERIALS

The *Hoplitis* Complex and Its Older Classification

The bees used in this study are the American species of the megachilid genera *Ashmeadiella, Anthocopa, Hoplitis,* and *Proteriades.* These four genera, for convenience, may be called the *Hoplitis* complex; they are part of the genus *Osmia* as it is understood by European workers. A total of 97 species were included in the study after the few forms known from only one sex were eliminated. We do not believe that exclusion of such forms affects the usefulness of this work as a pilot study to determine the suitability of the method. There must have been many species which are now extinct, which can never be included in any such study. Therefore, the exclusion of certain other species should not affect our evaluation of the method. Fortunately, among the 97 included species, there is at least one representative of every subgenus or species group recognized in the earlier systematic studies.

As has been explained elsewhere (Michener, 1941; Hurd and Michener, 1955) the genera of the *Hoplitis* complex are closely related. Species of two of them (*Hoplitis* and *Anthocopa*) are found in Eurasia (with a few species in Africa), but with one exception Eurasian forms belong to different subgenera than American forms. We believed this complex particularly appropriate for our study because previous investigations indicated that it contains some large groups of very similar species as well as some apparently isolated species.

The bees of this group are solitary rather than social; each female is a reproductive, there being no worker caste. Nests are made in the soil or in wood or stems. Most of our species inhabit western America although several occur in the east; none is known south of Mexico.

Table 1 (the formal classification) shows the names and code numbers of the 97 species and the way in which they were organized into genera, subgenera, and species groups prior to the present study. This classification is merely an attempt to place conveniently within the framework of a classificatory hierarchy, the ideas which are indicated with more accuracy and detail in the diagrams of relationships (figs. 1 and 2). These figures show relationships among the species groups including various details which cannot be indicated in the formal classification. Citations to the systematic papers which form the basis for these figures and table appear in Michener (1949b) and in Hurd and Michener (1955), with some

CHARLES D. MICHENER AND ROBERT R. SOKAL

TABLE 1. *Formal classification of 97 species in the* Hoplitis *complex*

Genus group	Genus	Subgenus	Species group	Species code number and name
1.	*Proteriades*	*Cephalapis*	—	1. *jacintana* (Cockerell)
		Proteriades s. str.	—	2. *semirubra* (Cockerell) 3. *deserticola* Timberlake & Michener
		Xerosmia	A	4. *xerophila* (Cockerell)
			B	5. *nanula* Timberlake & Michener 6. *pygmaea* Timberlake & Michener 7. *reducta* Timberlake & Michener 8. *cryptanthae* Timberlake & Michener 9. *caudex* Timberlake & Michener 10. *hamulicornis* Timberlake & Michener 11. *truicauda* Timberlake & Michener 12. *bidenticauda* Timberlake & Michener 13. *nigrella* Michener 14. *tristis* Michener 15. *palmarum* (Cockerell) 16. *seminigra* Timberlake & Michener
			C	17. *remotula* (Cockerell) 18. *incanescens* (Cockerell)
	Hoplitis	*Hoplitina*	—	19. *howardi* (Cockerell) 20. *bunocephala* Michener 21. *bullifacies* Michener
		Acrosmia	—	22. *laevibullata* (Michener) 23. *rufina* Michener
		Andronicus	—	24. *cylindrica* (Cresson)
		Monumetha	A	25. *fulgida* (Cresson) 26. *louisae* (Cockerell) 27. *viridimicans* (Cockerell)
			B	28. *albifrons* (Kirby)
		Dasyosmia	—	29. *biscutellae* (Cockerell) 30. *paroselae* Michener
		Cyrtosmia		31. *hypocrita* (Cockerell)
		Alcidamea	A	32. *producta* (Cresson) 33. *pilosifrons* (Cresson) 34. *grinnelli* (Cockerell) 35. *colei* (Crawford) 36. *elongaticeps* Michener 37. *uvulalis* (Cockerell)
			B	38. *sambuci* Titus 39. *truncata* (Cresson)
		Formicapis	—	40. *clypeata* (Sladen)
		Robertsonella	—	41. *simplex* (Cresson)

TABLE 1—*Continued*

Genus group	Genus	Subgenus	Species group	Species code number and name
2.	*Anthocopa*	*Atoposmia*	A	42. *triodonta* (Cockerell) 43. *hebitis* Michener 44. *pycnognatha* Michener
			B	45. *oregona* Michener 46. *anthodyta* Michener 47. *panamintensis* Michener 48. *elongata* (Michener) 49. *abjecta* (Cresson)
		Hexosmia	—	50. *copelandica* (Cockerell)
		Eremosmia	A	51. *hemizoniae* (Cockerell) 52. *viguierae* (Cockerell) 53. *enceliae* (Cockerell)
			B	54. *rubrella* Michener 55. *hurdiana* Michener
			C	56. *daleae* Michener 57. *beameri* Michener 58. *mirifica* Michener 59. *hypostomalis* Michener 60. *nitidivitta* Michener 61. *rupestris* (Cockerell) 62. *robustula* (Cockerell)
	Ashmeadiella	*Chilosima*	—	63. *holtii* Cockerell 64. *rhodognatha* Cockerell
		Ashmeadiella s. str.	A	65. *altadenae* Michener 66. *rufitarsis* Michener
			B	67. *titusi* Michener 68. *femorata* (Michener) 69. *truncativentris* Michener
			C	70. *rufipes* Titus 71. *prosopidis* (Cockerell) 72. *sonora* Michener 73. *occipitalis* Michener 74. *bucconis* (Say) 75. *cactorum* Cockerell 76. *bigeloviae* (Cockerell) 77. *difugita* Michener 78. *leucozona* Cockerell 79. *foveata* Michener 80. *gillettei* Titus 81. *maxima* Michener 82. *meliloti* (Cockerell) 83. *aridula* Cockerell 84. *californica* (Ashmead) 85. *opuntiae* (Cockerell)

TABLE 1—*Continued*

Genus group	Genus	Subgenus	Species group	Species code number and name
		Titusella	—	86. *cubiceps* (Cockerell)
		Rhamphorhina	A	87. *foxiella* Michener
			B	88. *timberlakei* Michener 89. *eurynorhyncha* Michener 90. *stenognatha* Michener 91. *salviae* Michener 92. *cazieri* Michener 93. *clypeodentata* Michener 94. *australis* (Cockerell)
		Corythochila	—	95. *breviceps* Michener 96. *inyoensis* Michener
		Cubitognatha	—	97. *xenomastax* Michener

FIG. 1. Diagram of relationships for species groups in the genera *Proteriades* and *Hoplitis* as visualized by C. D. M. at beginning of this study.

The diagram has been modified from the more conventional dendritic form in order to be comparable to later diagrams (especially figs. 5–8 and 12–15). Relative levels of forks on any one main branch are related to primitiveness of the forms concerned. In contrast with the later figures, however, length of the various segments of the lines and relative levels of forks on different lines are of no significance. Another difference between this and later diagrams is that the tip of each line represents a species group, not a species. This is because C. D. M. did not attempt to indicate the structure within species groups. The broken line terminated by arrows separates the two genera. Generic names are capitalized, subgeneric names are in lower case type. The numbers at the tips of the lines are species code numbers.
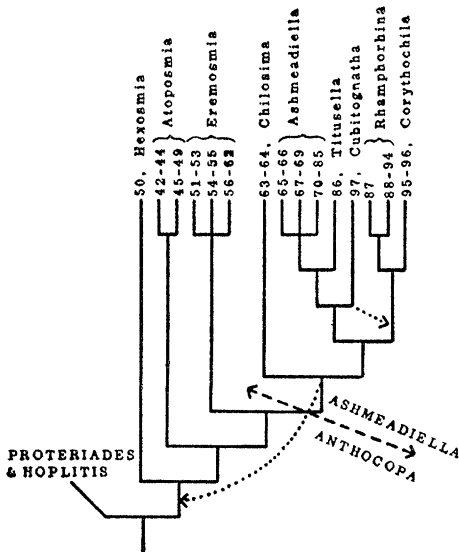
FIG. 2. Diagram of relationships for species groups in the genera *Anthocopa* and *Ashmeadiella* as visualized by C. D. M. at beginning of this study.

Explanatory comments as for figure 1. Dotted lines with single arrows point to possible alternative origins of certain stems.

significant changes in views being indicated in the latter paper.

These diagrams, constructed on the basis of systematic studies, undertaken and largely published before statistical work was contemplated, show the relationships as they were then understood. Since diagrams of relationships, phylogenetic trees and the like often have different meanings to different workers, we elaborate on the meaning of our diagrams as follows: They show that in the *Hoplitis* complex two major groups (1, *Proteriades, Hoplitis;* 2, *Anthocopa, Ashmeadiella*) arose from a common ancestor. The relative levels of forks on any one main branch are related to the primitiveness of the forms concerned. For example *Hexosmia* and *Robertsonella,* believed to have retained relatively primitive characters, should be more closely related than any other pair of subgenera in opposite major groups. Moreover, they should be more similar to the common ancestor than

others in their major groups. The subgenera *Alcidamea, Andronicus, Cyrtosmia, Dasyosmia,* and *Monumetha* were believed equally related to one another. The location of any branch to the right or to the left is of no significance. For example, *Alcidamea* was thought to be as closely related to *Monumetha* as to *Cyrtosmia.* Lengths of the various segments of the lines, and relative levels of forks on different lines are of no significance.

In figures 1 and 2, as well as in table 1, and throughout the remainder of this paper, generic and subgeneric names are freely used to avoid needless lists of species names or numbers. However, in all aspects of the work every effort was made to avoid bias due to the previous existence of such names. In the interest of brevity, the species of bees will hereafter be referred to by number only.

Characters Studied

The characters utilized were not the result of a new search, but were those that had been found of value in distinguishing any American species or group within the *Hoplitis* complex. Not all of them had been used or described within any one genus, but all had been used somewhere within the complex.

The word "character" has been used in two rather different ways by taxonomists. In its commonest usage, any feature of one kind of organism that differentiates it from another kind is a character. Thus the red abdomen of one bee is a character distinguishing it from another bee with the abdomen black. In this paper we use the word in a second connotation only; that is, as a feature which varies from one kind of organism to another. Now, to use the above example, abdominal color is the character; it occurs in two "states" or alternatives, red and black. The way in which the states were delimited will be more clearly understood from a perusal of examples detailed in the following paragraphs.

A total of 122 characters were found, each varying among the species so that
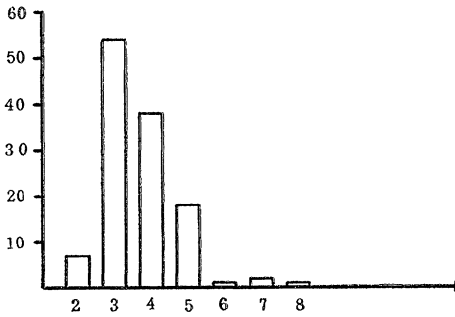
FIG. 3. Bar diagram showing number of states per character in the 122 characters of the study.

Abscissa: numbers of states per characters. *Ordinate:* frequency, number of characters per class. Average number of states per character equals 3.69.

there were from two to eight alternatives · or states per character. Figure 3 is a bar diagram showing the number of states per character for the 122 characters of the study. The average number of states per character was 3.69. Numbers ranging from 1 to 8, hereafter called "state codes," were used to record the state of all characters for each species. The total number of character states was 450. Some characters could easily have been more finely subdivided to make more states.

In the study of the 97 species we considered a total of 11,834 characteristics. Each of the many characters that are regularly correlated between the sexes, even though different in males and females, was treated as a single character. Thus, the width of the genal area in relation to the eye was determined for females and ignored in males. It is usually markedly less in males than in females of any one species and is often used in distinguishing species in both sexes. It could reasonably have been used as two separate characters, one for each sex, each with several states. The number of characters could easily have been increased by about one-third had all similar cases been treated in this way.

Table 2 lists the characters studied, with the number of states into which each was subdivided. Reference to this table will show that characters of many sorts were used. The following paragraphs illustrate by means of examples the way in which we dealt with most of the problems encountered. Of course the states were always arranged in some logical linear sequence. A few characters were meristic, such as number of segments of the maxillary palpi, which ranges from 5 to 2, with obviously 4 states (state 1 = 5-segmented; 2, 4-segmented; 3, 3-segmented; 4, 2-segmented).

Sizes of the body and of various structures were used as characters. In most such cases a complete range existed from the largest to the smallest; the range was arbitrarily divided into a convenient number of states and appropriate states recorded for each species; when a species fell on a borderline between states, it was placed arbitrarily on one side or the other.

Coloration was treated in a similar manner. For example, the metasoma varies from red to black; 4 states were employed: (1) black, (2) red on the first and second terga only, (3) red on the first to third or fourth terga, and (4) metasomal terga entirely red or nearly so. A few species have a borderline type of coloration or their range of coloration overlaps two states and hence could have been placed, for example, in either 3 or 4. They had to be placed arbitrarily in one or the other. Fortunately we were not confronted with forms in which only the apical terga were red; had such existed, an effort would have been made to distinguish the two types of coloration for all species and to recognize two characters, one for red expanding from the base of the metasoma, the other for red expanding from the apex.

Red markings of the same sort occur in some species of each of the four genera. This is an obvious example of parallelism (the presumed primitive forms of each genus have little or no red) but we used the same character state numbers for the same condition in each genus. To do otherwise would introduce a deliberately

TABLE 2

Characters, with the number of states into which each was divided shown at right. Characters marked by an asterisk (*) were observed in females but are so well correlated with the same male structures that no separate characters are listed for the males. Characters marked by a dagger (†) are for males only. Those that are unmarked are for females only.

*Size... 4
  Extent of projection of apex of clypeus over labrum.................................... 4
  Degree to which clypeal apex projects forward rather than downward...................... 4
  Breadth of clypeal truncation (broad to narrow and produced)........................... 7
  Ends of clypeal truncation (almost unrecognizable to much produced).................... 6
  Shape of clypeal truncation (convex to concave)...................................... 4
  Sublateral angles of clypeus... 5
  Emargination between truncation and sublateral angle of clypeus....................... 5
  Width of impunctate margin of clypeus... 4
  Convexity of clypeus.. 5
  Spaces between clypeal punctures (close to clypeus impunctate)........................ 4
  Longitudinal median line of clypeus (absent to distinct and raised).................... 4
  Coarseness of clypeal punctures (finer than to coarser than those of frons)............. 3
*Coarseness of punctures of vertex (finer than to coarser than those of mesoscutum)......... 3
*Ratio of distance from antennal bases to anterior ocellus to distance from anterior ocellus to
    posterior margin of vertex.. 4
*Ratio of distance between posterior ocelli to distance from one of them to eye margin......... 3
*Ratio of distance between posterior ocelli to distance from them to posterior margin of vertex.. 4
*Ratio of genal width to eye width.. 5
  Degree of convergence (or divergence) of inner margins of eyes......................... 3
  Clypeal brushes (absent to large)... 5
  Mandibular hairs (sparse to dense brush)... 4
  Width of mandibular apex.. 4
  Width of mandible at narrowest median region....................................... 4
  Length of mandible.. 3
  Size of upper mandibular tooth (absent to large)..................................... 3
  Inner subapical swelling of mandible (absent to tooth-like)............................ 4
  Ratio of distance from upper to middle mandibular tooth to distance from middle to lower tooth.. 5
  Shape of upper mandibular tooth (acute to absent).................................... 4
*Height of hypostomal carinae.. 3
  Posterior angles of hypostomal carinae... 2
  Shape of hypostomal areas (long to short).. 3
  Width of proboscidial fossa.. 3
*Segments of maxillary palpus (5 to 2).. 4
*Ratio of length of first to second segment of labial palpus............................... 4
*Setae on labial palpi and galeae (few straight, to many wavy).......................... 5
  Shape of labrum (broad rounded, to narrow bidentate)................................ 4
*Length of proboscis... 4
  Ratio of length to breadth of antennal pedicel.. 2
  Ratio of lengths of first to second flagellar segments.................................. 4
  Ratio of length of antennal scape to distance between antennal bases.................... 3
*Shape of thorax (elongate to nearly spherical).. 4
*Pre-episternal carina (absent to long).. 3
*Carina of pronotal lobe (absent to distinct).. 3
*Hind tibial spur (slender to robust).. 2
*Hind coxal carina (absent to high lamella)... 5
*Coarseness of punctures of mesoscutum (by comparison to those of episterna)............. 3
*Coarseness of punctures of episterna (by comparison to those of genal areas)............. 3
*Spaces between mesoscutal punctures.. 4
*Pattern of pubescence at anterior end of mesoscutum (none, pair of spots, merged spots)...... 4
*Size of papillae on wings... 3
*Wing color (clear to dusky)... 3
*Depth and definition of concavity of first metasomal tergum........................... 4
*Number of fasciae on metasomal terga... 5
*Coarseness of largest metasomal punctures (by comparison to those of mesoscutum).......... 3
*Coarseness of metasomal punctures in general.. 4

TABLE 2—*Continued*

*Width of smooth margins of metasomal terga.. 3
*Distinctness of basal grooves of metasomal terga... 2
 Length of scopa... 3
 Amount of scopa on sixth metasomal sternum.. 3
 Size of depression of sixth metasomal sternum (absent to broad)......................... 4
 Amount of projection of sixth metasomal sternum beyond corresponding tergum.............. 3
 Profile of sixth metasomal tergum (straight to strongly concave)........................ 3
 Width of apical flange of sixth metasomal tergum.. 3
*Integumental color (black to metallic).. 3
*Metasomal color (black to red)... 4
*Femoral color (black to red)... 4
*Color of tegulae (black to red).. 4
*Color of pubescence (white to black)... 4
†Breadth of clypeal truncation.. 4
†Ends of clypeal truncation... 4
†Shape of clypeal truncation (smooth to crenulate)...................................... 3
†Sublateral angles of clypeus... 3
†Width of impunctate margin of clypeus.. 3
†Convexity of clypeus... 4
†Spaces between clypeal punctures... 3
†Coarseness of clypeal punctures (by comparison to those of frons)...................... 3
†Direction of subantennal sutures (parallel to diverging upward)........................ 3
†Degree of convergence (or divergence) of inner margins of eyes......................... 3
†Density of clypeal vestiture... 4
†Clypeal brushes (absent to distinct)... 3
†Length of mandible... 4
†Size of upper mandibular tooth (absent to large)....................................... 3
†Length of lower mandibular tooth in relation to others................................. 3
†Apex of labrum (rounded to emarginate)... 2
†Antennal scape (slender to swollen).. 3
†Antennal apex (rounded, pointed, hooked, flattened).................................... 4
†Length of hair of antennal flagellum... 3
†Ratio of length to breadth of antennal pedicel... 4
†Thickness of antennal flagellum.. 3
†Ratio of length to breadth of median flagellar segments................................ 3
†Outer carina of hind coxa.. 2
†Ventral tooth of middle coxa... 3
†Lateral teeth of sixth metasomal tergum (absent to acute).............................. 4
†Median teeth of sixth tergum (absent to twice as long as broad)........................ 5
†Apical flange of sixth tergum (absent to broad).. 3
†Lateral margins of sixth tergum (straight to convex)................................... 3
†Extent that seventh tergum is exserted... 5
†Shape of apex of seventh tergum (medially pointed, rounded, to bilobed)................. 7
†Profile of first metasomal sternum (flat to protuberant)............................... 3
†Size of second sternum in relation to third.. 3
†Profile of disc of second sternum (flat to protuberant)................................ 3
†Apex of second sternum (flat to transversely elevated)................................. 3
†Margin of second sternum (concave to medially pointed)................................. 5
†Margin of third sternum (concave to convex).. 5
†Fringe of third sternum (absent to long across entire margin).......................... 5
†Margin of fourth sternum (concave to convex)... 4
†Fringe of fourth sternum (absent to long).. 3
†Amount that fifth sternum is exposed... 3
†Margin of fifth sternum (concave to convex).. 3
†Fringe of fifth sternum (absent to long)... 3
†Amount that sixth sternum is exposed... 3
†Longitudinal ridge of sixth sternum (absent to produced and expanded).................. 5
†Sixth sternum (planate to duplicated).. 3
†Margin of sixth sternum (convex, to concave, to concave with median point)............. 5
†Subapical lobes of sixth sternum (present or absent)................................... 2
†Margin of seventh sternum (convex to concave).. 5

TABLE 2—*Continued*

†Projection of eighth sternum behind lateral apodemes (absent to over twice as long as broad).... 5
†Curvature of gonocoxites (none to angular)... 3
†Apices of gonocoxites (pointed to truncate)... 3
†Shape of volsellae (short rounded, to long pointed).................................... 3
†Fusion of volsellae to gonocoxites (none to complete).................................. 3
*Distribution (desert only to high montane only)....................................... 8

subjective element and moreover, as explained elsewhere, the large number of characters used swamps the effects of occasional parallelisms.

Another type of coloration used is intensity of blue or green metallic hue. This is entirely independent of the red coloration; actually no species exist in the *Hoplitis* complex which are both red and metallic but this combination occurs in related bees and is presumably not physiologically impossible in the *Hoplitis* complex. On the other hand, we excluded color of the red pigment as a character because, had we used it, one state (lack of red) would have been the same for two characters (extent of red and color of red pigment). In all cases characters were reorganized or omitted to exclude correlations resulting from such situations, where a certain state of one character requires a particular state of another.

Proportions of various structures were widely used. For example, the ratio of the first to the second segments of the labial palpi was divided into four states as follows: greater than 1.0, 1.0 to .75, .74 to .5, less than .5. As with other continuously variable characters, various species fell on the border lines between states or the range of their variation overlapped two states; such species were placed arbitrarily in one or the other of the two states. No effort was made to establish character states that would contain equal numbers of species; actually, in this instance, there are few species in the first and last states and many in the two middle states. Rather, in so far as possible, character states with median points more or less equally different morphologically were employed. Characters of sculpturing (punctation) of the body were han-

dled in a similar manner. For example, clypeal punctation of the males was divided into four states, much finer than that of the frons, finer than the frons, equal to the frons, and coarser than the frons. In such cases, to insure uniformity of judgment, a certain species was selected in advance as typical of "much finer than frons," another as typical of "finer than frons."

Among some of the structural characteristics, such as form of the clypeus of the female, were encountered the only problems in arranging character states in a logical order. In this instance the difficulties disappeared when the clypeal form was broken down into several component elements as shown in the first part of table 2.

No character was included, of course, if it was the same in all species of the *Hoplitis* complex. However, characters were included in cases where only a single species differed from all the rest. There are several such instances of which the following exemplify two types: The inner subapical swelling of the mandible of the female is absent (state 1) in the majority of species, present and rounded or weakly angulate (state 2) in many *Anthocopa*, and produced to a distinct tooth (state 3) in but one species. An outer carina on the hind coxa is absent (state 1) except in one species (which exhibits, therefore, state 2).

After the data were collected, it was apparent that C.D.M. had had a strong tendency to begin the series of states for the various characters with number 1 as the commonest extreme or the extreme state nearest the commonest state. The consequences of this fact and of the variable number of states per character are

discussed in the section on "Choice of a Correlation Coefficient" by Sokal and Michener (1957).

PROCEDURES

Correlational Procedures

As already indicated, the justification for and details of our statistical procedures are being · presented elsewhere (Sokal and Michener, 1957). Here we offer merely a summary statement.

Two obvious ways suggested themselves for condensation and interpretation of the data discussed above. We could either correlate characters with each other (R-technique of factor analysts) or species with each other (Q-technique of factor analysts). The former is the usual technique in biological and psychological studies involving correlational analysis. For our purposes, however, correlations between species based on the characters studied, seems a more useful method. In so far as the characters used are indicative of the entire spectrum of potential variation of the species, we can say that the resulting correlation coefficient is representative of the real affinity between two species. When scanned for clusters of high correlation coefficients, the Q-type matrix (i.e., tabulation of species against species showing correlation coefficients of each species with all of the others) reveals species which are similar. It is thus especially suited to classificatory problems. In spite of this, except for the work of Sturtevant (1942) which involved not correlations but character differences, the authors are unaware of other Q-type studies in the biological sciences.

There are two evolutionary situations under which it is important to examine the two types of matrices. The first might be referred to as breakage of correlation. It occurs when two characters that were correlated in the ancestral line and are still correlated in related lines become independent of each other in a certain evolutionary branch. Under such conditions the R-matrix is a poor representation of the true relation between the two characters. There is no good way of representing such a correlation, close in one line, absent in the other. On the other hand a Q-matrix is not affected by such data.

Convergence of species for a number of characters is a second disturbing phenomenon. Here the R-matrix is not affected while the Q-matrix is affected if the convergent characters outweigh the non-convergent ones in numbers. As explained in the Introduction, we do not believe this is likely if an adequate number of characters is studied. In case of a preponderance of convergent characters and in the absence of paleontological data it is doubtful whether the systematist would be able to distinguish convergence from relationship by descent.

After considering the various possibilities, the product-moment correlation coefficient was adopted as the most suitable, as explained by Sokal and Michener (1957). It might have been desirable to code the data in such a way as to put all character states on the same scale. In a character with two states, the code "2" indicates one extreme for the character, quite unlike code "2" in a character, with, for example, 7 states. This situation is usually met by normalizing rows (characters) of the raw data. We did not perform this transformation for various reasons, including the prohibitive amount of IBM (International Business Machines) work that would have been necessary, since a one-digit code would not have sufficed for normalized data.

The authors are well aware that their methodology of coding and correlation could profit by refinement. It is, however, our point of view that in a pilot study of this nature such refinements are premature. Should the general method prove of value, significant results will surely emerge in spite of minor imperfections in technique.

Even though simplified by the use of a one digit code rather than actual meas-

TABLE 3. *Two-way frequency distribution of correlation coefficients against formal classification codes*

r-class code	r	Formal classification code (c)					Total f_r
		1	2	3	4	5	
1	−07					5	5
2	−02					20	20
3	03				3	36	39
4	08				9	46	55
5	13				20	79	99
6	18			1	27	162	190
7	23			5	51	298	354
8	28			5	140	384	529
9	33			7	150	431	588
10	38			33	181	440	654
11	43			86	189	275	550
12	48		2	121	141	92	356
13	53		13	154	88	22	277
14	58	1	10	120	73	6	210
15	63	5	17	94	59		175
16	68	14	27	53	10		104
17	73	22	32	. 32	4		90
18	78	63	49	17	4		133
19	83	87	36	4			127
20	88	57	16				73
21	93	23					23
22	98	5					5
Total f_c		277	202	732	1149	2296	4656

urements or any other device to express character states, the computation of a 97 species × 97 species matrix of correlation coefficients presents serious technical difficulties. An operation of this magnitude cannot be reasonably undertaken without punched card or electronic computing equipment.

The Matrix of Correlation Coefficients

The 4656 correlation coefficients among the species of our study ranged in magnitude from − .0626 for the correlation between species 26 and 92, to .9747 for the correlation between species 43 and 44.[3] Correlation coefficients equal to or larger than .23 are significant at the one per cent level. As our next procedure, a

frequency distribution of these correlation coefficients was set up with the help of the IBM tabulator. The correlation coefficients were grouped into 22 classes with class intervals of .05. The frequency distribution can be read from the right hand column of table 3. The modal class has a class mark of .38; this represents the most frequent class of correlation coefficients found between species in this study. However there is a second mode at r = .78. This bimodality would indicate that we are dealing with two populations of correlation coefficients: Those indicating close, possibly intrageneric relations and others representing more distant relations. That this is so can be verified from a two-way frequency distribution of the formal classification against the correlation coefficients, shown in the body of table 3. Codes across the top of the table represent relations between species based on the formal classification as

[3] For lack of space the matrix cannot be reproduced here. Microfilm or IBM-tape or card copies can be obtained through the Secretary, Department of Entomology, University of Kansas, Lawrence.

follows:

1. Species in the same species group.
2. Species in different species groups but in the same subgenus.
3. Species in different subgenera but in the same genus.
4. Species in different genera but in the same genus group.
5. Species in different genus groups.

It is clear that a large part of the second peak is formed by relations coded 1, i.e., between members of the same species group. It was encouraging to find that magnitude of the correlation coefficient was apparently an estimate of systematic relationship as indicated by the formal classification. This was confirmed by a calculation of the correlation coefficient between the computed correlation coefficients and the codes for the formal classification which was − .80.

Another way of examining these correlation coefficients is to study frequency distributions of the coefficients for any single species against all other species. Figure 4 shows histograms of such distributions for species 1, 4, 6, 19, 29 and 92. There are marked differences in form between the histograms. Species 1, 4, 6, 19 and 29 have two modes and species 92 has three modes. The principal modes are at .33 for species 1, 19 and 92 and at .38 for species 4, 6 and 29. The secondary mode is highest in species 6 (.78), lower in 1, 4 and 19, and lowest in 92 (.53). This latter species has a third mode at .86. The scatter is lowest in species 1 and 19, increased in species 4, 6 and 29 and highest in 92.

On the basis of these findings we can distinguish at least three types of species. The first type is a member of a closely related group of species. It therefore has a number of correlations considerably higher than the majority of its correlations. Species 6 and 92 with their high secondary modes are examples of this type and subsequent work has borne out this view. Close inspection reveals dif-
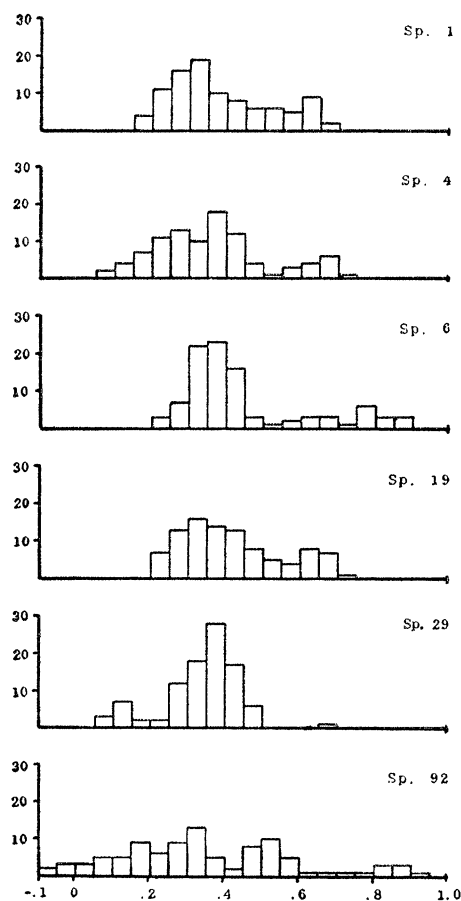

Fig. 4. Histograms showing levels of correlation coefficients for each of six species against all other species.

Abscissas: levels of r for given species. *Ordinates:* frequency, number of r's per class.

ferences between the two. Species 92 has three distinct modes, while species 6 has only two with a slight indication of a third one. Subsequent studies showed that species 92 was a member of a closely knit species group with a considerable gap between this group and the rest of its genus (see fig. 8). Species 6 is also a member of a species group but transition to more distant relations in the same genus is more gradual than in the former case (see fig. 5). It is interesting to note the tail of low correlations shown by species 92, while the correlations of species

6 are sharply truncated at the lower end. It would appear that species 6 is closer to the general type of the *Hoplitis* complex than is species 92.

The second type of species is illustrated by 1, 4 and 19. They are isolated species moderately well correlated with the other members of their genus, as illustrated by their lower secondary modes. Species 4, by its greater scatter, appears to deviate more from the norm than the other two. Reference to figure 5 will confirm these views.

The third type is the very isolated species, which is possibly in a genus by itself. The histogram of species 29 shows only one relatively high correlation (with species 30 at .60). The remaining correlations are far below that value. Our subsequent study suggested that 29 and 30 · may be different enough from the other species in the *Hoplitis* complex to warrant generic distinction (see fig. 6).

The absence of significant negative correlations from our matrix requires some discussion. Q-technique matrices of correlations between people (based on psychological tests) are quite likely to yield such correlations. If there are distinct, antithetical types of persons represented in the matrix, such as extroverts and introverts, it is likely that a high score for one type will be a low score for the other and vice versa. In our case evolutionary progress may be represented by either an increase or a decrease in state codes. In the majority of characters the supposedly primitive situation is an intermediate state code with two diverging evolutionary trends represented by the lower and higher code numbers. Furthermore, characters representing correlated trends were not necessarily coded on the same scale. It is clear that under such circumstances distantly related forms are likely to be uncorrelated rather than negatively correlated.

The Search for Group Structure

The matrix of correlation coefficients obtained as reported above can be put to a variety of uses and the analysis reported below represents merely an initial effort at exploitation of the data. The correlation coefficients serve as an absolute measure of relationship between any two species in our study, limited only insofar as the characters chosen do not represent the total correlated variation of the two species. Serious study of the matrix will center around two aspects: (1) a search for group structure (2) an analysis of the evolutionary factors in the systematic group under study. The second point will undoubtedly require some method involving the use of partial regressions, be this factor analysis, path coefficients or some other device. This study is presently being initiated and the authors are optimistic about the significance of the outcome (see Sokal, 1957).

The search for structure among the correlation coefficients of the matrix is of course no different in aim from the search by the systematist for a natural system in an array of species. Such a system consists of a hierarchy of groups. Various methods can be used for discovering a hierarchy in data such as ours. A customary, rather simple device of the psychometrician is so-called "cluster analysis," developed to a fine art by Tryon (1939).

Attempts to employ cluster analysis for finding structure in our matrix were only partially successful, since the resulting clusters were partly overlapping, i.e., a given species might be a simultaneous member of two clusters. This makes good sense for intermediate forms in an abstract scheme of relationships. In a systematic hierarchic classification, however, groups at the same level have to be mutually exclusive for practical as well as for theoretical reasons, except for low level groups exhibiting reticulate evolutionary patterns (rare above the species level in animals). A further reason for the unsuitability of cluster analysis is the complexity of the clusters as more species are added to them. Although clusters are therefore not convenient in an initial

search for structure, the diagram of relationships established by methods to be described below could be easily recognized in the clusters outlined by cluster analysis. A method essentially similar to cluster analysis is the ρ-group and ρF-group method of Olson and Miller (1951) applied to three paleontological R-technique matrices. It suffers from the same drawbacks as cluster analysis.

In view of the disadvantages of this and certain other methods, we decided on the following procedure: a nucleus of a group was established, using the two species having the highest coefficient of correlation. Then species would be added to this nucleus, one at a time, always adding first the species with the highest average correlation with the members of the group. The limit of the groups could be found by decreases $(\bar{L}_{n+1} - \bar{L}_n)$ in the level of the average correlation $\bar{L}_n$, where the subscript refers to the number of members in the group. A significant drop was empirically determined since sampling distributions of average correlations,
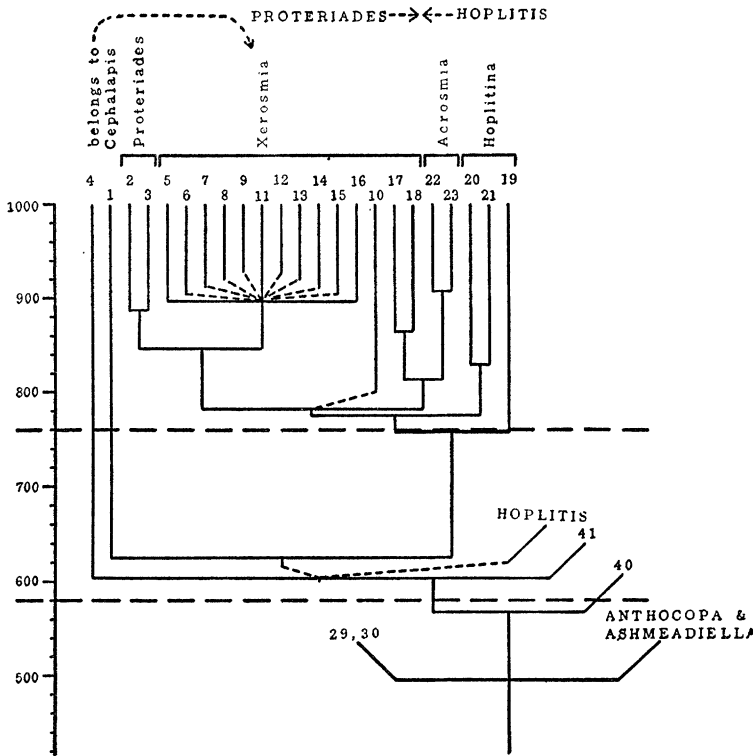

FIG. 5. Diagram of relationships for the genus *Proteriades* obtained by the weighted variable group method.

Ordinate: magnitude of correlation coefficient multiplied by one thousand. Correlations between any two joining stems can be found by reading the value on the ordinate corresponding to the horizontal line connecting the stems. This value becomes approximate and maximal in cases of multifid furcations. Broken lines where more than 2 lines join, are for convenience only; the horizontal connecting line has the same significance as elsewhere. "Roofs" over the species numbers at the tips of the lines delineate subgenera containing more than one species. These subgenera, identified in lower case print, are based on C. D. M.'s formal classification and not on the findings of this study. Generic names are capitalized. The boundary between the genera *Proteriades* and *Hoplitis* according to the older classification is indicated by two arrows meeting head-on at the top of the diagram. Horizontal broken lines are explained in the text, under Classificatory Problems.
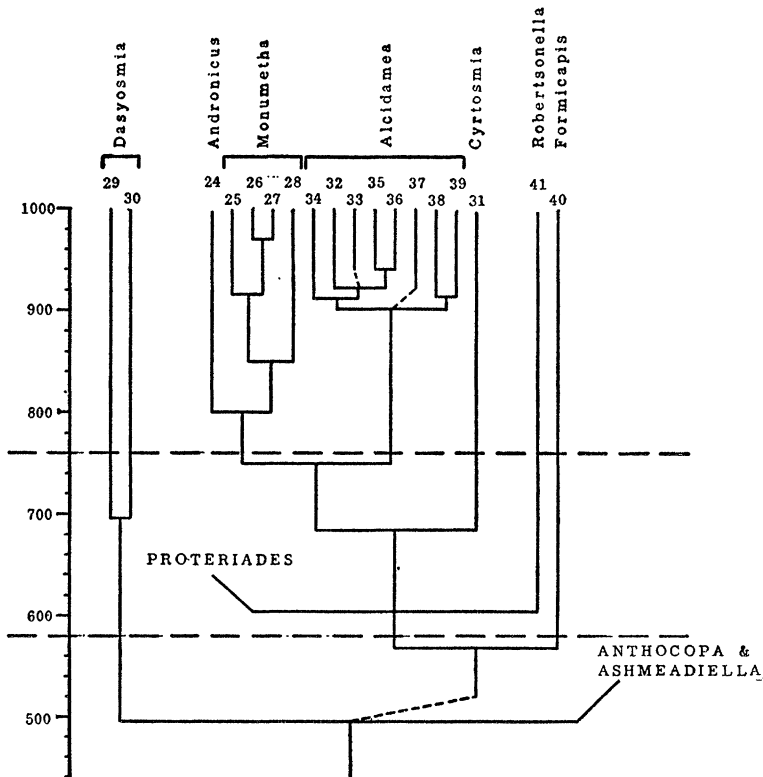
FIG. 6.   Diagram of relationships for the genus *Hoplitis* obtained by the weighted variable group method.
Explanatory comments as for Figure 5.

such as $\bar{L}_n$, are unknown.   First, lower groups (species groups) were developed. Then a new matrix of correlation coefficients, based on the groups as well as isolated species which had failed to enter any group, was calculated.   Then by the same method, it was possible to unite these into larger groups (sometimes subgenera), and these into still larger or higher groups, etc.   Thus it has been possible to develop a hierarchy of groups which is represented below by means of diagrams of relationships (figs. 5–8) for comparison with figures 1 and 2.

We might have admitted only one new member for each group at a given hierarchic level, thus obtaining a diagram of relationships consisting of bifurcations only (pair group method).   However, we decided that this was an artificial situation and, moreover, we are forced into multifid furcations by some decreases in the average correlation coefficients, too small to plot.   Thus a method using groups of various sizes, called a variable group method, was selected as more reasonable and practical than the pair group method.   A second consideration was how to weight the variables (species or groups) on calculating a new correlation matrix following each grouping procedure.   If A and B represent closely related species, C a more distant species, the problem is, should a still more distant species, D, be placed in relation to a homogeneous group ABC or in relation to the AB-C stem of a diagram of relationships?   The latter alternative was selected, so that not all species have the same weight in constructing the diagrams
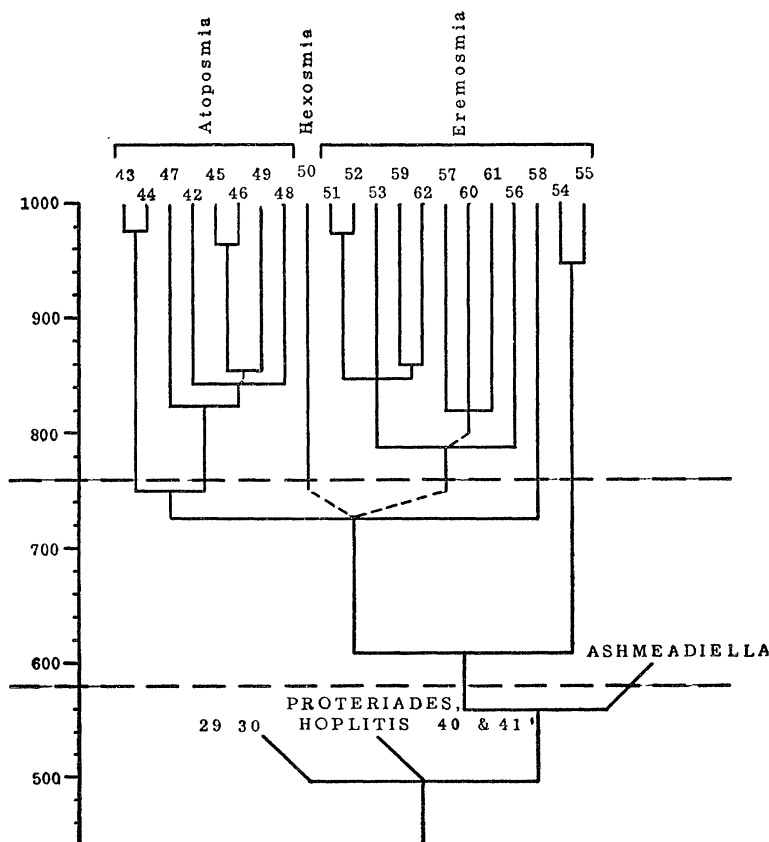
FIG. 7.   Diagram of relationships for the genus *Anthocopa* obtained by the weighted variable group method.

Explanatory comments as for Figure 5.

of relationships. We therefore termed our technique the "weighted variable group method." The greater part of another paper (Sokal and Michener, 1957) is devoted to a detailed explanation of this method so that others can readily follow it.

Once computed, the relations were represented as diagrams of relationships in figures 5–8, one diagram for each genus. The ordinate at the left of each diagram is graduated in units of 1000 times the correlation coefficient. The correlations between any joining stems in the diagram can be read by measuring the level along the ordinate of the horizontal line connecting the stems. Thus in figure 8, species 63 and 64 are correlated at a level of .908, while group 63–64 is related to group 67–72 at .702. Furcations involving more than three lines are shown by broken lines converging on the midpoint of the horizontal line as in group 88–94 of the above figure. This was done to avoid any impression that, for example, 88 and 89 are closer than 88 and 94. The tops of the figures are at a level of 1000 (correlation of 1) since obviously each species is perfectly correlated with itself. These diagrams represent relationships in a static, non-historical sense.
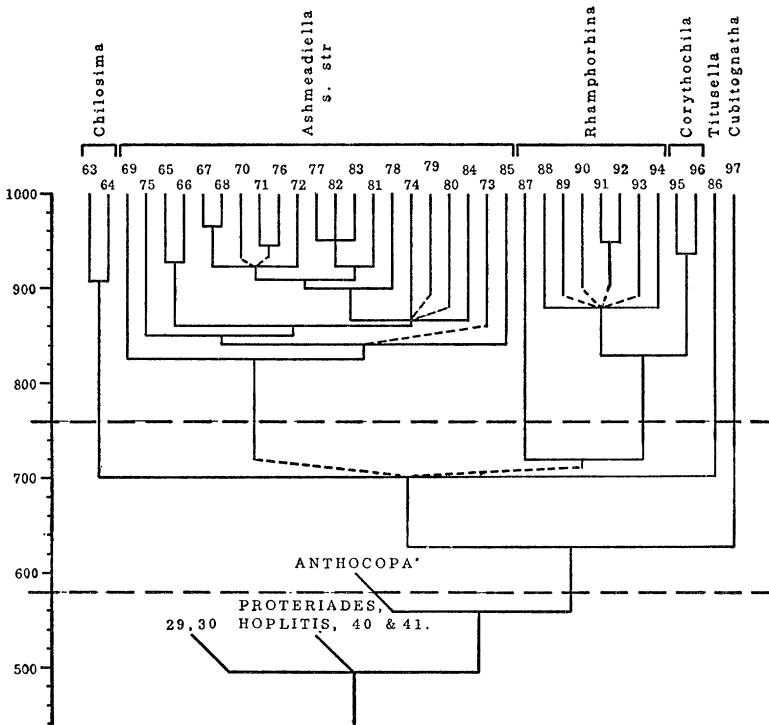
FIG. 8. Diagram of relationships for the genus *Ashmeadiella* obtained by the variable group method.

Explanatory comments as for figure 5.

COMPARISON OF DIAGRAMS OF
RELATIONSHIP

Relationships

A comparison of table 1 or figures 1 and 2, the older classifications (by C.D.M.), with figures 5–8 shows that the correspondence between the classical and the quantitative methods is good, with certain exceptions discussed below. The diagrams of relationships shown in figures 5–8 differ from those of figures 1 and 2 in that the vertical distances indicate degree of relation (or of divergence) as shown by the scale of average correlations ($\times$ 1000) at the left. Thus the degree of difference between any two forms can be judged by measuring from the top of the figure to the level where their branches meet.

The following is a list of the main changes indicated by the new diagrams (figs. 5–8) upon comparison with the old ones (figs. 1–2):

1) 17–18 closer to 22–23 than to 5–16.
2) 19 a separate group from 20–21.
3) 1 and 4 more isolated from *Proteriades* in general.
4) New boundary between *Proteriades* and *Hoplitis,* the subgenera *Hoplitina* and *Acrosmia* being transferred to *Proteriades.*
5) More structure in junction of *Dasyosmia-Cyrtosmia* group of *Hoplitis.*
6) *Dasyosmia* extremely isolated.
7) 40 and 41 in reversed positions in relation to the main stem of *Hoplitis.*

8) 54 and 55, 86, and 97 more isolated.

9) More (and sometimes different) structure in *Monumetha, Alcidamea, Atoposmia, Eremosmia, Ashmeadiella* s. str., and *Rhamphorhina*.

10) 50 not as isolated as claimed.

There are several instances (listed in item 9 above) of groups not or but little subdivided in figures 1 and 2 which now show more structure; that is, we now have evidence of the varying degrees of relationship among the species of the groups. In part this is merely due to the fact that in the original diagram no effort was made to show relationships below the level of species group. In *Monumetha*. the relationships shown in figure 9 are the same as those that would have been suggested had a "tree" for the species of this group been attempted before the statistical study. In the other groups, usual systematic methods did not provide the worker (C.D.M.) with the bases for attempting diagrams of relationships. Certain species (e.g., 67, 68) were known to be closely related, so that it is not true that these groups were "structureless" in the eyes of the systematist, but it was not obvious how the pairs and little groups of very similar species, together with the more isolated species, were related. It is also true that the three groups recognized in *Ashmeadiella* s. str. (fig. 2) were very informal and the failure of the statistical method to support them is not surprising to C.D.M. In short, the diagrams for the groups listed in item 9 above, except for certain matters in *Eremosmia* discussed later, seem satisfactory and shed new light on some of the relationships. The same may be said also of the relationships between the subgenera *Dasyosmia, Andronicus, Monumetha, Alcidamea,* and *Cyrtosmia* (item 5 of above list), shown as equally related in figure 2. Figure 6 provides an idea of the relationships of these groups which seems acceptable to

C.D.M. The most striking fact here is the isolated position of *Dasyosmia*.

The new location of the dividing line between *Proteriades* and *Hoplitis* (item 4, above list) is not at all surprising. The old diagram of relationships (fig. 1) shows *Proteriades* as a mere offshoot of *Hoplitis,* having close relationships (pointed out by Hurd and Michener, 1955) to *Hoplitina* and *Acrosmia,* which were placed as subgenera of *Hoplitis. Proteriades* was separated as a rather large group having specialized hairs on the galeae and labial palpi for extracting pollen from the slender flowers of *Cryptantha*. This was a striking and convenient character and was correlated with some others, but the discovery of males of 17 and 18, which nearly or entirely lack such hairs, already weakened it.

There are several instances where a species or small group is shown as very much more isolated by our statistical method than by the classical methods of systematics. This is noted in items 3, 6, and 8 of the above list and item 2 may represent the same phenomenon. We have selected species 4 as a basis for further discussion. There is no doubt that this is indeed a strange insect and that it differs by many characters from its relatives. However, one of us (C.D.M.) maintains that in spite of its numerous unique characters it shows definite evidence of relationship to species 1–16 and that it did not branch from the stock ancestral to all the other *Proteriades* as our method (fig. 5) might suggest. The bases for this contention are that species 4, like the rest of the group 1–16, has the following characters: (1) proboscis unusually short; (2) galeae and labial palpi provided with hooked hairs in both sexes; (3) maxillary palpi short; (4) middle trochanter of the male provided with toothlike lamella; and (5) the hind coxae provided with a strong carina. That this combination of characters arose independently, in parallel fashion, for species 4 and the other species from 1–16 is possible,

but unlikely. Of course the characters listed above are only some of those in which species 4 agrees with the others mentioned; numbers (1), (2) and (4) are peculiarly significant because they do not occur in the *Hoplitis* complex except among the species 1–16.

Our method shows species 4 to be quite distantly related to other members of the *Hoplitis* complex. However, the method cannot indicate whether a species differing greatly from others does so because it is a *pregroup* derivative, i.e., a form which has branched from the ancestral stem of a group; or because it is an *exgroup* species, i.e., a highly specialized derivative of the group in question with many features peculiar to it. The difference between these conditions is shown diagrammatically by phylogenetic trees in figure 9. In figure 9A, species X is shown as a derivative of a primitive stem; in 9B species X is a specialized member of a cluster of related species. The correlation coefficients between X and members of the cluster in 9A and between X and members of the cluster in 9B might be equally low. Yet the phylogenetic relationships are very

TABLE 4. *Relations among relatives of isolated species*

Isolated species		Relative number:											
		1	2	3	4	5	6	7	8	9	10	11	12
1.	Relative species	2	14	8	15	13	9	17	11	16	7	5	3
	r_{1j}	.69	.69	.65	.65	.64	.63	.62	.62	.61	.61	.61	.60
	baseline		.85	⟶				.78	⟶				
4.	Rel. sp.	2	6	8	14	15	16	9	5	11	13	12	7
	r_{4j}	.71	.70	.70	.69	.69	.67	.67	.65	.65	.63	.62	.60
	baseline		.85	⟶									
19.	Rel. sp.	17	5	22	14	20	23	18	11	6	13	16	12
	r_{19j}	.74	.69	.68	.68	.68	.68	.67	.66	.65	.65	.65	.64
	baseline		.78	⟶									
29.	Rel. sp.	30	24	27	34	35	36	52	38	59	62	25	32
	r_{29j}	.70	.49	.49	.47	.46	.46	.46	.45	.45	.45	.44	.44
	baseline		.50	⟶									
31.	Rel. sp.	27	25	26	37	28	32	24	34	35	33	36	38
	r_{31j}	.70	.69	.68	.62	.61	.60	.58	.57	.56	.56	.56	.56
	baseline		.92	⟶	.75	⟶							
35.	Rel. sp.	36	32	33	34	37	39	38	27	25	24	26	28
	r_{35j}	.94	.93	.92	.89	.85	.84	.83	.63	.61	.60	.60	.58
	baseline		.92	⟶	.91	.90	⟶		.75	⟶			
40.	Rel. sp.	28	25	26	27	17	37	50	41	94	18	1	24
	r_{40j}	.61	.58	.57	.53	.52	.52	.51	.50	.49	.49	.48	.47
	baseline		.85	⟶		.61	⟶	.50	⟶				
41.	Rel. sp.	25	19	18	17	27	39	34	28	35	5	26	33
	r_{41j}	.60	.58	.55	.55	.55	.54	.53	.53	.52	.52	.52	.52
	baseline		.61	⟶									
43.	Rel. sp.	44	42	45	46	49	48	47	61	51	60	53	52
	r_{43j}	.97	.80	.73	.72	.67	.66	.64	.61	.59	.58	.57	.55
	baseline			.84	⟶				.72	⟶			

TABLE 4—*Continued*

Isolated species		Relative number:											
		1	2	3	4	5	6	7	8	9	10	11	12
50.	Rel. sp.	51	48	58	84	52	45	59	46	80	47	74	77
	r_{50j}	.73	.72	.70	.68	.66	.65	.65	.64	.64	.63	.62	.61
	baseline		.72	→	.56								→
54.	Rel. sp.	55	80	77	85	82	73	79	59	72	58	75	81
	r_{54j}	.95	.65	.64	.63	.62	.62	.62	.61	.61	.60	.59	.59
	baseline			.56									→
58.	Rel. sp.	52	51	50	59	62	53	56	61	57	60	54	80
	r_{58j}	.76	.74	.70	.69	.68	.67	.63	.61	.61	.60	.60	.59
	baseline		.98	.73							→	.61	.56
63.	Rel. sp.	64	70	96	76	67	71	68	95	78	72	65	80
	r_{63j}	.91	.75	.70	.70	.69	.68	.68	.68	.68	.67	.65	.64
	baseline		.70										→
69.	Rel. sp.	72	68	85	82	81	67	70	75	79	77	83	76
	r_{69j}	.82	.82	.80	.79	.78	.77	.77	.77	.77	.76	.74	.74
	baseline		.92	.84									→
86.	Rel. sp.	84	74	83	80	77	79	71	81	67	78	73	76
	r_{86j}	.75	.75	.70	.69	.67	.66	.65	.64	.64	.64	.64	.63
	baseline		.86								→	.84	→
97.	Rel. sp.	92	89	91	63	64	90	93	96	72	88	95	65
	r_{97j}	.66	.63	.61	.60	.60	.60	.59	.59	.58	.58	.58	.58
	baseline		.88	→	.70								→

different. For convenience we refer to this problem as the "pregroup-exgroup problem."

A consideration of this matter and of the correlation coefficients among species of the *Hoplitis* complex suggests that
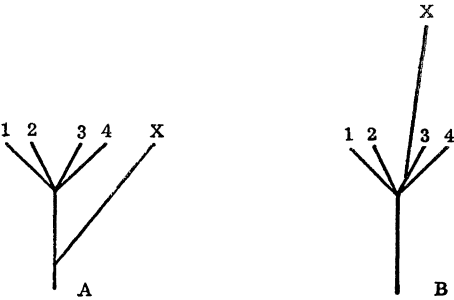

FIG. 9. Hypothetical diagrams of relationships to illustrate the problem of differentiating between primitive and specialized differences, or pregroup and exgroup species.

For explanation see test.

quantitative ways of solving this dilemma may exist for particular species. An array of species most closely related to a given isolated species can be listed in order of decreasing magnitude of their correlation coefficients with that species. The "base line" relating these relatives to each other can be read off figures 5 to 8 as the highest horizontal line connecting the species concerned. Clearly, as more species are admitted to the group of relatives the base line level is bound to decrease. The original level of the base line, the presence or absence of a decrease and its magnitude appear to reveal phylogenetic relationships of the isolated species. For purposes of analysis we have chosen an arbitrary number of 12 nearest relatives for each of 16 isolated species and listed them in decreasing order of correlation with the isolated species in table 4. Line 1 for each isolated species
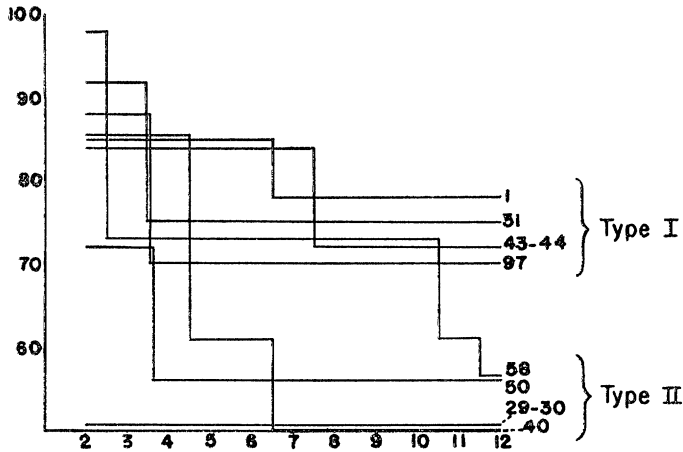
FIG. 10. Base lines of relatives of isolated species types I and II (data from table 4).

Abscissa represents decreasing rank of correlation of isolated species with related species. *Ordinate* is in correlation coefficient units multiplied by one hundred. For explanation see test.

identifies the code numbers of the most closely related species while line 2 gives the value of the correlation coefficient of the relative with the isolated species concerned. Line 3 shows the level of the base line for the group of relatives up to and including the species listed in line 1. Thus in the case of species 4 the base line for species 2 and 6, its closest relatives, is .85 and remains at that level upon addition of the next ten relatives (species 8, 14, 15, . . . 7).

Figures 10 and 11 show the levels of the base lines of table 4 in graphic form. On careful study the lines fall into a number of more or less definite types. Type I, as exemplified by species 1, 31 and 97 and group 43–44, includes species whose nearest relatives are themselves a closely knit group, while more distant relatives (among the 12 studied) are still fairly highly related. Species 1, for example, is clearly closer to the 2–16 group (of fig. 5) than it is to group 17–18. There is,
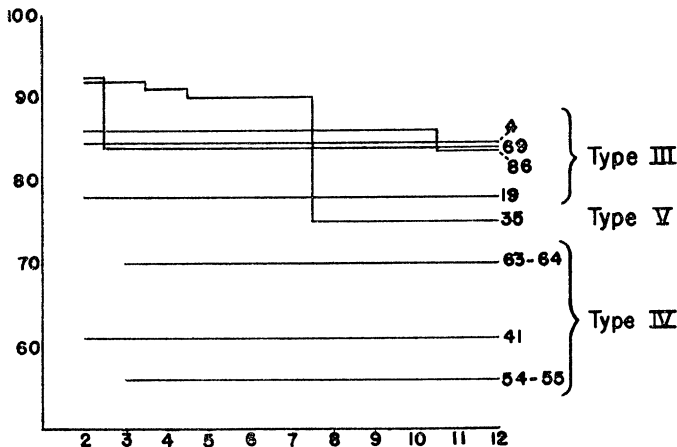

FIG. 11. Base lines of relatives of isolated species types III and IV and of a non-isolated species, type V (data from table 4).

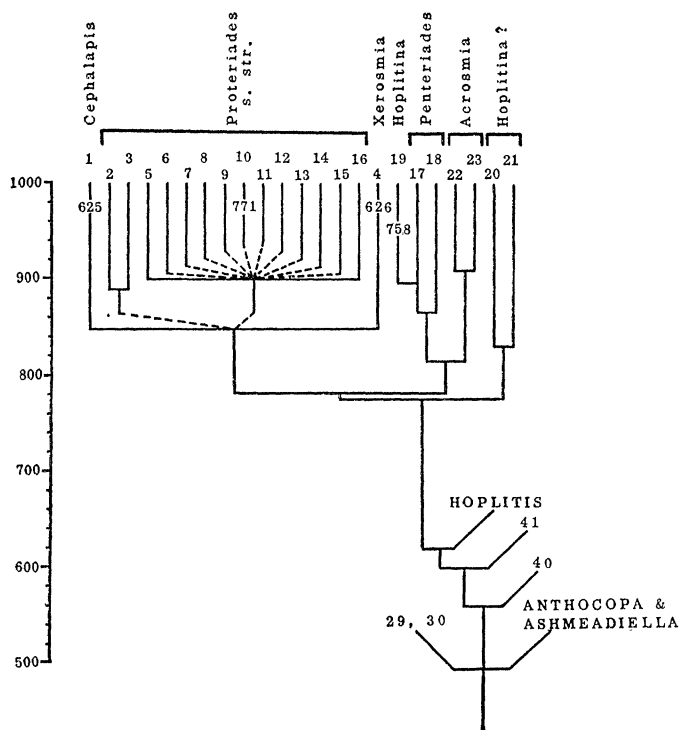*Abscissa* and *ordinate* as in figure 10. For explanation see text.

FIG. 12. Diagram of relationships for the genus *Proteriades* modified from figure 5 after consideration of the pregroup-exgroup problem.

Isolated species 1, 4, etc., shown as pregroup species in figure 5, are now shown as exgroup derivatives of various stems. Numbers inserted in the lines leading to these species indicate the levels, in terms of the scale at the left, at which the lines for these species joined their relatives in figure 5. These numbers therefore give an indication of the unusually large amount of specialization that has occurred in these species.

however, no evidence linking it to either 2–3 or 5–16 in preference to the other group. While the previous study had shown that 2–16 and 17–18 were much closer to each other than they were to 1, and consequently species 1 had been depicted as isolated, the present findings indicate that 1 has closer affinities to the 2–16 stem than to the 17–18 stem. This suggests that species 1 is a greatly modified offshoot from the 2–16 stem. In relation to species 1 the pregroup-exgroup terminology is not too appropriate. Species 1 is pregroup to groups 2–3 and 5–16, but exgroup to 2–16. As long as the correct relations are depicted, however, the terminology is secondary; moreover all intermediates must exist between pregroup and exgroup species. Reasoning

similar to that employed in the case of species 1 leads us to regard species 31 as a modified offshoot of the 25–27 stem, species 97 as derived from stem 88–94, and group 43–44 as derived from stem 47–48, possibly from species 42.

In type II, including species 40 and 50 and perhaps species 58, as well as group 29–30, the near relatives of the isolated species are not too closely related while the more distant relatives are equally distant from each other. The isolated species concerned would more likely be pregroup species to their closer relatives. Species 40 appears closest to stem 25–28, while species 50 appears to have been placed correctly by the variable group method. Species 58 apparently branched from the 51–52 stem. Group 29–30 may
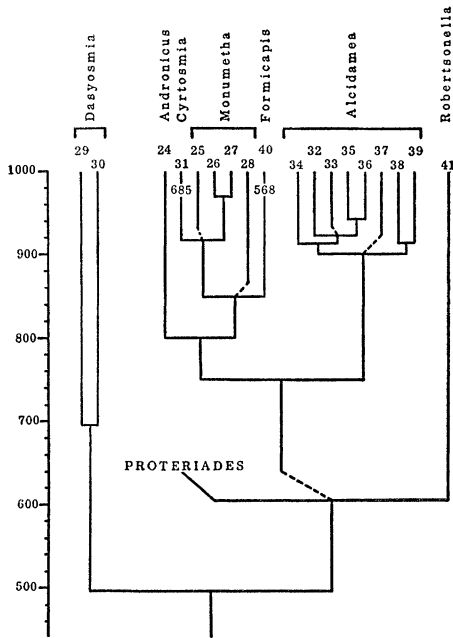
FIG. 13. Diagram of relationships for the genus *Hoplitis* modified from figure 6 after consideration of the pregroup-exgroup problem. (*Anthocopa* and *Ashmeadiella* stem omitted by error at level 500.)

Explanatory comments as in figure 12.

be a much altered branch off the 24–39 stem.

Type III includes species 4, 19, 69 and 86. Each of these seems equally and fairly highly related to a large number of species indicating that they originated from a large species group. Species 4 appears equally closely related to 2–3 and to 5–16 and is therefore represented as originating from the 2–16 stem. Species 19 has close relations to the 5–16 and 17–18 groups but curiously not to 2–3. Since it has a higher correlation coefficient with species 17 than with any other species it might be suggested that it originated from an ancestor of that species. Species 69 appears closest to 67–72, while 86 shows relations with 67–84.

We have designated as type IV species such as 41 or groups such as 54–55 and 63–64, which do not show close relations to any species. Those species to which

they are somewhat related are not in a group by themselves but are scattered throughout the study. Such evidence would indicate that species of this type are isolated pregroup species. Their positions on the diagrams of relationships, therefore, have not been changed.

Finally species 35 was chosen as an example of a non-isolated species, type V. An analysis similar to the ones carried out above corroborates the position assigned by the variable group method.

New diagrams of relationships were constructed (figs. 12–15) based on a composite view of morphological resemblance as shown by the variable group method and relationship by descent as found by the above procedure. These diagrams are reasonably close to the original figures 1 and 2, indicating a surprising amount of agreement between the subjective and the more objective methods.

It should again be emphasized that the variable group method cannot by itself supply a diagram of relationships by descent. It can indicate static, nonhistorical relationships only. However, when another technique is applied in conjunction with the variable group method a diagram of relationships which is phylogenetically meaningful can be obtained.

Classificatory Problems

A difficulty arising in any large systematic study is the question of assigning hierarchic rank to taxonomic categories. A systematist has two purposes in mind when establishing a hierarchy of categories: (1) He would like to set up a so-called natural system of relationships which would indicate *both* common descent as well as degree of diversification (or amount of evolutionary change subsequent to the divergence of phylogenetic lines); (2) he uses categories as a convenient means of organizing his information (i.e., the many different species with which he has to deal). It is regrettable but true that the present system of nomenclature cannot easily meet both purposes. Moreover even in purpose (1)
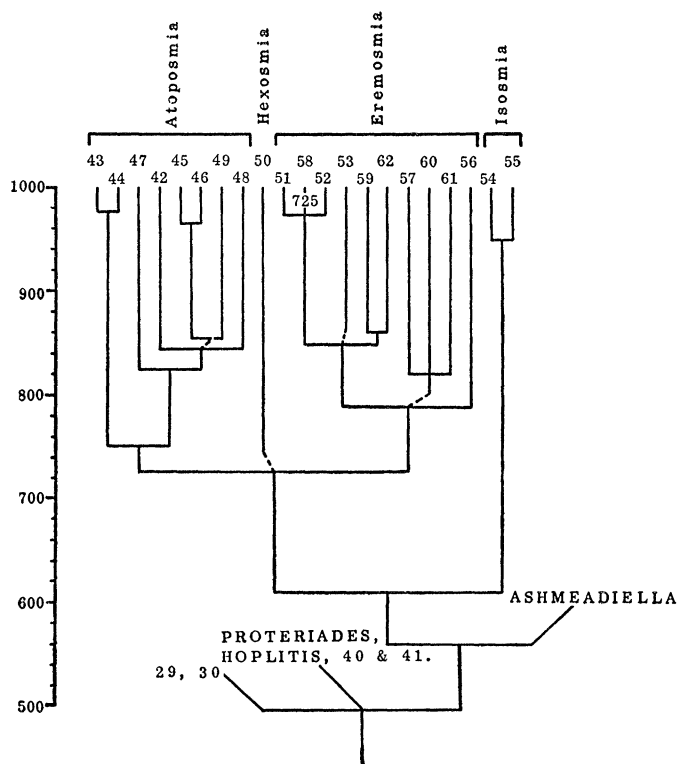
FIG. 14. Diagram of relationships for the genus *Anthocopa*, modified from figure 7 after consideration of the pregroup-exgroup problem. Explanatory comments as in figure 12.

the present nomenclatorial system cannot always show both common descent and evolutionary divergence. Thus, referring to figure 9, it will be found difficult to devise a nomenclature to differentiate situations A and B. If we call both A and B genera, for example, then we reveal nothing of the different origin of X in the two diagrams. On the other hand calling 1–4 and X separate genera in A but the same genus in B implies that X is closer to 1–4 in B than it is in A, which is true from the viewpoint of time of divergence, but not from that of evolutionary change.

The two authors have divergent views on this issue. C. D. M. would prefer a more or less subjective approach, utilizing information gained by the statistical method but modifying it with knowledge of probable actual phylogenies obtained in a variety of ways. He considers the

convenience of a classificatory scheme important. For example, he would prefer for the sake of convenience, that the size of groups recognized be more or less inversely proportional to the degree of difference between them. He is joined in these views by many present day systematists whose position might be summarized in the words of Mayr, Linsley and Usinger (1953, p. 48): "A genus is a systematic category including one species or a group of species of presumably common phylogenetic origin, which is separated from other similar units by a decided gap. It is suggested for practical reasons that the size of the gap be in inverse ratio to the size of the unit. The latter qualification will prevent the recognition of unjustified monotypic genera." Thus while *Proteriades* (species 1–24) was labeled a genus by C. D. M.,

FIG. 15.   Diagram of relationships for the genus *Ashmeadiella,* modified from figure 8
after consideration of the pregroup-exgroup problem.
Explanatory comments as in figure 12.

Dasyosmia (species 29–30), which is equally different from the other genera, was not accorded generic rank.

On the other hand, R. R. S. would like to see uniform and objective standards applied to the recognition of categories. Not being a practicing taxonomist he would prefer that the classification reflect the quantified relationships rather than considerations of convenience. R. R. S. feels that the only method that even approximates a realistic description of relationships is a pictorial one and that the hierarchy of categories can try to indicate either descent or amount of divergence (or relationship) between stems, but not both. He maintains that in the example cited above the nomenclature reflects descent in one case (by including *Dasyosmia* in the genus *Hoplitis*) but indicates divergence in another case (calling *Pro-*

teriades, which also descended from *Hoplitis,* a separate genus). Rather than use categories in this ambiguous manner he would prefer to utilize them to show divergence or relationships only, leaving the description of descent to diagrammatic representation. He would therefore urge both "lumping" and "splitting" wherever indicated by the variable group analysis, i.e., he has no *a priori* objections to either monotypic or inordinately large genera or subgenera.

Proceeding on the above line of reasoning R. R. S. drew two arbitrary horizontal lines across figures 5 to 8. The levels of these lines, which are the same for all four figures, were determined by having them cut the ordinate at a point which would result in the number of categories (i.e., genera and subgenera) closest to the number established by

C. D. M. prior to the present study. The upper broken line might thus be called the subgeneric line and the lower, the generic line. By this method he would establish five genera: a joined *Proteriades-Hoplitis, Dasyosmia, Formicapis, Anthocopa* and *Ashmeadiella.* A subgeneric classification based on the upper line would include more monotypic subgenera than before (about half of them) as well as one newly constituted very large subgenus resulting from joining of most of (old) *Proteriades* with *Acrosmia* and *Hoplitina.*

C. D. M. as the advocate of the subjective approach, however, maintains that size of gaps as well as degree of difference is important and variable. He finds complications such as in the first matrix where species 54 and 55 had higher cor- relations with species 80 (in *Ashmeadiella*) than with any other species of *Anthocopa.* Obviously the intergeneric gap here is small. He would therefore prefer the flexibility of the subjective approach which permits consideration of degree of difference, size of gaps between groups, number of species in the groups, etc.

Taxonomic Consequences

The taxonomic consequences of the application of the objective approach can be determined by anyone who wishes to study figures 5 to 8 and 12 to 15. The classification proposed below is based on an admittedly subjective interpretation, but it makes extensive use of results of the objective statistical work summarized in figures 12–15.

In a number of cases a final decision was delayed because of excluded species which might shed light on certain dubious relationships. This classification should be examined in relation to table 1.

1. Genus *Proteriades.* Retained as a genus and expanded to include *Hoplitina* and *Acrosmia.* Can be distinguished from *Hoplitis* in its new, more restricted sense by the apical process, bearing spe- cialized hairs, on the sixth metasomal sternum of the male; near absence of an apical flange on the sixth tergum of the female; broader and shallower basal sul- cus on the first tergum than in *Hoplitis,* etc.

a. Subgenus *Cephalapis:* Unchanged. Seems to be an exgroup species from *Proteriades* s. str. but is sufficiently different to merit a name.

b. Subgenus *Proteriades* s. str.: Expanded to include group B of subgenus *Xerosmia* (table 1). Species 10 is merely a specialized "exgroup" member of this group.

c. Subgenus *Xerosmia:* Includes species 4 (group A) only, which has been amply demonstrated to be very different from the 5–16 (B) group. It is probably an exgroup species from *Proteriades* s. str. but is so different as to merit a name.

d. Subgenus *Penteriades* new subgenus. Type species: *Osmia remotula* Cock- erell. Includes species 17 and 18 of *Proteriades.* Can be distinguished from other *Proteriades* by the five- segmented maxillary palpi, the ab- sence or near absence of hooked or wavy hairs on the galeae and labial palpi of the males, and the bulging upper part of the clypeus of the females.

These bees are doubless similar to the primitive stock from which spe- cies 1 to 16 arose. They are also similar to *Acrosmia,* whose species evolved along quite different lines.

e. Subgenus *Acrosmia:* Unchanged in content.

f. Subgenus *Hoplitina:* Unchanged in content.

From figures 5 and 12 it appears that this group should be divided. We are inclined to take this action, but have delayed until we obtain both sexes of species known from only one sex, which were excluded from our study and which may be important in this decision.

2. Genus *Hoplitis. Hoplitina* and *Acrosmia* removed to *Proteriades.*

The species 41, 40 and 29 and 30 are quite atypical. However, the relations of *Hoplitis* to *Anthocopa* should be further investigated in Eurasia before a decision on the status of these species is reached. At the moment it appears most useful to recognize *Hoplitis* as a large primitive group from which *Anthocopa* arose at a rather low level on the diagram, *Proteriades* at a higher level. Subgenera of *Hoplitis* remain unchanged in content.

3. Genus *Anthocopa*. Unchanged in content.

a. Subgenus *Atoposmia:* Unchanged. Species 47 is probably an exgroup member of group 42–48.
b. Subgenus *Hexosmia:* Unchanged. .
c. Subgenus *Eremosmia:* Species 54 and 55 excluded. Species 58, probably an exgroup member, is included in *Eremosmia.*
d. Subgenus *Isosmia,* new subgenus: Type species: *Anthocopa rubrella* Michener. Includes species, 54 and 55; can be distinguished from other *Anthocopa* by having the form and appearance of *Ashmeadiella;* second to fourth metasomal sterna of male subequal in size; apical margin of sixth metasomal tergum of female in the form of an unusually broad flange.

4. Genus *Ashmeadiella*. Unchanged in content.

a. Subgenus *Chilosmia:* Unchanged.
b. Subgenus *Ashmeadiella: Titusella* placed as synonym by C. D. M. since 86 appears to be an exgroup representative of 67–73. Species 69 is doubtless an exgroup form, phylogenetically related to the large central group of the subgenus (see fig. 15).
c. Subgenus *Titusella:* Merged with *Ashmeadiella.*
d. Subgenus *Rhamphorhina:* Subgenus *Corythochila* placed as synonym.

Species 87 is included for the present. It is isolated and has primitive features (a pregroup species) but is connected to 88–94 by a few species known only as females and hence excluded from our study.

e. Subgenus *Cubitognatha:* Unchanged because of very distinctive features although 97 is an exgroup representative of 88–94.

Applicability and Objections

Applicability of Method

A number of implications and applications of our procedure as well as some objections to it have been discussed in previous sections. It remains for us to enter upon the general question of applicability of the method and to discuss several objections which we have so far not had occasion to mention.

The method described above provides a means whereby relative similarities among species of any group may be determined on an objective basis provided numerous characters differentiating the species are known or can be discovered. The relative similarities can then be used (with or without subjective interpretation) to form a classification which in our case resembled that devised previously by classical systematic methods. This similarity might be used by an advocate of either method to support his favorite technique.

For many groups already studied by competent systematists, our procedure will corroborate and quantify relations already well-known. It is satisfying to the systematist to have his findings substantiated in such a manner and at the same time our method can serve as a check against erroneous conclusions. Taxonomy, more than most other sciences, is affected by subjective opinions of its practitioners. Except for the judgment of his colleagues there is virtually no defense against the poor taxonomist. Above the species level there are no experiments that can be repeated and shown

invalid, no mathematical or symbolic reasoning that can be demonstrated to be in error. It is here that our method has advantages since it will give results disregarding subjective opinions if only characters are chosen properly (see below) and computations performed correctly. Our method may be no better than the work of the best systematist, although it provides much information useful to him; but it is far better than the work of the poor one. (A few years ago C. D. M. placed species 4 in *Anthocopa*. The error was detected and corrected by standard methods; it would have been detected by the statistical method also.)

Serological or chromatographic studies may perform a similar function if the serum proteins are sufficiently representative of the genotype of the animal; more likely they merely represent some additional characters to be evaluated along with the rest.

Presumably our method could be applied to almost any group of organisms, although among lower categories in plants, where reticulate evolution has occurred, the interpretations might differ. Even in groups having extensive fossil records, our method could provide a quantitative measure of amount of difference, hence of evolution, between various types. It could not be applied to groups exhibiting few characters, but we believe that all groups, except possibly some microorganisms, have enough characters if students will undertake to find them, although in some groups the characters will be largely internal or for some other reason more difficult to find than in bees.

Presumably any of the characters could be used, whether morphological, physiological, or behavioral. When using morphological characters we would recommend choosing characters from all parts of the body and would suggest that any duplicate characters (same character for different sexes, growth stages, or on serially homologous organs) be included as single characters. Characters that are consistently and obviously physiologically

coupled should be considered as a single character. Characters should never be organized in such a way that a particular state of one obviously requires a particular state of another. A maximum number of 9 states per character would simplify calculations and should provide an adequate number of classes for most characters.

The number of characters used in such a study as ours should not be less than sixty, as the confidence limits for the correlation coefficients become too large with fewer characters. Also, the sampling of the genetic spectrum might become too incomplete with fewer characters. Since confidence intervals decrease slowly and genetic differences between species in a group are not inexhaustible, additional characters do not proportionally increase the value of such a study. However, little computational labor is added by more characters; if they can be found and recorded, they might as well be used.

As the method is independent of matrix size, it would work with fewer species than we used. We see no reason why it should not be used in a study of relationships among higher categories, such as genera and families. Among very high categories (classes, phyla) difficulty would be encountered, we suspect, in finding enough characters which are (1) known to be homologous among the various categories and (2) not so variable within some of the categories as to make coding of character states an impossibility.

As we have indicated already, we do not consider the method definitive; we hope to improve the method ourselves by factor analytic study and we hope to undertake similar or more refined studies. We do believe that it is a step in the right direction, that is, it decreases the subjective bias in systematic procedure.

As to its economic feasibility, we can point out that our total costs for assistants and IBM machine rental were about $300. Costs would decrease rapidly with fewer species and smaller matrices. Moreover, high speed electronic computers are in-

creasingly widespread. Such machines could handle the entire computation, from raw data to completed correlation matrix, without human intervention in less than an hour, or about one-twohundredth of the time it took us to obtain the same information with punched card machines! Obviously the computational difficulties for such work are rapidly diminishing.

Objections to the Method

While the method here suggested provides an opportunity to quantify the degrees of relationship between various species or groups, the biological reliability of this quantification has yet to be examined. A consideration of certain species pairs gives some idea of the reliability of the information on levels of the forks found in figures 5 to 8. Species 67 and 68 are shown as correlated to a high degree (.96), while species 63 and 64 are less correlated (.91). A study of the actual differences shows that the principal reason for the lower correlation of the latter pair is that one of them has both the femora and the abdomen largely red, the other has these parts black. Thus in two characters, each having four states, these species fall at opposite extremes. Therefore their correlation coefficient was appreciably lowered. Probably the coloration of the femora and of the abdomen are causally correlated in 63 and 64, so that they ought to have been treated as a single character for this species pair, although this is obviously not so elsewhere in *Ashmeadiella* where forms with red femora and black abdomens, and with black femora and red abdomens, are well known. In any event, 63 and 64 are actually very similar and may not be specifically distinct. This shows that at the higher levels in the tree more characters are needed (whenever they can be found) to make the levels of forks more reliable. There were only five characters showing any difference in states between 63 and 64 and seven between 67 and 68. Also it shows that base levels as different as .91 and .96 may not be significant bio-

logically even though they are statistically. These species pairs were selected for comment because the information regarding them in figure 8 seemed to C. D. M. contrary to his impressions as to relationships. Other similar pairs in our diagrams agreed better with his views. At lower levels in the tree we believe that smaller differences in correlation coefficients (levels of forks) are more significant biologically because of the much larger number of differentiating characters involved.

Another commentary on our method, if not an objection, at least suggests that the method is unnecessary. By working first at one classificatory level, then at another, the systematist is able to work with a large number of characters. He does this by ignoring, at each level, all of the characters that are constant within the group he is studying and also those that are variable within the subgroups. Thus if he is studying relationships in a species group, most of the characters are constant and he has only to consider the few that he can find varying among the species. If he is studying generic relations, he ignores characters constant within the subfamily and also, to a considerable degree, those variable within genera, considering only those varying among the genera. Obviously systematists can do good work by their method; we do believe, however, that a method considering the totality of relationships, using all characters, has advantages. In addition to those mentioned earlier, it makes possible consideration of significant trends among specific characters at generic and higher levels. Such trends are commonly ignored by systematists except for the statements that a certain genus *usually* exhibits a particular character.

We have no belief that our method will replace standard systematic procedure, although we believe it may frequently be useful to check such procedure by a statistical approach. The question arises, whether the statistical method could be applied to a systematically unworked or

little worked group. It is obvious that it could not be used until the species are accurately distinguishable, sexes properly associated (unless we are to base the systematics on only one sex), etc. Also, finding numerous characters is a major undertaking, even in groups like the bees which are relatively replete with characters. However, the method should have great advantages for groups containing numerous recognized species whose relationships are unstudied.

A final difficulty with the method concerns species exhibiting intraspecific variation for certain characters that have value in distinguishing species elsewhere in the group. In our study this was a minor problem. Only for a very few species which vary, for example, in the extent of red on the abdomen, was it necessary to record a character state code characteristic of the most widespread subspecies (or the commonest color form) and ignore variability among subspecies (or individuals). However, the problem is certain to be a major one, as already indicated, in any such study in which the objective is to determine relationships of higher categories without consideration of the included species.

SUMMARY

The authors have developed a statistical method for describing relationships among species and higher systematic categories. The method is based on an analysis of a matrix of species-$\times$-species correlation coefficients computed from a large number of characters for each species and has been called the weighted variable group method. Diagrams of relationships resembling phylogenetic trees are constructed and, on the assumption that morphological resemblance is with certain exceptions inversely related to amount of evolutionary divergence, phylogenetic relationships may be hypothesized.

The method is applied to a group of four genera of bees in the family Megachilidae. One hundred and twenty-two characters were studied for 97 species of this group. Relationships between the species in this study as visualized by Michener before the statistical analysis are compared with relationships as found in the analysis by Sokal. Good agreement between the two sets of data is found. The few exceptions to this rule are discussed. Some changes in the classification of these bees are made as a result of the analysis and two new subgenera are named. However, the paper serves primarily to illustrate a method that can be used to remove some of the subjective bias from taxonomy.

LITERATURE CITED

HURD, P. D., AND C. D. MICHENER. 1955. The Megachiline bees of California. Bull. Calif. Insect Survey, 3: 1–247.

MAYR, E., E. G. LINSLEY, AND R. L. USINGER. 1953. Methods and Principles of Systematic Zoology. McGraw-Hill, New York.

MICHENER, C. D. 1941. A generic revision of the American Osmiinae with descriptive notes on Old World genera (Hymenoptera, Megachilidae). Amer. Midland Nat., 26: 147–176.

——. 1949a. Parallelisms in the evolution of the saturniid moths. EVOLUTION, 3: 129–141.

——. 1949b. Records and descriptions of American megachilid bees (Hymenoptera). Jour. Kansas Ent. Soc., 22: 41–59.

——. 1953. Life history studies in insect systematics. Systematic Zoology, 2: 112–118.

OLSON, E. C., AND R. L. MILLER. 1951. A mathematical model applied to a study of the evolution of species. EVOLUTION, 5: 325–338.

SOKAL, R. R. 1957. Quantification of systematic relationships and of phylogenetic trends, Proc. Tenth Internat. Congr. Ent., in press.

SOKAL, R. R., AND C. D. MICHENER. 1957. A statistical method for evaluating systematic relationships, Univ. Kansas Sci. Bull., in press.

STURTEVANT, A. H. 1942. The classification of the genus Drosophila with descriptions of nine new species. Univ. Texas Publ., 4213: 1–51.

TRYON, R. C. 1939. Cluster Analysis. Edwards Bros. Ann Arbor, Mich.