Hunt Institute for Botanical Documentation
5th Floor, Hunt Library
Carnegie Mellon University
4909 Frew Street
Pittsburgh, PA 15213-3890
Telephone: 412-268-2434
Email: huntinst@andrew.cmu.edu
Web site: www.huntbotanical.org

The Hunt Institute is committed to making its collections accessible for research. We are pleased to offer this digitized item.

Usage guidelines
We have provided this low-resolution, digitized version for research purposes. To inquire about publishing any images from this item, please contact the Institute.

Statement on harmful and offensive content
The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences. Due to the wide range of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions. The Hunt Institute for Botanical Documentation does not endorse the views expressed in these materials, which are inconsistent with our dedication to creating an inclusive, accessible and anti-discriminatory research environment. Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections' creators.

Many of the historical collections in the Hunt Institute Archives contain personal correspondence, notes, recollections and opinions, which may contain language, ideas or stereotypes that are offensive or harmful to others. These collections are maintained as records of the individuals involved and do not reflect the views or values of the Hunt Institute for Botanical Documentation or those of Carnegie Mellon University.

About the Institute
The Hunt Institute for Botanical Documentation, a research division of Carnegie Mellon University, specializes in the history of botany and all aspects of plant science and serves the international scientific community through research and documentation. To this end, the Institute acquires and maintains authoritative collections of books, plant images, manuscripts, portraits and data files, and provides publications and other modes of information service. The Institute meets the reference needs of botanists, biologists, historians, conservationists, librarians, bibliographers and the public at large, especially those concerned with any aspect of the North American flora.

Hunt Institute was dedicated in 1961 as the Rachel McMasters Miller Hunt Botanical Library, an international center for bibliographical research and service in the interests of botany and horticulture, as well as a center for the study of all aspects of the history of the plant sciences. By 1971 the Library's activities had so diversified that the name was changed to Hunt Institute for Botanical Documentation. Growth in collections and research projects led to the establishment of four programmatic departments: Archives, Art, Bibliography and the Library.

Taximetrics Lab., Armory 101 December 29, 1969

Dear Frank:

I finally found time to read your ms, The Selection and Weighting of Characters in Angiosperm Taxonomy. In general, I agree with the paper, as to its inclusiveness and its efforts to explain what we mean by character analysis.

Two points seem to need some discussion, however. The first is the word "weighting." Since this word has so many connotations, why not substitute the word "evaluation?" This seems to me to reflect more accurately what a taxonomist actually does, and relieves us of the burden of defense against those who have other connotations. I believe Mr. White would agree with such a substitution. You might discuss in the text the relations of evaluation to weighting, but don't put the word weighting in the title.

The second point is, perhaps, a matter of my interpretation, but I was startdd by your conclusion (p. 16, 2nd para.)".....it is possible that such a taxonomist may proceed with the remaining stages of his work without further help from taximetrics." In the very large genera, where we believe that clustering methods such as ours are significant, we (or I) think that you have only begun when you have your characters straightened out. I really don't see how you can defend your statement.

Other than these two comments, I have no other disagreements. Even these two aren't very serious, but probably merit some discussion with Mr. White.

I have another mather which I wonder if you can help me with. I have kept the Manihot specimens from Kew much too long, and have just had a rather pointed letter from Sir George Taylor to get them back. We need about 6 months more to finish study of them, and I wonder if, through your connections with Mr. Polhill, you could aid in explaining that we are actively engaged in study of the specimens, they are in good condition and care, and that our computer development has slowed down the work beyond that which would normally been required, but that now we are in the final stages of the study, and hope to complete it within the six month time mentioned above. I intend to write to Sir George myself, but I am sure that some additional knowledge, or explanation,of the work from you to Polhill (and hopefully from him on to the director) would aid immeasurably. Perhaps you want to talk this over with Mr. White before going ahead with something like this.

Best wishes for the new year, and my regards to Mr. White.

Sincerely,

TELEPHONES OXFORD
55757 PROFESSOR
57891 DEPARTMENT

DEPARTMENT OF FORESTRY
COMMONWEALTH FORESTRY INSTITUTE
UNIVERSITY OF OXFORD
OXFORD OX1 3RB

10 December 1969

Professor David J. Rogers,
Taximetrics Laboratory,
Dept. of Biology,
Armory 101,
University of Colorado,
BOULDER, Colorado 80302

AIR MAIL

Dear Dave

Hello! How are things in Boulder?

At last I have got the Character Analysis paper completed and I shall try to submit it as soon as possible. I enclose a copy for you. If you have any major suggestions or disagreements with what I have said or done, could you possibly let me know very quickly indeed, as I am trying to hurry this through as quickly as possible. Mr. White has put pressure on me to try and eliminate as many as possible of the technical terms of taximetrics and has insisted on my using words that are meaningful to ordinary botanists, such as "weighting", etc.

Do let me know what is happening in Boulder now. I am trying very hard to get a foothold on the Atlas Computer at Chiltern, so as to get your programmes established there. One of the reasons for this is that the Williams & Lance type programmes are getting a big foothold over here.

Best wishes for Christmas and the New Year to you and all the Lab.

Yours sincerely,

Frank

F.A. Bisby

(missing TABLE 2)

THE SELECTION AND WEIGHTING OF CHARACTERS IN ANGIOSPERM TAXONOMY

BY F. A. BISBY

Department of Forest Science, University of Oxford

SUMMARY

The selection and weighting of characters is of great importance in Angiosperm taxonomy. Fifty-two morphological characters, observed in two hundred and seventy-three African species of *Crotalaria* L., are analysed using a taximetric procedure, "Character Analysis", which provides a quantitative measure of their potential taxonomic value, the "information contribution". The characters are listed in descending order of information contribution and compared with the relative importance given to them by Polhill in his previously published orthodox study of the genus. The four characters used by Polhill for his major division are among those highest on the list, and only two of the twenty that he used for the delimitation of sections occur in the lower half of the list.

The taxonomic value of a character is dependent on how it is defined. The way in which the information contribution measure can be used in the production of better definitions of poor characters is described, and one very successful example of "Character Improvement" is given.

The potential value of these taximetric procedures in taxonomic revisions of large genera and families with complex reticulate patterns of variation is discussed.

INTRODUCTION

The availability of computers with large memory stores and high rates of computation has in recent years opened up the possibility of performing very rapidly, large quantities of data comparisons. This is of potential importance in classification where it is receiving increasing attention. In biological classification it has lead to the study and development of automated procedures intended to replace some or all of the steps in orthodox taxonomy. This subject, originally described as numerical taxonomy, is now frequently referred to as taximetrics. The latter word describes the use of measurement and arrangement, and avoids the implication of the earlier phrase, that this subject involves a different kind of taxonomy.

Despite some common goals, such as the precise and objective treatment of data, avoidance of a priori character weighting, and the replacement of time-consuming herbarium work by automated procedures, workers in taximetrics have produced methods which vary both in procedure and in their underlying concepts concerning classification.

Taximetrics first reached the general public when Sokal and Sneath published The Principles of Numerical Taxonomy in 1963. They envisaged the complete replacement of orthodox taxonomy, particularly in the complex field of Angiosperm taxonomy, by a revolutionary automated process, numerical taxonomy, designed on a phenetic model of classification. The book is as much concerned with pheneticism as with taximetrics and, being the only book available on the subject, has left many biologists with the mistaken notion that pheneticism and a process completely automated from beginning to end are the essential characteristics of taximetrics. Two particularly important features of this approach are the use of large numbers of characters and the implicit use of the same characters for all levels of classification.

The first of these features arises from the assumption that the characters to be used in a study are a sample from the set of possible characters and that these provide an approximation to the genotype of the organisms under study.

Consequently it is proposed that by increasing the number of characters used, the resulting taxonomic arrangements should approach asymptotically a stable configuration. In practice this approach can lead to the uncritical use of very large numbers of characters. However, in Angiosperm taxonomy the number of characters available, especially above the species level, is strictly limited and collecting data for the number of characters sometimes advocated by pheneticists, often in the order of one hundred, is simply not feasible. Mayr (1965) makes it clear that many of the characters that at first sight appear to be potentially available in higher organisms are in fact not available because of incomplete material or because they are found to be highly variable within taxa of lower rank. Another difficulty is that where more characters can be found, this is usually achieved by observing characters of a new type and their introduction en bloc introduces bias towards the type of characters introduced. In a particular case it may be that the outcome of a classification is determined by the ratio of, say, pollen morphology characters to wood anatomy characters, and yet any method of introducing characters "at random" must involve some arbitrary decision as to the appropriate ratios of different character types.

A second, implied, feature in the phenetic approach is that the same characters are used at all or several levels of classification. Thus Sokal and Sneath suggest that higher taxa should be represented by data for the full number of characters used at lower levels scored for a chosen "exemplar" individual from the higher taxon in question. They write "Thus the error introduced by choosing a single representative of a taxon should not be large enough to seriously affect the estimation of the similarity among the taxa of the study". This is in sharp contrast to the method used by orthodox taxonomists, in which, at any level, only variation between and not within the taxa of lower rank is used in their classification. For instance, in the present study of Crotalaria, species are taken as the lowest taxonomic unit and only characters that are not variable within species are used in the study.

Other workers in taximetrics have asked whether taximetric methods can be used in a stage by stage process of classifying rather than in trying to replace the whole process with a single automatic one. Rogers and his collaborators at Boulder and elsewhere are shewing that taximetrics is of value in this stage by stage process in which the taxonomist retains control of each stage. They write (Prance, Rogers and White, 1969) "that taximetrics should serve as a tool to the taxonomist and that he alone is able to evaluate its results." The methods that they have devised try not only to emulate the activities of a good orthodox taxonomist, but also to separate into distinct stages the various logical steps in orthodox work (Rogers, Fleming, Estabrook, 1967). What is of particular importance to Angiosperm taxonomists is that they include in these logical steps the selection of characters.

The importance of character selection in taxonomy was discussed critically by Cain (1959) who emphasized the distinction between a priori weighting and a posteriori methods such as phyletic weighting or selection of characters with high covariation (see also Cain and Harrison, 1959). In chapter 14 of The Origin of Species Darwin writes "The importance, for classification, of trifling characters, mainly depends on their being correlated with many other characters of more or less importance". The characters that are most useful to taxonomists are those that are highly correlated with others and they can sensibly be weighted or selected after they have been found to be highly correlated. Davis and Heywood (1963) write of the subconscious assessment of correlations "The ability to do this is probably the most characteristic and useful attribute of the good taxonomist". On page 139, discussing the numerical approach, they write "Character selection is the weak link in this whole approach".

In weighting or selecting characters according to their correlations with others, it is important to be cautious of correlation caused by logical and functional dependence. The logical dependence of the states or values of characters causes correlation which can be attributed to the way in which the characters have been devised. Cain and Harrison (1960)

describe situations in which seemingly different characters are functionally dependent because of the precise functional relationships between parts. Such characters are necessarily totally correlated and can only be treated as one character. In the Angiosperms floral mechanisms are often highly adapted for pollination by particular insects or birds, as in the subfamily Papilionoideae of the Leguminosae. However, a glance at the floral characters used in this group shows that, although correlations and functional relationships do exist, the characters themselves occur in a bewildering number of combinations. Such characters are only partially correlated through functional relationships, and, without using them, the whole basis for a natural classification would break down and major Angiosperm families, long regarded as natural and monophylletic, would disappear.

Mayr (1965) writes "A character that is not part of a functional complex, and yet is correlated in its presence or absence in several taxa obviously has a much higher predictive value and should, therefore, get a much higher weight in a classification than a character that is distributed more randomly among related taxa". and "It should be an easy matter to use this method for an a posteriori weighting of characters.....".

In simple taxonomic situations good taxonomists are likely to detect those sets of characters that are highly correlated unaided by taximetrics. However, there remain many taxonomic problems of great complexity, such as the supra-specific classification of Senecio with over 2,000 species and Solanum with over 1,500 species. In these genera, even though the species may, for the most part, be easily separable on the basis of many characters, the problems of evaluating possible characters and of recognising natural groupings involve a formidable number of comparisons. Monographic revision of this kind is an urgent necessity and, despite the reluctance of orthodox taxonomists to use taximetric methods, it may be in this field that their greatest contribution will be made to the efficient use of taxonomic manpower.

The genus Crotalaria L. in the Leguminosae, subfamily Papilionoideae, approaches the type of genus described above. Although it contains only about 600 species, the variation between species in the genus is highly reticulate and presents a pattern of great complexity. The genus is pantropical with the greatest aggregation of species in Africa. Supra-specific classification of the genus involves a very large number of comparisons, both between potential characters and between species, and is a good subject for taximetric studies.

For many years Milne-Redhead and then Polhill worked on the taxonomy of the Crotalaria species found in Africa. Their work culminated in the delimitation of 432 African species (Milne-Redhead 1961, Polhill 1968) and a classification of these into eleven sections based on flower morphology. Their work is based on a very extensive knowledge of the genus, both in the field and herbarium, and is highly regarded by other Angiosperm taxonomists.

In 1967 I started a study of the African Crotalaria species based on the same material as that used by Milne-Redhead and Polhill, principally the large collections at Kew. The purpose was to classify the African species using taximetric methods and using the same material as was available to Polhill. I collected the data for the taximetric study quite independently but used a large number of characters suggested by Polhill. His suggestions lead to the use of 39 characters which varied between species, but rarely or never within species. These included those that he had found valuable and those that he found less valuable for his classification. In addition 13 other characters were used. The species delimited by Milne-Redhead and Polhill were accepted as the basic units of the classification. The number of species used (273) was limited by the availability of sufficiently complete material for the description of this large number of characters.

This parallel study differed from Polhill's only in the methods used and not in the information and material available. Taximetric methods were substituted for the orthodox approach so that differences between the results achieved must be interpreted as consequences of the methods used and not of the material or the author's taxonomic background.

concordance it can be assumed that both are providing the best possible results. This may not be true, but it can only be disproved by investigations not yet made. Where the two studies are not in concordance, clearly, at least one of the studies has been unsatisfactory and both studies must be carefully scrutinised. Either the taximetric methods have failed to take into account some relationship or assessment correctly used by Polhill, or they have revealed some situation that he has overlooked.

The information Theory Model used here was originally described by G.F. Estabrook (1967). Hawkesworth, Estabrook and Rogers (1968) have discussed the potential uses of the model for character analysis and indicate its usefulness in the relatively small genus Arceuthobium. In the present paper the model is used for character analysis in a problem of much greater magnitude, the usefulness of the results is demonstrated by comparison with a parallel orthodox study, and a new technique for character improvement is described and demonstrated.

THE INFORMATION THEORY MODEL

In the present paper a character is a formal entity distinct from raw observational descriptions. The character is a basis for comparison. For each character there exist exclusive character states which partition the individuals being compared, so that like individuals are in the same state, unlike objects in different states. Given a particular set of observations it is often possible to formalise different characters so as to include different information. For example, from observations on the leaf forms of members of the Leguminosae it is possible to formalise one character with the following states: simple, trifoliolately palmate, trifoliolately pinnate, impari-pinnate, pari-pinnate, but a second character could be formalised with states: leaflets 1, leaflets 3, leaflets more than 3. By formalising the states of a character the objects are partitioned (divided up into groups) and, as in the example, different formalisation of characters from the same observations can give different partitions.

For each character there is a quantity called its information content which can be calculated as follows.

$-H = a \text{ Ln } a + b \text{ Ln } b + c \text{ Ln } c \ldots\ldots\ldots n \text{ L n}$

where a is the fraction of individuals in the a th state, b is the fraction in the b th state, up to n states in all. Ln = logarithm to the base 2.

As with intuitive ideas of information imparted it should be noted that this quantity rises with the number of states (for only one state $-H = 1 \text{ Ln } 1 = 0$ and there is no information) and is at a maximum for a given number of states when their occurrence is equi-probable ($a = b = c \ldots\ldots = n$).

If the characters divide the individuals into the same partition, so that all of the objects in the same state for one character are also in the same state of the second character, then not only do they have the same quantity of information content, but also all of this information is identical in the two partitions, and they are said to have all of it held in common. Given to which state of one character an individual belongs, there is 100% probability of assigning it correctly to a state of the second character. More frequently two characters will have some part of their information in common. The actual quantity (called the "mutual information ", by Orloci) (1968) is calculated from the conditional probabilities for the states of one character, given to which state of a second character the individuals belong.

Details of the computations are published in an Appendix to a paper by Estabrook (1967).

The Analogy

In the rare case of two characters having the same information content and all of their information held in common, the whole set of individuals is divided identically by the two characters and from a taxonomic standpoint the situation is ideal. An Angiosperm taxonomist encountering such complete correlation would undoubtedly give the partition very great importance in his deliberations. Although the information in question is repeated identically in the two characters, it is of great usefulness by virtue of this repetition.

A more likely case is that of two characters with different information contents, but with all of the information of the one (with a smaller content) shared with the other (with a larger content). This is the case of partitions that fit together hierarchically so that the states of one (with less information content) wholly contain the members of states of the other. Here again a taxonomist will be most interested in the information that is common to both characters. For instance where a particular character divides the individuals of a study into four groups, and another divides them into two, so that each of the two groups contains the members of two of the groups from the other partition, the taxonomist will consider that he has weighty evidence for dividing them into two. In the model this important partition is represented by the information of the second character, which is the information held in common between them.

The usual case with two characters is that they have in common an amount of information which is less than the complete information of either. As in the previous cases it is clearly their shared information that is of interest. This represents the groupings that are in agreement between the partitions by both characters. The differences in the two partitions are represented by the information in either that is not shared.

It emerges from this examination of the analogy between the model and taxonomic situations that the taxonomist is in fact interested in obtaining a partition (or structured set of partitions) which is most closely related to the body of information held in common by the characters in a study. Consequently, to approach this body of information most closely the taxonomist will wish to examine the formal characters and adjust the complete set so that he uses only characters with as high as possible a fraction of their information held in common with other characters. It is with this end in view that character analysis and character improvement are described.

CHARACTER ANALYSIS

The characters are taken two at a time and for each, its information content and its table of conditional probabilities with the other are calculated. These conditional probabilities are the probabilities of an individual being in a certain state, given to which state of the other character that individual belongs. Table 3 shows two examples. From the two tables of conditional probabilities, the quantity of information held in common by the two characters is calculated. This quantity is then expressed in two ways, as a fraction of the information content of the one character, and as a fraction of the information content of the other.

Where the number of individuals and characters is large, the conditional probabilities may be of great interest themselves, although with smaller studies these relationships will present no problems. It is important in the formalisation of characters, especially if they are to be used in a taximetric study, to realise how small an information content is possessed by, say, a two-state character with one of the states being very rare, and how large a content is possessed by a many state character with approximately equiprobable states.

For each individual the fractional information held in common with each of the other characters is summed. This total will be referred to as the "information contribution" of a character. In the earlier discussion of the Information Theory Model, it was shewn that the information of greatest interest for taxonomy was represented by the information held in common among the characters. To approach this information, characters with large fractions of the information held in common with many other characters will be more useful than characters with lower fractions shared. Thus although high values of fractional information in common with one particular character may be of interest *per se*, the major assessment of the usefulness of a character will depend on the total for that character of fractional information quantities held in common with each of the other characters, the "information contribution".

The character analysis is now complete and the taxonomist has in front of him a measure for each character of its usefulness in revealing the structure in which he is interested. In most instances these values will reveal a disparity among the characters and will motivate the taxonomist to try to improve the character set so as to eliminate those characters of low usefulness. This does not necessarily mean that any of them will be abandoned, but rather that the information will be formalised in a more useful way. Again the formalisation of leaf-form characters in the Leguminosae can be used as a simple example. If there is correlation between several characters in a study and the presence or absence of a terminal leaflet, then there will be a very low information contribution from a character formalised with these states; leaflets 3 or less, leaflets 4 or more. However, by changing the formalisation but using the same observational data (possibly leaflet counts or arrangement diagrams) a two-state character such as "pari-pinnate or impari-pinnate" can be found which, on repeating some of the computations, is found to have a very much higher information contribution. This process of changing characters to obtain the largest possible amount of useful information from the raw observational data is described below.

CHARACTER IMPROVEMENT

Improving a character may involve testing several alternative formalisations before the most satisfactory one is found. The taxonomist will draw on impressions of conditional probabilities from the material and on conditional probabilities calculated in the initial character analysis for suggestions on how to improve the formalisation of a character. He may also be stimulated to collect more, or more detailed, observational data from the plant material.

Each experimental formalisation is treated as a potentially useful character and is compared with the other characters in the study by computing the same quantities as in the initial character analysis. By summing the fractional information quantities in common with the other characters (omitting the character that the experimental one replaces) the information contribution of the experimental character is obtained.

Before comparisons can be made the information contributions of the other characters must be corrected by subtracting their fractional information held in common with the replaced character and adding their fractional information held in common with the experimental character. This correction is necessary because by replacing one character with an experimental one, the fractional information quantities of other characters shared with it will be changed, as well as its fractional information shared with the others. Thus if the fractional information quantities held in common are tabulated in columns for each character, replacing a character involves the substitution of a complete row as well as a complete column of values, and consequently the information contributions (column totals) will all be altered.

After this experimental cycle of computations, the taxonomist has the same computed quantities for the experimental character as for all the other characters. He can compare the performance of the experimental formalisation with that of the character that it is hoped to replace and if it is judged to be markedly more useful than its predecessor, data using the new formalisation is used to replace that of the previous formalisation. Otherwise the search for a more useful formalisation is continued and the experimental cycle of computation is repeated. Where a poor character cannot be successfully improved by this method the taxonomist must consider whether or not there is actually any useful information to be gained from the observational data in question and if not, he may choose to eliminate it from the study.

CHARACTER ANALYSIS IN *Crotalaria*

I had already collected data from 273 African species of *Crotalaria*, using 52 formalised characters, before the present study was begun. All characters were of the qualitative type for which the information theory model is appropriate. Of the 52, 39 were characters suggested by Polhill and used by him (Polhill 1968).

The set of 52 characters was analysed using the computer programme CHARANAL, which is described briefly in the APPENDIX on page .

For each character the fractional information held in common with each of the remaining 51 characters was tabulated and summed to obtain its information contribution (see Table 2). The information contributions of these characters varied from 0.75 to 6.14.

The quantities of fractional information in common were low and even where an impression of good correlation was obtained from the material, values were usually less than a half. This underlines the main problem of supra-specific classification in Crotalaria, which is that variation among the large number of species is highly reticulate despite the discrete nature of the species themselves. Consequently the major divisions within the genus have to be made on combinations of characters and the problem of recognition and circumscription of groups is very real. The highest value for fractional information in common is the 0.78 of the information in character 3 (style simple or bifurcate) shared with character 35 (legume inflated or laterally compressed). Both of the characters have a very rare second state and these are both found in the species C. leptocarpa Balf. f. the only member of Polhill's section Schizostigma Polhill included in the study.

The clearest demonstration of the potential value of character analysis comes not from listing the multitude of individually useful facts that emerge, but from examining the correspondence between those characters with a high information contribution and those actually selected by Polhill for use in his circumscriptions and keys at the sectional level.

In Table 2, the 52 characters analysed are listed in descending order of information contribution. Those characters used by Polhill are marked "P" if used for his major division of the genus and "p" if used by him to separate sections. Characters marked "B" were introduced by myself and those marked "Ba" were the major characters used by Baker in his classification of the genus in 1914 (Baker 1914).

From Table 2, it is clear that Polhill's major characters all have relatively high information contributions. It is also interesting that the major characters of E.G. Baker have, with one exception, very much lower information contributions than those of Polhill. Polhill rejects Baker's system as unsatisfactory and criticises the use of these particular characters for the major divisions (Polhill 1968).

They were among the characters studied by Polhill but rejected in favour of others, so confirming the general concordance between high information contribution and usefulness as judged by him. He did not consider character 57 (filament length) which is in fact well correlated with character 4 and, like that character has one of the states confined to his section Dispermae Wight and Arn..

CHARACTER IMPROVEMENT IN Crotalaria

Those characters with low information contribution values were carefully examined and experimental formalisations tested, again using CHARANAL, to obtain values which can be compared with those of the original characters. The improvement achieved for character number 7 is described below.

Character 7 is the formalisation of observations of the indumentum on the upper margin of the keel. Three states were originally defined — margin glabrous; margin with an indumentum of hairs less than 1 mm long (usually almost ciliate and clearly distinct from the next state); and, margin with hairs greater than 1 mm long (these are woolly white hairs visible to the naked eye). In the initial character analysis its information contribution was 1.45, as in Table 1.

Close examination of these observations revealed that the second state contained not only species with short hairs along some length of the upper margin, but also some species with a very few short hairs (up to 10) arranged distinctively at the proximal end of the upper margin. Two experimental formalisations were devised and then analysed in comparison with each of the remaining 51 characters. In one of these formalisations (7b) the species with few hairs were placed in the first state, which is then defined as "margin glabrous or with a few short hairs at the proximal end". This was tried because of an impression, gained from the material, that these species had states of other characters in common with a group of species, all of which had state one of character 7. This formalisation gives an information contribution of 1.96. The other experimental formalisation (7c), in

which the species with few hairs were placed in a separate fourth state, gives an information contribution of 2.30, a considerable improvement on the original 1.45.

It was decided to replace the original character 7 with the new formalisation 7c. The information contributions of the other 51 characters were then adjusted for the replacement of character 7 by 7c.

The way in which this change increases the information held in common with another character can be seen from the conditional probability distributions of character 40 (position of the inflorescence) given to which state of character 7 or 7c a species belongs. For character 7 only the third state is decisively correlated with character 40. (See Table 3). However the separation into four states of character 7c gives much more structure with, in addition to the third state, states two and three being strongly correlated with character 40.

Unfortunately, when this work was being carried out at Boulder, the herbarium material used in the initial states was not available so that character improvement was of necessity limited to those characters for which I had already recorded extra observations. Alternative formalisations of characters 50 and 51 were also tested, but their information contributions remained virtually unchanged. For character 50 it rose from 1.40 to 1.51, and for character 51 it dropped from 1.25 to 1.14.

CONCLUSIONS

In Crotalaria the correspondence between information contribution and prominence in Polhill's classification is evidence that information contribution can be used as an assessment of the potential usefulness of a character. The information contribution is a quantity calculated solely from properties of the data, and is not dependent on circular argument such as consistency with groupings of individuals. The comparison with Polhill's classification provides an external test with which was completed (Polhill 1968) just before my own study began.

The measure may be used for *a posteriori* weighting as described in Davis and Heywood (1963).

A taxonomist who turns to taximetrics because of the problems presented by large amounts of data, and uses character analysis to discover the most useful characters, can save a large amount of time. As this is one of the most difficult and time-consuming stages in taxonomic work, it is possible that such a taxonomist may proceed with the remaining stages of his work without further help from taximetrics.

In the description of the information theory model, observations and characters were carefully distinguished. In higher organisms, where structures are frequently complex and variable this distinction is real, and the process of formalising characters necessarily precedes any activity that uses the characters. The procedure for improvement is an iterative one in which the consequences of one formalisation are examined and used in trying to devise an improved formalisation. This method aims at extracting as many highly useful characters as possible from the observations. In practice some characters, especially those with very low information contribution and low information content, will not respond to improvement and may have to be abandoned.

ACKNOWLEDGEMENTS

I wish to thank Professor D. J. Rogers and Mr. G. F. Estabrook for many of the ideas incorporated in this paper, Mr. F. White for much constructive criticism of the manuscript and Mr. R.M. Polhill for allowing me to draw on his information and personal knowledge of the genus *Crotalaria*. This work was supported by the Taximetrics Laboratory, University of Colorado and was carried out during the tenure of a Christopher Welch Scholarship at Oxford University.

REFERENCES

Cain, A.J. (1959). Deductive and Inductive Methods in Post-Linnaean Taxonomy. Proc. Linn. Soc. Lond. 170, 185.

Cain, A.J. and G.A. Harrison (1960). Phyletic Weighting. Proc. Zool. Soc. Lond. 135, 1

Darwin, C. (1859). The Origin of Species, London.

Davis, P.H. and V.H. Heywood (1963). Principles of Angiosperm Taxonomy, Edinburgh and London.

Estabrook, G. F. (1967). An information Theory Model for Characters Analysis. Taxon 16, 86.

Hawkesworth, F. G., F. G. Estabrook and D. J. Rogers (1968). Application of an Information Theory Model for Character Analysis in the genus Arceuthobium (Viscaceae). Taxon 17, 605.

Mayr, E. (1965). Numerical Phenetics and Taxonomic Theory. Syst. Zool. 14, 73.

Milne-Redhead, E. (1961). Miscellaneous notes on the African species of Crotalaria L.: I. Kew. Bull. 15, 157.

Orloci, L. (1969). Information Theory Models for Hierarchic and Non-hierarchic Classifications. In: Numerical Taxonomy (Ed. by A. J. Cole), London and New York.

Polhill, R. M. (1968). Miscellaneous Notes on the African species of Crotalaria L.: II. Kew. Bull. 22, 169.

Prance, G. T., D. J. Rogers and F. White (1969), A taximetric study of an Angiosperm family: Generic Delimitation in the Chrysobalanaceae, New Phytol. 68, 1203.

Rogers, D. J., H. S. Fleming and G. F. Estabrook (1967), Use of computers in studies of taxonomy and evolution. In: Evolutionary Biology (Ed. by Th. Dobzhansky, M. K. Hecht. and W. C. Steeve), p. 169, New York.

Sokal, R. R. and P.H.A. Sneath (1963), Principles of Numerical Taxonomy, San Francisco and London.

APPENDIX : CHARANAL, A COMPUTER PROGRAMME FOR CHARACTER ANALYSIS

This programme performs comparisons between pairs of characters. For each pair the following quantities are calculated and printed out :- Conditional probabilities; information content of each; information quantity held in common and information quantity held in common expressed as a fraction of the information content of each. The individuals missing data in one or both of the characters are listed and eliminated from the computation.

The user specifies which of the characters he wishes to compare. In an initial character analysis this will usually be all of the possible $n(\frac{n-1}{2})$ comparisons amongst n characters. In subsequent runs one or a few experimentally formalised characters will be compared with the remainder.

The programme CHARANAL is written in FORTRAN for the C.D.C. 6400 and requires only a minimal amount of core storage. Details and a listing are available from :- Professor D. J. Rogers, Taximetrics Laboratory, 101 Armory, University of Colorado, BOULDER, Colorado 80302, U.S.A.

	Character No.	Information Contribution	Character Description	Number of States	
1	4	6.14	Anther shape	2	p
2	3	5.27	Stigma shape	2	p
3	57	4.84	Filament length	2	Bi.
4	40	4.65	Raceme attachment	3	p
5	25	4.55	Mature calyx deflection	2	
6	15	4.32	Standard appendage attachment	3	P
7	2	4.19	Style shape	4	p
8	10	4.08	Beak twisting	3	P
9	8	4.04	Presence of crest on keel	2	P
10	5	4.00	Keel shape	2	p
11	29	3.82	Markings on petals	2	p
12	38	3.41	Bract persistence	3	
13	48	3.41	Whether a tree	2	
14	1	3.37	Style pubescence	4	p
15	24	3.33	Calyx lobing	3	P
16	35	3.20	Legume inflation	2	
17	19	2.75	Seed thickness	3	
18	17	2.55	Seed number	3	p
19	22	2.45	Hypanthium development	2	p
20	34	2.42	Legume shape	4	p
21	43	2.38	Presence of spines	2	Ba
22	6	2.32	Keel pubescence	3	p
23	21	2.31	Aril development	2	p
24	31	2.27	Stipe attachment	2	p
25	36	2.24	Bracteole position & form	5	p
26	28	2.24	Calyx lobe number	2	
27	37	2.05	Bud inflection	3	p
28	26	1.95	Calyx lobe shape	4	
29	32	1.95	Stipe/calyx tube length	4	
30	27	1.77	Calyx lobe/tube length	2	

31	14	1.69	Adaxial pubescence of standard	3	
32	46	1.65)	Development of petioles	2	
33	47	1.65)	Development of stipules	3	
34	23	1.55	Calyx/keel length	3	
35	13	1.51	Shape of standard	5	
36	7	1.48	Keel upper margin pubescence	3	
37	9	1.43	Keel length	3	Ba
38	52	1.40	Prominence of lateral ridges on keel	2	Bi
39	30	1.39	Flower colour	2	p
40	44	1.33)	Leaf-form	3	Ba
41	33	1.33)	External pubescence of legume	3	
42	51	1.27	Reflexing of calyx lobes	2	Bi
43	12	1.26	Wing/keel length	2	
44	20	1.16	Seed surface	2	
45	45	1.15	Development of leaf fascicles	2	
46	41	1.14	Raceme form	3	
47	18	1.11	Seed shape	3	
48	42	1.07	Flower number per inflorescence	2	
49	50	1.04	Seed colour	5	Bi
50	16	0.93	Internal pubescence of legume	2	Ba
51	11	0.80	Wing shape	2	
52	39	0.75	Bract/pedicel length	2	

TABLE 2

TABLE 3

Given that a species for character 7 belongs to:-	Probability of the occurrence of states of character 40		
	state 1	state 2	state 3
state1	.68	.06	.25
state 2	.80	.05	.13
state 3	1.00	0.00	0.00

Given that a species for character 7c belongs to:-	Probability of the occurrence of states of character 40		
	state 1	state 2	state 3
state 1	.68	.06	.25
state 2	.86	.06	.07
state 3	1.00	0.00	0.00
state 4	.14	0.00	.85

FAH/C

Forest Herbarium,
Dept. of Forestry
Commonwealth Forestry Inst.,
University of Oxford.

18 February 1970

Professor D. Rogers,
Taximetrics Laboratory,
Dept. of Biology,
University of Colorado,
Boulder, Colorado 80302.

Dear Dave,

I wonder if you could possibly let me know what programming facilities you now have for the Taximetrics Laboratory, as we may need some help in converting the CHARANAL programme for the Atlas computer at Chilton. For instance, do you still have either George or Bob with you, who would be able to communicate with us on some of the details of the programme? I hope you realise that this is in your own interest and that I am trying as hard as possible to make a name for some of your programmes over here.

I was over at Cambridge the other day, talking to Jardine and he seems very friendly towards our approach and encourages me in my plans for future research. I should very much like to go to Cambridge and, of course, they have an Atlas computer themselves, so, once established, we could use the programmes at both places.

Do let me know any news about people from the Laboratory. I have not heard anything for some time.

Yours sincerely,

Jardine

P.S. I did some talking on your behalf at Kew. Don't worry about them too (L) remember they're civil servants, mostly rabid bureaucrats (Said you were working on specimens and would return them in 6 months.)

Forestry Dept
Oxford
OX1 4RB

3rd February 70

Dear Dave,

Thanks for your letter and comments. It's really sad to hear that financially things are about as bad as th could be for the Taxinometrics Lab. What a George and Bob and Gill doing now? Also I ha no reply to a Christmas letter to the Wongs. I could be that they're just poor correspondents, but if they have moved from 10th street Boulder could you let me know their address?

I accept both of your criticisms of my paper and have changed the text accordingly. It's been accepted by the New Phytologist, so it should appear about June-July.

Did I ever tell you how successful that seminar series at Cambridge turned out? We really got down to some of the central problems and at one stage I described Charanal which went over very well, especially with Jardine and Co.

Will you be a referee for some more job applications please Dave? I've already used your name in an application to Merton College and I'd like to use it in some more.

Have you had any programme enquiries from Dr Howlett, Director of the ATLAS

Professor D.T. Rogers
Taximetrics Lab
Biology Department
Univ of Colorado
Boulder, Colo, 80302
USA

BY AIR MAIL
AIR LETTER
PAR AVION AEROGRAMME

OXFORD 7.15PM 3 FEB 1970

SECOND FOLD HERE

AN AIR LETTER SHOULD NOT CONTAIN ANY ENCLOSURE; IF IT DOES IT WILL BE SURCHARGED OR SENT BY ORDINARY MAIL

SENDER'S NAME AND ADDRESS

F.A. Bisby
St. John's College
Oxford
OX1 3JP

TO OPEN SLIT HERE

computing lab at Chilton, Didcot (near Oxford)? I've got him interested in getting a series of taximetric programmes going and of course I've plugged your name. He's an influencial man on our Science Res. Council so it's in our interest to get in there.

We seem to have had weeks of rain! – a really warm wet winter, and I've been missing Boulder life – the winter trips in the hills and those lovely people I lived with in Whitney Place! Yours [illegible]

Mr.F.White, Forest Herbarium, Commonwealth
Forestry Inst. University of Oxford OX1 3RB

FW/C

15 July 1970

Professor David J.Rogers,
Taximetrics Laboratory, Dept. of Biology,
Armory 101, University of Colorado, BOULDER, Colorado. 80302

Dear Dave,

First of all I would like to say how very sorry I was to learn, some time ago, that you had not been able to raise the funds necessary to keep your taximetrics team in existence. It seems to me that in times when money is short that it is the most worthy projects which are axed first. This must have been a most bitter blow to you after your remarkably successful pioneering efforts in such an important field.

I should also, in this letter, like to try to clarify the position concerning the request Frank Bisby recently made to you.

My research student, Francis Ng, who is studying Malesian Ebenaceae, at one stage in his work was finding, and indeed is still finding some difficulty in selecting and evaluating the best characters to classify his species. After some considerable discussion with Bisby and myself we thought that he had a problem to which your programme CHARANAL might be relevant. I understood from Bisby that, although he was arranging for this programme to be made available in this country on the SRC Atlas computer at Harwell, there was likely to be some appreciable delay. Because of this we discussed the possibility of sending the punched cards to you. However, it was never my intention that Bisby should do this on his own initiative. I had intended to write myself officially, as head of the herbarium, and was very surprised when I learned that Frank had done this on his own initiative, although I must admit that I possibly did not make my intentions sufficiently clear.

It so happens, however, that a few days before Frank received your last letter he received a telephone call from the programmer at Harwell, saying that your programme had been re-written for the Atlas machine and that she wanted some test data. Just at present Frank is fully occupied completing his thesis but this will be finished at the end of this month. I have impressed upon him that before he leaves Oxford at the end of August he has a great obligation to you to make quite sure that CHARANAL is running properly at Harwell and that this must be his highest priority. I am sure that this is a commitment which he will fulfil. This being so I think it would be most suitable for Francis Ng's data to be run at Harwell, particularly as Frank has shown a good deal of interest in this work.

I am, however, quite sure that the problem itself will be of interest to you and when we have got some results I shall write to you again. As you know, my own interest in taximetrics is very much that of a monographer and I think in this particular case we have a very interesting situation where a capable but relatively inexperienced taxonomist is tackling a substantial problem de novo.

/over

With best wishes,

Yours sincerely,

Frank

F. White

TO OPEN SLIT HERE

SENDER'S NAME AND ADDRESS

Mr F. White, Forest Herbarium
Dept. of Forestry
Commonwealth Forestry Inst.
University of Oxford OX1 3RB

AN AIR LETTER SHOULD
NOT CONTAIN ANY ENCLOSURE;
IF IT DOES IT WILL BE SURCHARGED
OR SENT BY ORDINARY MAIL

SECOND FOLD HERE

BY AIR MAIL
AIR LETTER
PAR AVION AEROGRAMME

OXFORD
9.45AM
16 JLY
1970
A

9p

Professor David J. Rogers,
Professor of Biology,
Taximetrics Laboratory,
Dept. of Biology,
Armory 101,
University of Colorado,
BOULDER, COLORADO 80302

U.S.A.

Air Mail

DEPARTMENT OF FORESTRY
Commonwealth Forestry Institute
University of Oxford

Professor David J.Rogers,
Taximetrics Laboratory,
Department of Biology,
Amory 101,
University of Colorado,
Boulder,
Colorado 80302, U.S.A.

6 August 1969

Dear Dave,

A Taximetrics Study .... Chrysobalanaceae

I have today received, corrected and returned the page proofs.

I suggest that you might like to distribute to the taximetrics men in America and that Iain might distribute to the American botanists. I will do the European distribution. Would you let me know please, by return, how many offprints you require?

With best wishes,

Yours sincerely,

p F White p Magrave-Secretary

(F. White)
Curator, Forest Herbarium

Signed in Mr White's absence

12 August 1969

Mr. Frank White, Curator
Forest Herbarium
Department of Forestry
University of Oxford OXI 3RB
United Kingdom

Dear Frank:

I have received the copy of your letter to Dr. Manning, and very deeply appreciate the good words you said therein. I hope that this will be the needed statement but I cannot be sure yet.

I also have your letter asking how many offprints we would like to have. We would like to have two hundred (200) offprints of the article, "A Taximetrics Study ... Chrysobalanaceae."

I look forward to seeing the paper.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Taximetrics Laboratory

Armory 101

31 July 1968

Dr. Frank White:
Department of Forestry
Commonwealth Forestry Institute
University of Oxford
Oxford, England

Dear Frank:

We are generally well satisfied with the write up, but as you might expect, have found some places which didn't seem to say what probably should be said. I will give the areas below, and some suggestions for them.

The first one is on page two of the manuscript, the last paragraph, starting "It might be argued that since the taxonomist . . . ." and ending with the last line on page 2, ". . . taxonomist's brain." We suggest that all of that be removed, and replaced with some statement including the following ideas:

It might be argued that were a worker to construct two classifications, one by conventional means, the other with the aid of a computer, the two might not differ significantly. If this were the case, what advantage does the computer give us? There are at least two major advantages. The computer is capable of manipulating data with a speed and accuracy far exceeding the capabilities of a conventional worker. Further, the classifications suggested by the computer are interpretable and communicable in the context of the methodolcgies under whose auspices the machine operates. END OF STATEMENT.

Return to the presentation as it now stands on the top of page three.

The next item which we feel very strongly about is found in the 2nd paragraph, page 4, beginning with "Natural classifications should . . ." and ending with " . . . by some workers in this field." Since many of our investigations have been directed towards removing the types of hurdles which are presented there, and since we have found ways in which the hurdles can be completely removed from the track, I think that this statement, at this time, is untimely, and not in keeping with what we know about the methods of structuring characters. However, to completely cover this important area we would need much more room than is available in this manuscript, and we should indeed write a separate paper on characters, their purpose, methods of construction, and practical considerations of their employment. So if you don't mind, let's just drop out that second paragraph on page 4, I don't think the flow of the argument is disturbed by its removal.

The last item where we wish to modify the manuscript is on page 6, second paragraph, the sentence which now reads, "The second is to develop logically a set of biological rules in the taxonomic area, discover the necessary and sufficient conditions to satisfy these rules, find mathematical models consonant with these rules, and then develop a practical working procedure which can be programmed to follow the mathematical model." We want to revise this sentence to read, "The second is to develop logically a set of biological rules in the taxonomic area, develop mathematical models consonant with these rules, and then discover a practical working procedure which can be programmed to follow the mathematical model."

Aside from these modifications, which I trust you find acceptable, the paper reads very well, and we'll be happy to have it submitted with author arrangement as you suggest.

I have corresponded with Mr. Bisby, giving him the necessary details of working with us, and one or two other points about the computer program he wanted to know about.

Thanks for your efforts - the time lapse you mentioned was not too bad, considering all the obligations which I know you have.

Best regards,

David J. Rogers
Professor of Biology

DJR:gm

Forest Herbarium
Dept. of Forestry
Commonwealth Forestry Inst.
University of Oxford OX1 3RB

16 July 1969

Professor David J. Rogers,
Professor of Biology,
Taximetrics Lab.,
Dept. of Biology,
Armory 101,
University of Colorado,
BOULDER, COLORADO 80302

Dear Dave,

Many thanks for your letter of 8 July which I received this morning.

I should be very pleased indeed to write a letter to Dr. Manning as you suggest. I hope, however, that it is not a matter of extreme urgency, since later today I am setting out on a fortnight's visit to the continent. Immediately on my return I will gladly attend to this matter.

I am very pleased to know that you think so highly of Frank Bisby. I am sure he has benefited immeasurably from his visit to your Laboratory.

The page proofs of our joint paper have been promised for "mid-July". I expect to find them on my desk when I return from the continent.

With best wishes,

Yours sincerely,

Dictated by Mr. F. White & Signed in his absence

Kathleen M. Cowley secretary

JUL 29 1968

TELEPHONE:
OXFORD 57891

DEPARTMENT OF FORESTRY
COMMONWEALTH FORESTRY INSTITUTE
UNIVERSITY OF OXFORD

FW/C

24 July 1968

Professor David J. Rogers,
Taximetrics Laboratory,
Dept. of Biology,
University of Colorado,
Boulder, Colorado, U.S.A.

Dear Dave,

When I wrote to you last on 7 May I fully anticipated being able to complete our Introductory remarks for The New Phytologist paper, within a few days. Unfortunately, my teaching load and other responsibilities were so arduous that I did not in fact find time to do this. Term is now over and I gave this job top priority and have recently completed what I hope will be the final draft. I enclose a copy for your ... comments.

Since a few of the points you made in your contribution over-lapped with points I have made in my introductory remarks I have taken the liberty of slightly altering your text in a few places. I hope in doing this I have not in any way misrepresented your views and that what I have done is acceptable.

I hope to deliver the entire work by hand to the editor of The New Phytologist in Cambridge within about a week. I don't think it will be going to press immediately so there should be time for any alterations you would like to be made, to be arranged.

Now that this introductory part is written I think it would be better published as an Introduction to the work on Chrysobalanaceae, rather than as a separate item. This being so I would like to suggest that the combined paper should be attributed to Prance, Rogers and myself, the names arranged like that in alphabetical order.

I am sorry that I have taken so long getting this piece of work ready for publication.

Mr. Bisby has recently prepared an account of what he has achieved so far in his research programme, and a statement of what he would still like to do. He is at present on holiday but when he returns he will be writing to you.

Best wishes,

Yours sincerely,

Frank

F.White

INTRODUCTION

This is an account of the way in which some taximetric techniques were used in the solution of the problem of generic delimitation in an Angiosperm family, the Chrysobalanaceae. One of us (F.W.) first became aware of the problem 19 years ago when supervising the work of a student on the West African species of Parinari (Trappes-Lomax, 1950). It was quite clear that the single genus Parinari was more heterogeneous than the remainder of the family, but equally clear that the problem could only be solved by studying all available material of all known species in the group, throughout its entire range. Some years later C.T.P. undertook this task and when his work was far advanced, motivated by curiosity, decided to submit his data to numerical analysis to see whether the techniques available would give taxonomically meaningful results and would provide further insights into the taxonomic situation. Subsequently, on moving to the New York Botanical Garden, in collaboration with the third author (D.J.R.), he subjected his data to analysis using the Wirth, Estabrook and Rogers model of clustering analysis, and so began a long and fruitful period of collaboration.

During the last few years there has been much debate concerning such things as the aims and purpose of numerical taxonomy or taximetrics,* the methods that can most properly be used, and the number, selection, weighting and treatment of characters. Many of the questions are still unresolved, and, such is the diversity of variation patterns within the plant kingdom that it is to be expected that different techniques will be appropriate

* These terms were independently coined about 10 years ago to cover the type of investigation in which computers are used to manipulate the data of taxonomy. Although neither term adequately conveys the operations undertaken we prefer the latter. The term 'numerical taxonomy' implies that computers are concerned with a special kind of taxonomy, a view to which we do not subscribe.

species level he is obliged to accommodate every species and every taxon of higher rank. Every species he omits could be a 'missing link'. As Burtt points out (1965) a numerical classification of the Gesneriaceae based on a 50% sample would probably leave out half of the really difficult problems. The reason why it is so easy to improve on the classifications of the Nineteenth Century is not because their authors were incompetent but because their material was incomplete. Any subsequent classification based on an incomplete but different sample is bound to be different, but not necessarily better, however many characters are used.

It is for reasons such as these that we think that our contribution may be of some interest, especially to those who are concerned about the potential application of taximetrics to practical taxonomy, especially the writing of world monographs. This is a field of endeavour at present greatly neglected, partly because of the effort involved, though it must be realised that whatever help may be provided by computers, at least for the foreseeable future, the greatest task is likely to be concerned with the assembly of material and scoring of data.

If taximetrics has any significant part to play in classificatory taxonomy it must either suggest classifications which are acceptable to taxonomists and their customers, or, if the classifications they suggest are not acceptable to orthodox taxonomists, they must be more acceptable to the customers than any alternatives the taxonomists propose. At present there is no indication that the second alternative will materialise.

It is our belief that taximetrics should serve as a tool to the taxonomist and that he alone is able to evaluate its results. In the present case, two of us (C.T.P. and F.W., especially the former) have lived with the problem of generic limitation in the Chrysobalanaceae for eight years and feel that we are qualified to do this.

It might be argued that since the taxonomist selects and scores the characters in the first place the computer is bound to confirm the classification he has arrived at without its aid, but this would only be true if the human mind was as reliable as a computer and the computer could be programmed exactly to simulate the workings of a taxonomist's brain.

In dealing with complex patterns the taxonomist is usually aware of a number of possible classifications and may not be able to decide between them, nor to evaluate quantitatively the different possibilities. Since under certain circumstances the computer may reveal irrational weighting of characters, inconsistency or error objectivity is increased.

Before we used the computer we were reasonably satisfied that most of our conclusions were correct, but there were a number of anomalous species about which we were undecided. The four different techniques taken collectively, abundantly confirmed our provisional conclusions concerning generic delimitation and provided insight on the relationships of the anomalous species which enabled us to decide their status with confidence. When a particular technique failed to do so it was usually found that the technique in question was not well adapted to handle that particular situation.

Rather than expect the reader to take on trust our statements concerning the part played by taximetric analysis we have given a fairly detailed account of our provisional taxonomic conclusions, and shown how they were confirmed or caused to be modified by the different methods. All the methods used contributed to our understanding of the group but the results obtained by the Wirth, Estabrook and Rogers method contained the greatest amount of taxonomically significant information. This is not surprising since it was specially designed to simulate taxonomic procedure as closely as possible.

Since to the lay public, at least, much confusion surrounds the subject of ~~numerical taxonomy~~ taximetrics and its exponents frequently disagree as to its aim and methods the following pages are intended to introduce the subject and make our own position clear. Although they have been mostly written by one of us (D.J.R.) they reflect the view of all three.

For generations taxonomists have produced classifications without paying much attention to their underlying assumptions and thought processes. From the time of Linnaeus, and particularly Jussieu, classifications have been increasingly 'natural', though that term has often meant different things at different times and to different persons.

Natural classifications should be based on a consideration of all available characters though at any particular level only certain characters are appropriate. In classifying species into genera characters that vary within species or are constant within the family cannot be used. This may place a stringent limit on the number of characters available, and it may not be possible to use the large numbers regarded as necessary by some workers in this field.

For most groups of higher plants taxonomists have done much of their work 'by eye', especially in the preliminary stages, sorting and resorting into groups on overall resemblance and, if experienced, making use of powerful and richly stored memories. At a relatively late stage the groups formed by eye are subjected to detailed analysis and comparison on the basis of the relatively few characters which have emerged as being of potential taxonomic worth. In the course of such a procedure countless comparisons have been made, many of them subconsciously or imprecisely. Although most classifications made in this way have proved to be useful they are still incomplete since botanical exploration and discovery are still incomplete and some parts of the classification have proved intractable. For these and other reasons taxonomists for a long time have attempted to discover procedures by which they could make their classifications more "objective". For example, Adanson attempted to improve the classification of plants by making many separate single character classifications and then attempted to merge them into one classification, preserving the information from as many of the separate classifications as possible. He was overwhelmed with the magnitude and complexity of the job because he did not have machines available with which he could manipulate the heavy load of data. For this reason, Adanson and subsequent taxonomists have had to fall back on "subjective" classifications, where all the data manipulation was done in the head, without a clear realisation of the actual steps in

the process. As a result, it was difficult for a taxonomist to explain to someone else just how he had reached his conclusions; the most valid test which could be applied to the value of his work was whether his published results actually supplied the information necessary for another person to identify the unknown. The large number of classifications which actually served this need testifies to the preponderantly large number of good intuitive classifiers.

The advent of computers within the last decade and a half has stimulated renewed interest in methods to manipulate the vast data banks available to taxonomists in their normal activities. A surprisingly large number of taxonomists have made more or less intense efforts to employ these incredibly fast machines, and two names have become associated with this type of effort, taximetrics and numerical taxonomy. Neither of these terms adequately conveys the operations undertaken but they both are intended to indicate the efforts to reflect various parts of the taxonomic process in sets of directions (programmes) for the computer to follow with a particular set of data.

It is well to recognize that "the taxonomic process" is made up of a number of quite different processes, each with its own rules and procedures. Therefore, to "do taxonomy on a computer" is a meaningless phrase by itself. One of the healthy aspects of attempting to use computers in taxonomy has been to force us to recognize the various parts of the discipline more clearly. We now define the process of classification, for example, and separate that process from identification. We recognise that classifications precede studies of phylogeny, and that the procedures used to produce a classification are not identical to those used when the phylogeny of a group is under study. Likewise, we identify more precisely the activities involved in nomenclatorial studies.

In a recent publication (Rogers, Fleming and Estabrook, 1967) an attempt is made to illustrate the diversity of taxonomic activities. The development of that illustration was absolutely essential, because we could not proceed with the development of useful computer programmes until we delineated the different aspects of the science, and could identify the types of activities we, as taxonomists, participate in when accomplishing

the tasks necessary for either monographic or floristic types of work. The illustration was not claimed to be a unique and absolute division of the science of taxonomy. Another systematist could produce a different type of breakdown though the major activities would certainly be similar.

The key factor in all of these efforts is the development of the stored programmes which are the sets of instructions to the machine as to what the investigator wants done with the data. The process of development can take one of two routes, both of which are being used in one laboratory or another. The first approach is to adopt mathematical methodologies already established, and modify the methodologies as results seem to dictate. We might term this approach as "heuristic". The second is to develop logically a set of biological rules in the taxonomic area, discover the necessary and sufficient conditions to satisfy these rules, find mathematical models consonant with these rules, and then develop a practical working procedure which can be programmed to follow the mathematical model. The latter, although perhaps more difficult, is preferable, inasmuch as one is aware at all stages of precisely what the model is doing, and what the results mean, both mathematically and biologically. It is probably true that we all start off in the heuristic approach, trying first this and then that procedure, and each time we fail or succeed we learn a little more about what may be sets of rules or what kind of mathematics may be appropriate.

Either procedure requires an interdisciplinary approach, wherein the taxonomist, the mathematician and the programmer work together. Each of the team members must be aware, though not necessarily an expert, of the processes inherent in the other disciplines. Perhaps the next generation of taxonomists will have all the necessary skills to perform adequately in each of the disciplines, but until we have given the proper background to our students,, we (as taxonomists) will have to work together with the mathematician and programmer. This is recognized as a difficult process because the vocabularies are frequently very distinct, and before effective communication can be established, we must be certain that we understand what words mean. The word "character" is a good example: it means a descriptor to the taxonomist, an alphanumeric symbol to the computer programmer, and has

no general meaning to a mathematician until defined. Taxonomists must even be sure which type of mathematician he is addressing, for different mathematical disciplines use the same word to mean different concepts. For example, the word normal: in algebra, normal is said of a subgroup N of a Group G whenever G/N is again a group; in topology, it is said of spaces with certain properties of regularity; in analysis, said of a line perpendicular to a surface at its point of intersection; and in probability, said of a distribution (limit of a repeated trials process)!

When speaking of taximetrics ~~or numerical taxonomy~~, we must be certain that it is understood that we are not substituting this type of work for the work, say, of biosystematics or chemotaxonomy, or classical morphological taxonomy. The data gathered by an individual is the information he thinks requisite or pertinent to solve some taxonomic problem. How ~~he uses~~ used should be decided by that individual that information is ~~in the realm of the taximetricist~~. This is, in a sense, theoretical taxonomy, and the application of computers is merely the end of the line for his work. We must consider the computer as just another instrument available to make the work of the taxonomist more objective, or more reliable, or more understandable. An analogy here is useful: when a stained section of a plant part is prepared and upon examination under the microscope we discover that we cannot distinguish a particular structure, we don't throw the microscope away, but rather go and make another slide, using different techniques which will make visible those structures for which a microscope is required. Likewise, the computer can array a vast quantity of information in ways which will allow us to "see" results which would otherwise be obscure.

But if computers are to be used sensibly and yield taxonomically useful information those who use them must acquire several skills. The aims of a course in taximetrics given in the Taximetrics Laboratory at Boulder, University of Colorado are to give the student a knowledge of the various mathematical, computer, and biological insights necessary to carry on his own investigation. In the biological part of this course, we try to define the concepts useful in constructing characters and attributes which will be useful in making a classification, determine what sort of classifications

are useful and desirable, how to prepare descriptions of characters which reflect biological thought processes, how to interpret the results of the computer programme in terms of taxa, etc. In the mathematical portion, we attempt to show how the mathematical thought processes can be useful in biological considerations, and in the computer portion we merely give the student sufficient understanding that he does not think the machine is some sort of magical black box. Above all, we want the student to understand that his role as a biologist is the most significant, and that he must not (and cannot) abrogate his responsibility to some mechanical device.

The most intensive investigations in computer taxonomy have been in the classificatory areas. Within this framework, attempts and some successful procedures have been developed to (1) analyze and standardize the characters used, (2) produce various measures of overall similarity, (3) assist in the establishment of clusters which may be assigned as taxa. In addition to the classificatory studies, efforts are now underway in many schools to produce computerized identification routines (keys), to develop programmes for phylogenetic studies (largely in the area of cladistics), and information retrieval and data manipulation systems to help curators with the mountains of clerical routine and to help investigators keep track of their specimens and data.

The computer print-out of clustering programmes can take several forms. Some prefer to have a "phenogram" where a series of lines drawn by the printer represent the individual taxa, the lines joining taxa at certain similarity values. We prefer a different format to present the results of the graph-clustering programme, wherein a series of partitions are given for the different levels of relationships, allowing the investigator to follow the building of the hierarchical levels.

Some examples of the application of computer methods to actual taxonomic problems are worth much more than volumes of theorising to convince (or vice versa) working taxonomists of their value. Such an effort is that of White and Prance, on the classification of the genera of Chrysobalanaceae which forms the bulk of this paper. Irwin and Rogers' paper on section Apoucouita of the genus Cassia (1967) gives a good description of one particular

procedure where a taxonomic problem was solved with the aid of a clustering programme. This gives the taxonomist a much more readily understood publication, couched in his own terminology. Several other taxonomists have been aided by the Boulder group in a variety of classificatory problems, but these papers have not yet been published (Hawksworth on Arceuthobium, for example). In all of these, the attitude taken ~~by our computer group~~ is that the specialist must (1) understand the workings of the computer programme, (2) recognize that the data which he prepares (as characters and attributes) will be the most important feature determining the computer results, and (3) the computer results are intended as "hints" about the classification which the competent specialist may either accept or reject. We must emphasize this last point. It would be patently ridiculous to insist that the specialist accept the results of the computer analysis if he did not agree with the results. The specialist must, however, have an open mind and be willing to determine (when the results disagree with his own decisions) whether some new insight may be gained from the machine's computations.

FW/C

Forest Herbarium, Dept. of Forestry,
Commonwealth Forestry Inst.,
University of Oxford.

11 April 1968

Professor David J. Rogers,
Taximetrics Laboratory,
Dept. of Biology,
Armory 101,
University of Colorado,
Boulder, Colorado, U.S.A.

Dear Dave,

I apologise for not having written to you before now but my last ten days in New York were pretty frantic, preparing thousands of specimens I had had on loan, for return.

I would like to thank you for the very warm reception you and your wife gave me while I was in Boulder a short while ago. I enjoyed my visit immensely and learned a great deal, although my stay was of such short duration.

I am now putting the finishing touches to Iain's paper on Chrysobalanaceae and hope to get it to the editor of the New Phytologist by the end of this month. I think it would be a good thing if the brief introductory paper could accompany it. I should be extremely grateful if you could let me have the few pages giving an outline of your approach to the problems of Numerical Taxonomy so that I can integrate them with my introductory remarks.

With best wishes,

Yours sincerely,

Frank White

F.White

Taximetrics Laboratory

Armory 101A

19 April 1968

Prof. Frank White
Forest Herbarium
Department of Forestry
Commonwealth Forestry Institute
University of Oxford
OXFORD, England

Dear Prof. White:

Enclosed is my informal set of remarks for your inclusion or exclusion as you see fit in the introductory comments. As you can see there is little description but I feel that the comments say about the kind of generalities you ought to make. Certainly the example (you and Prance) will go very far in making a meaningful illustration of computer methodologies.

As I said above, you may use this in any way that seems to fit with your own discussions. You may wish to put me on as junior author for the introductory paper. I trust that we can look forward to some more collaborative efforts.

Have you any more specific word about your student who was to join us for a while. We can foot the bill for his maintenance money while here. We look forward to having him with us.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

INTRODUCTORY COMMENTS
(For Frank White)

Taxonomists have attempted for a long time to discover procedures by which they could make their classificatory procedures more "objective". For example, Adanson attempted to improve the classification of plants by making many separate, level classifications and then attempted to merge the separate classifications into one classification, preserving the information from as many of the separate classifications as possible. He was overwhelmed with the magnitude and complexity of the job because he did not have machines available with which he could manipulate the heavy load of data. For this reason, Adanson and many other subsequent taxonomists have had to fall back on "subjective" classifications, where all the data manipulation was done in one's head, without expressing or realizing the actual steps in the process of correlating many data to reach a conclusion. As a result, it was difficult to explain to someone just how he had reached his conclusions, and about the only valid test which could be applied to the value of his work was whether his published results actually supplied the necessary information so that another person could identify an unknown. The large number of classifications which actually served this need testifies to the preponderantly large number of good intuitive classifiers and useful classifications.

The advent of computers within the last decade and a half stimulated renewed interest in methods to manipulate the vast data banks used by taxonomists in their normal activities. A surprisingly large number of taxonomists have made more or less intense efforts to employ these incredibly fast machines. As more work was done in this area two names have become associated with this type of effort, taximetrics and numerical taxonomy.

Neither of these epithets adequately reflect the operations undertaken under these titles, but they both are intended to indicate the efforts to reflect various parts of the taxonomic process in sets of directions (programs) for the computer to follow with a particular set of data.

It is well to recognize that "the taxonomic process" is made up of a number of quite different processes, each with its own rules and procedures. Therefore, to "do taxonomy on a computer" is a meaningless phrase by itself. One of the healthy aspects of attempting to use computers in taxonomy has been to force us to recognize the various parts of the discipline more clearly. We now define the process of classification, for example, and separate that process from identification. We recognize that classifications precede studies of phylogeny, and that the procedures used to produce a classification are not identical to those when the phylogeny of a group is under study. Likewise, we more precisely identify the activities involved in nomenclatorial studies.

In a recent publication (Rogers, Estabrook and Fleming, 1967) we attempted to illustrate the diversity of taxonomic activities. The development of that illustration was absolutely essential, because we could not procede with the development of useful computer programs until we delineated the different aspects of the science, and could identify the types of activities we, as taxonomists, participate in when accomplishing the tasks necessary for either monographic or floristic types of work. The illustration was not claimed to be a unique and absolute division of the science of taxonomy. Another systematist could produce a different type of breakdown though the major activities would certainly be similar. Indidentally, this "flow-chart" of taxonomy serves a very useful purpose in teaching students.

it gives a student a way to know what he must study in order to become a well-rounded systematist.

The key factor in all of these efforts is the development of the stored programs which are the sets of instructions to the machine as to what the investigator wants done with the data. The process of development can take one of two routes, both of which are being used in one laboratory or another. The first approach is to adopt mathematical methodologies already established, and modify the methodologies as results seem to dictate (we might term this approach as "heuristic"). The second is the logical development of a set of biological rules in the taxonomic area, discover the necessary and sufficient conditions to satisfy these rules, find a mathematical model consonant with these rules, and then develop a practical working procedure which can be programmed to follow the mathematical model. The latter of these two is preferable (though perhaps more difficult), inasmuchas one is aware at all stages precisely what the model is doing, and what his results mean, both mathematically and biologically. It is probably true that we all start off in the heuristic approach, trying first this and then that procedure, and each time we fail or succeed we learn a little more about what may be sets of rules or what kind of mathematics may be appropriate.

Either procedure requires an interdisciplinary approach, wherein the taxonomist, the mathematician and the programmer work together. Each of the team members must be aware, though not necessarily an expert, of the processes inherent in the other disciplines. Perhaps the next generation of taxonomists will have all the necessary skills to perform adequately in each of the disciplines, but until we have given the proper background to our students, we (as taxonomists) will have to work together with the mathematician and programmer. This is recognized as a difficult process because

the vocabularies are frequently very distinct, and before effective communication can be established, we must be certain that we understand what words mean. The word "character" is a good example: it means a descriptor to the taxonomist, an alphanumeric symbol to the computer programmer, and has no general meaning to a mathematician until defined. Taxonomists must even be sure which type of mathematician he is addressing, for different mathematical disciplines use the same word to mean different concepts. For example, the word normal: in algebra, normal is said of a subgroup N of a group G whenever G/N is again a group; in topology, it is said of spaces with certain properties of regularity; in analysis, said of a line perpendicular to a surface at its point of intersection; and in probability, said of a distribution (limit of a repeated trials process)!

When speaking of taximetrics or numerical taxonomy, we must be certain that it is understood that we are not substituting this type of work for the work, say, of biosystematics or chemotaxonomy, or classical morphological taxonomy. The data gathered by an individual is the information he thinks requisite or pertinent to solve some taxonomic problem. How he uses that information is in the realm of the taximetricist (taximetrician?). This is, in a sense, theoretical taxonomy, and the application of computers is merely the end of the line for his work. We must consider the computer as just another instrument available to make the work of the taxonomist more objective, or more reliable, or more understandable. An analogy here is useful: when a stained section of a plant part is prepared and upon examination under the microscope we discover that we cannot distinguish a particular structure, we don't throw the microscope away, but rather go and make another slide, using different techniques which will make visible

those structures for which a microscope is required. Likewise, the computer can array a vast quantity of information in ways which will allow us to "see" results which would otherwise be obscure.

While it would be nice to be able to give a precise review of the work now being done with computers for taxonomy, this is clearly an impossible task. Although I am in the middle of one of the most intensive studies of taximetrics, I still cannot be certain that I know what all other students are doing in the field. What anyone must do is to try to review the literature available and make some efforts himself with the processes. Much as the taxonomist has to have someone to show him the procedures to use an electron microscope, (must he) have someone to work with him in the necessary processes to use computers in classification. We offer a course in taximetrics, the aims of which are to give the student an over-view of the various mathematical, computer, and biological insights necessary to carry on his own investigation. In the biological part of this course, we try to define the concepts useful in constructing characters and attributes wh which will be useful in making a classification, determine what sort of classifications are useful and desirable, how to prepare descriptions of characters which reflect biological thought processes, how to interpret the results of the computer programs in terms of taxa, etc. In the mathematical portion, we attempt to show how the mathematical thought processes can be useful in his biological considerations, and in the computer portion merely give the student sufficient understanding that he does not think the machine is some sort of magical black box. Above all, we want the student to understand that his role as a biologist is the most significant, and that he must not (and cannot) abrogate his responsibility to some mechanical device.

The most intensive investigations in computer taxonomy have been in the classificatory areas. Within this framework, attempts and some successful procedures have been developed to (1) analyze and standardize the characters used, (2) produce various measures of over-all similarity, (3) assist in the establishment of clusters which may be assigned as taxa. In addition to the classificatory studies, efforts are now underway in many schools to produce computerized identification routines (keys), to develop programs for phylogenetic studies (largely in the area of cladistics), and information retrieval and data manipulation systems to help curators with the mountains of clerical routine and to help investigators keep track of their specimens and data.

The ~~form of the~~ computer print-out of clustering programs can take several forms. Some prefer to have a "phenogram" where a series of lines drawn by the printer represent the individual taxa, the lines joining taxa at certain similarity values. We prefer a different format to present the results of the graph-clustering program, wherein a series of partitions are given for the different levels of relationships, allowing the investigator to follow the building of the heirarchical levels.

Some examples of the application of computer methods for taxonomy to actual problems in classification are worth much more to convince (or vice versa) working taxonomists of their value. Such an effort is that of White and Prance, on the classification of the genera of Chrysobalanaceae. (Cite paper as following.) Irwin and Rogers' paper on section Apoucouita of the genus Cassia (cite) gives a good description of one particular procedure where a taxonomic problem was solved with the aid of a clustering program. This gives the taxonomist a much more readily understood publication, couched in his own terminology. Several other taxonomists have been aided in a

variety of classificatory problems, but these papers have not yet been published (Hawksworth on Arceuthobium, for example). In all of these, the attitude taken by our computer group is that the specialist must (1) understand the workings of the computer program, (2) recognize that the data which he prepares (as characters and attributes) will be the most important feature determining the computer results, and (3) the computer results are intended as "hints" about the classification which the competent specialist may either accept or reject. I must emphasize this last point. It would be patently ridiculous to insist that the specialist accept the results of the computer analysis if he did not agree with the results. The specialist must, however, have an open mind and be willing to determine (when the results disagree with his own decisions) whether some new insight may be gained from the machine's computations.

THE NEW YORK BOTANICAL GARDEN
BRONX • NEW YORK 10458 212/933-9400

Dear Professor Rogers,

I am working at the New York Botanical Garden until mid-April and in about a month's time will be flying down to Mexico to do some field-work with my former student, T. D. Pennington, who is at present on secondment there.

I should very much like to pay you a brief visit on my way to Mexico or on the way back, preferably the former, to discuss one or two matters of numerical taxonomy, chiefly concerning the work of another of my former students, G. T. Prance with whom you have collaborated, and the work of a present student, who is just beginning, F. A. Bisby.

I must apologize for the long delay in the publication of Francis' work, for which I am partly responsible. The draft he sent me two years ago was not very clearly presented and, although what he had to say might have been intelligible to the few people actually working in this field, it would not have conveyed much to the general botanist or to the taxonomist not well acquainted with mathematics. I think there is a great need for a paper or series of papers, specially written for such people and based on genuine taxonomic problems which in the course of the work have been satisfactorily solved — not as has so often been the case (e.g. Williams and Watson on Ericaceae / Epacridaceae) bogus problems which don't exist.)

I discussed this matter with one of the editors of the 'New Phytologist', which caters for the well-informed general botanist rather than narrow specialists. He said he would be prepared to publish such a paper (or papers)

provided it was likely to be intelligible to his readers. Ever since then I have worked on Prance's MS when I could spare the time with this object in mind. The main text is now complete and I am sending you a copy. I would very much appreciate your comments.

It is intended to be the first of a series. The second will deal with Pennington's work on Meliaceae, and the third and fourth will be based on the work Bisby is doing on rather different problems in Crotalaria and Diospyros. Then I think it might be possible to write a general paper critically discussing the value of numerical techniques.

The draft I enclose still needs some introductory matter relevant to the series as a whole and some concluding remarks on what we are at present working, but is otherwise complete. You will

notice that in the section dealing with your model of clustering analysis I have explained your technique in some detail by paraphrasing part of your Oncidiinae paper. I think this is desirable since many readers of the New Phytologist will not have ready access to the original. If I have misrepresented your method in any way I hope you will let me know.

It occurred to me that you have put in a great deal of trouble helping Iain with this aspect of his work and I think his account, since together with the subgraphs so beautifully demonstrates the application of your method ~~it~~ it would not be unreasonable for your name to be formally associated with the paper, either as a co-author or ... in collaboration with ... or in some such way. Iain agrees to this. I hope you ~~agree to this~~ do also. In any case, if the second paper, on the Meliaceae is to be satisfactory, it would I think require your collaboration which I hope you will be willing and able to give. But

THE NEW YORK BOTANICAL GARDEN
BRONX • NEW YORK 10458 212/933-9400

(5)

there are matters we can discuss later.

The man Bisby I mentioned earlier is a good mathematician, besides being a good botanist and is very intelligent. I hope that in the course of his work he would be able to visit your Department. I should like to discuss this possibility with you also. Bisby says that of all the numerical taxonomists with whose work he is familiar, you are the one he would most like to meet.

I could visit you on one of the following dates — Thurs. 22 Feb. Fri. 23 Feb. Sat. 24 Feb. Monday 25 Mar. Tuesday 26 Mar. I hope that one of them will prove convenient.

Yours sincerely

F. White

NUMERICAL TAXONOMY OF THE ANGIOSPERMS

I. GENERIC DELIMITATION IN THE CHRYSOBALANACEAE

by

F. White

Curator of the Forest Herbarium, University of Oxford

and

G. T. Prance

Associate Curator of Amazonian Botany, New York Botanical Garden

Second copy

Since 1960 we have been engaged on a study of the Chrysobalanaceae (Rosaceae, Chrysobalanoideae), especially concerning the circumscription of its component groups of species at and near generic level. Previous studies have been based on incomplete material of relatively few species. In particular fruit and seedling characters have been neglected. Our own work is based on herbarium material of 349 species, pollen slides of 65 species, correlated slides of secondary xylem of 35 species and seedling material of 25 species.

When we began, generic concepts were unsatisfactory. The differences separating groups of species in the single genus Parinari were often much greater than those separating other genera which have been kept apart for nearly 200 years.

choices - alternatives come only in pairs!

Once this was recognised we were faced with three alternatives.

a) To leave things as they are, in which case the degree of difference between genera and their internal homogeneity would vary greatly and the classification would fail to reflect relationships.

b) To merge all genera with the first described, Chrysobalanus.

c) To recognise certain groups of species in Parinari as distinct genera.

The reasons for adopting the last course seemed overwhelming. The proposed segregates from Parinari are based on several closely correlated characters and are easier to recognise and key out than several genera which have been universally recognised since the time of Linnaeus, both in the Chrysobalanaceae and in other families. On practical grounds the reasons for preferring the third to the second alternative are equally compelling. If the new genera are recognised only 31 new combinations are necessary. If the family is reduced to a single genus nearly 400 new names would be needed.

choice!

Although we had no doubt that some groups of species deserved to be recognised as genera there were a few borderline cases where the differences were less striking and we were undecided as to their best treatment. Since it was our intention to produce as consistent and objective a classification as possible we decided to test our tentative conclusions by subjecting the data on which they were based to numerical analysis using

an electronic digital computer,

The following account is concerned only with the numerical taxonomy of one of the two tribes, the Hirtelleae. No reference is made to the Chrysobalaneae since generic delimitation in that tribe was largely found to be acceptable.

During the first phase of our work which was done in the Forest Herbarium of Oxford University we used an Association Analysis programme written by Professor W. T. Williams (not discussed here) and two principal component analyses, one devised by Professor Williams, the other by Mr. J. N. R. Jeffers. Subsequently one of us (G.T.P.) moved to the New York Botanical Garden where he subjected the same data to two methods of cluster analysis, one invented by Wirth, Eastabrook and Rogers, the other by Rubin.

TAXONOMIC HISTORY

The Chrysobalanaceae is a woody group almost confined to the tropics. Although it was given family rank by Robert Brown as long ago as 1818, all the authors of the best-known and widely used systems of classification (De Candolle, Bentham and Hooker, Engler and Prantl and John Hutchinson) have treated it as a tribe or subfamily of Rosaceae. Nevertheless, nearly all workers with specialist knowledge of the group, e.g. Fritsch (1888) and particularly those who have studied its anatomy, e.g. Hallier (1903, using the results of the comprehensive study on leaf-anatomy by Kuster (1897)), Juel (1915, ovary-structure) and Bonne (1928, floral anatomy) have considered it to be sufficiently different from Rosaceae to be treated as a separate family. Both Metcalfe and Chalk (1950) and Erdtman (1952) imply that no objections could be raised against treating the group as a family on the basis of anatomy and pollen-grain structure respectively.

The most recent comprehensive treatment at generic level is that of Focke (1891) in Engler and Prantl's 'Die naturlischen Pflanzenfamilien'. The most recent world-wide treatment at specific level (De Candolle, 'Prodromus', 1825) was written at a time when less than ten-per-cent of species now known had been described.

Focke included the following genera: Chrysobalanus, Grangeria, Moquilea, Licania, Hirtella, Couepia, Parinari, Acioa, Angelesia, Parastemon, Lecostemon ('Lecostomion') and Stylobasium. The last two were included with some reservations and he suggested a relationship for them with Phytolaccaceae. Since Focke's time three additional genera have been described - Geobalanus, Magnistipula and Afrolicania.

In the first phase of our own work we assembled comprehensive taxonomic data based on morphology, anatomy, palynology and blastogeny for as many species as possible, with a view to deciding:

1) whether the group should be given family rank or treated as a subfamily of Rosaceae.

2) whether Lecostemon and Stylobasium should be included or not.

3) the number and circumscription of the remaining genera.

4) their best arrangement into tribes.

The results of this investigation comprised the major part of a doctoral thesis (Prance, 1963) and have been subsequently published elsewhere (Prance,) or will be published shortly (Prance, in press a, b).

Briefly our conclusions were as follows:

1) the group should be given family rank.

2) that true Lecostemon is a synonym of Sloanea (Tiliaceae), but that Lecostemon of Focke is the same as the subsequently described Rhabdodendron of Gilg and should be placed in its own family, close to Phytolaccaceae.

3) that Stylobasium should be placed in its own family, Stylobasiaceae, close to Sapindaceae.

4) that the following genera should be recognized: Chrysobalanus ~~(including Geobalanus)~~, Licania (including Angelesia, Geobalanus and Moquilea), Afrolicania and Parastemon in the tribe Chrysobalaneae and Hirtella, Couepia, Parinari, *Maranthes, *Cyclandrophora, *Neocarya, *Exellodendron, *Bafodeya, *Kostermanthus, Acioa, Magnistipula, Grangeria and *Hunga in the tribe Hirtelleae.

Those names preceded by an asterisk refer to new genera described by Prance (in press, b) or, in the case of Maranthes and Cyclandrophora, ~~an~~ old genera which never gained wide acceptance, but in our opinion should be revived. They had all been formerly included, at least in part in Parinari. The present paper is chiefly concerned with them.

PROBLEMS OF GENERIC DELIMITATION

Parinari was first described by Aublet in 1775 for two species from South America. Only four species were known

fifty years later when De Candolle described the Rosaceae for his 'Prodromus'. Although he did not appear to attach undue weight to this character, De Candolle mentioned that the drupe (and by implication the ovary) is 2-loculate. Ever since then most authors have used this character to diagnose the genus. As a result it has become increasingly heterogeneous.

De Candolle placed his four species in two sections, Petrocarya (Parinari if the International Rules are followed) to accommodate Aublet's original species and Neocarya for the other two. One of these, P. excelsa, should have been placed in Petrocarya; the other P. senegalensis, which had been described as P. macrophylla by Sabine the year before, is superficially similar to the other species but differs in several important respects and, in our opinion, should be placed in a separate genus.

Although it was De Candolle who first defined Parinari, perhaps inadvertently, on the basis of an artificial character which cuts across other more important resemblances, it was Bentham who consolidated the situation. In dealing with the Chrysobalanaceae for Hooker's 'Niger Flora' (1849), Bentham adopted the sections of De Candolle and added a third, Sarcostegia, to accommodate P. polyandra from West Africa and another species from tropical Asia. These two species are very different, both in structure and appearance from the original species of Parinari and Bentham did seriously consider making a new genus to accommodate them. Had he done this the history of generic delimitation in the Chrysobalanaceae would certainly have been very different. By choosing to characterise Parinari by the spurious dissepiment separating the ovules, Bentham paved the way for the increasing heterogeneity of the genus. Such was the authority of Bentham that subsequently all newly described Chrysobalanaceae, irrespective of other resemblances and differences were placed in Parinari if they possessed this character. Even so, Parinari as circumscribed when our work began was not even an artificial genus, clearly defined on the basis of a simple artificial character. In some species the partition is incomplete and traces of a partition can sometimes be found in species of other genera. A few species of Parinari, e.g. P. myriandra and P. heteropetala have unilocular ovaries.

Most authors of regional and local floras during the last

hundred years or so have either failed to notice the heterogeneity of Parinari or, if they did, have made no comments. A number of specialists however have drawn attention to the unsatisfactory situation, but without attempting to improve it. Juel (1915) expressed his doubts as follows: "Es scheint mir indessen zweifelhaft, ob die in allgemeinen angenommenen Gattungen der Chrysobalanoideen wirklich gut begründed sind, und ob nicht etwa die Gattungen Hirtella, Couepia, Parinari und Acioa zu vereiningen sind, oder auf andere Weise aufzuteilen." Hauman (1952) clearly had similar views. Of Parinari he said "très hétérogène, il conviendra sans doute de le diviser."

As our own work progressed it soon became apparent that Parinari was much more variable than any other genus in the Chrysobalanaceae and that it contained several groups of species which were more sharply defined than a number of other groups whose right to generic status had never been questioned. It seemed that the section Parinari and the section Sarcostegia differed from each other more than the members of any pair of currently accepted genera in the family. It was also clear that section Sarcostegàa had much more in common with the genus Couepia than with any other members of Parinari. If a consistent classification was to be achieved then either Parinari must be split, or most, if not all, of the genera in the family must be united.

It seemed that the first course was preferable but some groups of species were clearly much more distinct than others. We had little doubt that some groups should be given generic rank, but for others, where the resemblances were greater and the differences less, we were undecided. It was at this stage that we decided to test our tentative conclusions using the techniques of numerical analysis and an electronic digital computer. For this investigation we used all species of Hirtelleae for which we had adequate material. In addition to Parinari, Magnistipula, which has been treated differently by different authors, is the only genus which has been the subject of serious taxonomic disagreement since the time of Focke.

THE CHARACTERS USED

Since the purpose of the numerical analyses was to test a classification already produced by traditional taxonomic procedures, only those characters already known or believed to be of value in defining groups of species at and near the level of the genus were used. It is possible that if more, or different, characters had been used a better classification would have been produced, but there is nothing to indicate that this would have been so and the extra time involved would have been prohibitive. The characters are fully described in the original thesis (Prance, 1963) and more briefly below. Some of the double state characters could have been scored as multiple state characters, e.g. 1, shape of calyx - lobes and 2, distribution of hairs inside the receptacle tube. This would have been more objective, but it is unlikely that it would have seriously affected the results.

(a) Qualitative (double state) characters

1. (i) Calyx lobes acute (1) or rounded (0).
2. (ii) Receptacle hairy inside to the base (1) or not (0).
3. (iii) Ovary bilocular (1) or unilocular (0).
4. (iv) Bracts and bracteoles enclosing the flowers in small groups (1) or not (0).
5. (v) Stamens far exserted (1) or not (0).
6. (vi) Staminodes united into a "comb" (1) or not (0).
7. (vii) Ovary terminal at the mouth of the receptacle (1) or not (0).
8. (viii) Stipules enlarged (1) or not (0).
9. (ix) Two glands present at the base of the lamina or on the petiole (1) or not (0).
10. (x) Lenticels prominent on young flowering stem (1) or not (0).
11. (xi) Stamens united to form a single ligule (1) or not (0).

(b) Qualitative (multi-state) characters

12. Receptacle shape: elongate, symmetrical, hollow (I); elongate symmetrical, solid; (2), ventricose (3); saccate (4).
13. Fertile stamens occupying the whole (1), two thirds (2), or less than a half (3), of the perimeter.
14. Leaf-undersurface with stomatal cavities (1); softly and densely hairy (2); glabrous (3); with stiff but not dense hairs (4).
15. Bracts and bracteoles with many sessile or stalked glands (1), two sessile basal glands (2), or none (3).
16. Epicarp verrucose (1), smooth (or with a few hairs) (2), with a dense rusty tomentum (3), a dense covering of crustaceous warts (4).
17. Endocarp smooth (1), very rough and fibrous (2), hard and roughish (3).
18. Sēdling-escape by basal stoppers or plates (1), a single line of weakness (2), three or more lines of weakness (3), no special mechanism (4).
19. Inflorescence a panicle (1), corymbose panicle (2), elongated raceme (3), short subcorymbose raceme (4).

(c) Quantitative characters

20. Stamen number
21. Flower size measured in millimetres from articulation to apex of calyx-lobe.

TENTATIVE TAXONOMIC CONCLUSIONS

Before performing the numerical analyses a tentative classification of the genera had been formulated. This is summarised below. We were reasonably certain that the conclusions preceded by an arabic numeral in the following account were fully justified but were doubtful about those preceded by a small letter.

1. Acioa was found to be the only genus which had not been confused with other genera by previous workers. In our opinion its circumscription is satisfactory.

2. Magnistipula, which was described in 1905, has been much discussed and variously defined by subsequent workers. Most species of Magnistipula have at one time or other been placed in Hirtella. It appeared to us that this was unsatisfactory and we agreed with the conclusions of Graham (1957), that several African species of Hirtella should in fact be transferred to Magnistipula.

3. Hirtella and Couepia were frequently confused during the Nineteenth Century. The majority of species of both genera differ in a large number of characters. However, a few species of Couepia and Hirtella are somewhat similar and because of this the definition of these two genera has been unsatisfactory in the past largely due to the undue emphasis which has been placed on the single character of stamen number. Stamen number overlaps in the case of 3 species only. However, there are other correlated characters, particularly of the fruit, which permit a better distinction between Couepia and Hirtella to be made. Because of this they should be kept apart.

4. Granger ia has sometimes been confused with Hirtella. On the basis of the characters we examined Granger ia was found to be distinct from Hirtella. The fruit structure is very different from all other genera in the family except Parastemon (Chrysobalaneae).

5. The subdivision of Parinari. Parinari was found to be heterogeneous. Parinari subgenus Parinari (Parinari sens. str.) is very different from Parinari subgenus Sarcostemis (henceforth called Maranthes which is the earliest name for this group as a genus). These two subgenera are more different than any pair of genera in the family. Maranthes appeared much more closely related to Couepia than to the rest of Parinari. It was also concluded that other groups in Parinari, some already given subgeneric recognition and others not yet recognised, should be elevated to generic rank. For some other small groups of species and isolated

single species we were undecided as to whether they should be recognised as genera or not. These putative segregates of Parinari are each discussed below.

6. Parinari sens. str. The majority of species in this group form a closely knit group with a large number of correlated characters. There were three species usually placed in this group about which we were undecided, viz. —

a. Parinari benna. Although it agrees with Parinari in the majority of characters it differs in some fruit characters and in its receptacle shape. Since the differences are comparatively small we could not decide whether it should be segregated or not.

b. Parinari argentea-sericea and P. canaricoides. These two Asiatic species differ only slightly from Parinari but the leaves lack the stomatal cavities so characteristic of other Parinari sens. str.

7. Neocarya (Parinari subgenus Neocarya) macrophylla. It was clear that this species should be elevated to generic rank. It differs greatly from Parinari sens. strict. in receptacle shape, in the large number of stamens and in the glabrous interior of the receptacle.

8. Maranthes (Parinari subgenus Sarcostemia) is very distinct from the rest of Parinari, differing in fruit structure, the numerous exserted stamens, the solid receptacle, the epigeal germination, etc.

9. Cyclandrophora (Parinari subgenus Cyclandrophora) is worthy of generic rank. It differs from the rest of Parinari in having a very distinct fruit, in its receptacle shape, in the leaf undersurface and in the number of stamens. Although originally described in 1842, few authors since then have upheld it as a genus.

10. After dividing Parinari as outlined above, there remain a few species whose taxonomic position is not clear.

a. Parinari tessmannii which Hauman (1951) placed in subgenus

Pellegriniella differs from the rest of Parinari in a number of characters. The fruit structure is unlike that of all other species. This species has several important characters in common with Magnistipula.

b. Parinari barbata, P. coriacea, P. gardneri. These three species also formed part of Hauman's subgenus Pellegriniella although they are very different from P. tessmannii. Their flower structure is similar to that of Parinari sens. strict. but they differ in the fruit, the leaf-undersurface and in the bracts and bracteoles.

c. Parinari myriandra and P. heteropetala are obviously not true species of Parinari even in the broadest sense since they lack the bilocular ovary. They have united stamens as in Acioa and a general facies similar to some Cyclandrophora. At this stage we were undecided how to treat them.

NUMERICAL TAXONOMY

In natural classification a taxon is said to be closely related to another if they have a large number of characters in common, and distantly related if only a few characters are shared, so that taxonomic relationship or its converse, taxonomic distance, can be regarded as a measure of resemblance based on all available characters. In traditional taxonomic procedure, at least for the Angiosperms, the almost infinite number of comparisons that are necessary when a group is classified is largely done visually and is usually supported at a relatively late stage by a detailed analysis of the relatively few characters that have emerged from the precursory study as being of likely taxonomic worth. Taxonomists using these methods usually produce results acceptable to other workers and the independently obtained classifications of different workers are frequently similar, but taxonomic agreement is

far from universal and in critical groups, or when radically original classifications are proposed there is an advantage if the degrees of relationship can be assessed objectively and expressed quantitatively.

The quantitative approach to problems in classification is based primarily on the mathematics of matrices and although powerful statistical methods have been available to taxonomists for half a century the arduous and time-consuming computations involved set a severe limit to their use. The development of electronic digital computers has facilitated an extension of these methods and recent developments in this field have given rise to the description of numerical taxonomy or taximetrics.

Until recently the most usual approach to the problem of quantifying affinity has been to consider taxonomic resemblance as a function of the distance between the taxonomic units in multidimensional space, the co-ordinates being the characters. In the past, before computers were available, investigations using these procedures have generally been based on a study of the inter-relationships between characters utilising an "R" type matrix, i.e. a table of correlation coefficients between each pair of characters. These methods were originally applied, and are still best applied, only to certain restricted categories of quantitative data, and are therefore most useful in the investigation of relationships within relatively homogeneous groups.

They are of somewhat limited value to museum taxonomists who are frequently concerned with more heterogeneous groups whose members differ in both qualitative and quantitative characters.

Perhaps the most significant feature of numerical taxonomy has been the introduction of procedures which enable quantitative assessments of similarity between the objects to be classified (usually referred to as 'Operational Taxonomic Units' or OTUs) to be based on both quantitative and qualitative data. Consequently mathematical procedures can now be extended to the classification of relatively heterogeneous assemblages.

In outline the methods employed are extremely simple. An OTU/data table is drawn up in which all the characters are listed against the OTUs in coded form. The data table used in the present study is not reproduced here but is included in the original thesis (France, 1963). This

coding procedure may be in some cases rather complicated, and is discussed in detail by Sokal and Sneath (1963), but for some programmes it is relatively simple. In the case of two-state characters, for example, the presence or absence of a particular character, one state may be coded 0 (absent) and the other 1 (present). Similarly multistate characters are coded 0, 1, 2, 3 .. .. corresponding to the variation. Multistate quantitative characters (measurements of organs or the actual number of structures present) may be dealt with by dividing the range arbitrarily ~~with~~ into a number of (not necessarily equal) parts and coding as in the case of multistate qualitative attributes.

The coded data are used either to compute an R-type matrix of coefficients of similarity between all pairs of characters attributes or to compute a Q-type matrix of similarity between all pairs of the OTUs under consideration, i.e. a table depicting the similarity of each unit with each of the others.

Having drawn up the matrix of similarity coefficients the problem is to condense the multidimensional relationships into a comprehensible pattern in two or three dimensions with the minimum amount of distortion and loss of information.

PRINCIPAL COMPONENT ANALYSIS

Accounts of the principal component method of analysis are given by Thomson (1951), Kendall (1952), Rao (1952), Williams and Lambert (1961), Gardiner and Jeffers (1962) and Sokal and Sneath (1963). Gardiner and Jeffers applied it to higher plants but were concerned with species discrimination (in Betula) not with the taxonomy of higher ranks.

In the type of principal component analysis used, an R-type matrix of correlation coefficients between attributes is used to calculate for each component

a latent vector (eigen vector) giving the weightings needed to transform the data from attributes to component axes and a latent root (eigen value) giving the percentage of the total variance accounted for by a variate measured along that component axis.

The method can be explained in terms of a geometric model. Each of the attributes is represented as an axis in a multi-dimensional space and each individual is represented by a point in this space with attributes as co-ordinates. Individuals with identical attributes are represented by the same point, those with different attributes by different points. The dispersion of the points in the space is representative of the pattern of similarities between the individuals, and the distance between points is an actual measure of their dissimilarity.

In principal component analysis new axes are characterised, fewer in number than the original attributes, with which the data may still be adequately described. These axes are placed so that the variates measured along them have the maximum possible variance and the sum of the squares of distances of the points from them are minimum. Thus the first component will be the axis of the greatest variance in the space, the second component will be the axis, (perpendicular to) independent of the first, with the ~~second~~ greatest variance and so on. By this method a large fraction of the variance in the space is defined by the first few principal component axes. Then the model is examined visually by plotting two dimensional projections of the points as they occur on the principal component axes.

In these projections individuals that appear close together on all of them are closely related and those that appear distant on all of them are distantly related. However in many cases there are structural gaps in the model which are only observed in a few of the projections. These may be of importance, and the actual scored attributes (characters) important in the principal components showing the discontinuity, may be identified from their transformation weightings. Usually the projections of the first two principal component axes display the greatest discontinuities as they were selected to describe the maximum variance. These projections are objective displays of taxonomic relationship on which the

taxonomist may base subjectively his circumscriptions.

When ~~I~~ we started ~~my~~ his work few suitable programmes were available and none had been applied to problems of generic delimitation.

In 1963 Professor W.T. Williams who was then Professor of Botany at Southampton University kindly made available a programme for principal component analysis devised by him for us on a Ferranti Pegasus Computer. This programme would only use multistate and quantitative characters so that only 10 of the 21 characters originally scored could be used.

Shortly afterwards, Mr. J.M.R. Jeffers of the Forestry Commission Research Station, Alice Holt Lodge, Farnham, Surrey, allowed ~~me~~ us to use his principal component analysis programme which is able to use double-state, multistate and truly quantitative characters. Because of the limited storage space of the Ferranti Sirius Computer at Alice Holt Lodge, only 30 characters can be used, but since ~~I~~ we had only scored 21, this didn't matter.

So leave this sentence OUT

Principal component analysis of characters 12-21 using the programme of Williams

The proportional weightings were calculated for the first two principal components as only these two were sufficiently significant to be useful. This made it possible to plot the final values on a single two-dimensional scatter diagram (fig. 1).

The proportional weightings obtained by calculating the latent vectors of the original data required for transformation to the principal component axes are tabulated below. Those characters which contribute

most to a component have high values (ignoring the sign) and are underlined.

Weighting for original components, Williams PCA

Character	Component I	Component II
12	-0.0724	+0.1424
13	+0.5657	-0.2605
14	+0.4563	+0.7925
15	-0.6617	-0.3773
16	+0.2283	+0.7040
17	+0.6776	+0.0906
18	-0.1272	+0.5348
19	-0.0966	+0.4583
20	-0.7927	+0.2189
21	-0.8028	+0.3328

Component I (The characters involved are +13, -15, +17, -20, -21)

The position of the stamens and the type of endocarp contrasted with the nature of the glands of the bracts and bracteoles, the number of stamens and flower size.

Component II (+14, +16, +18)

The type of leaf under-surface, the type of epicarp and the method of seedling escape.

Once the components were found, the original data ~~was~~ were normalised by calculating

$\frac{x_{ij} - \bar{x}_i}{s_{xi}}$ for each value x_{ij} (where x_{ij} is the value in the original data for the character i in the individual j, $\bar{x}_i$ is the mean of the values of character, i for all individuals, and s_{xi} is the standard deviation of the values for character i)

Since the original data values are expressed in various absolute units, they are replaced by normalised values which are relative to their means and expressed in units of their standard deviation.

Using this normalised data the value of each component was computed for each of the 108 species and these values plotted on a scatter diagram (fig. 1) using the two components as axes.

The taxonomically significant groupings which are detected by this method are summarised below.

1. GRANGERIA. All three species occur together forming an isolated group.

2. PARINARI sens. str. All species occur together forming an isolated group.

3. COUEPIA. All species occur together forming a group close to Magnistipula.

4. HIRTELLA. All species occur together forming an isolated group.

5. CYCLANDROPHORA. All species occur together and form an isolated group, except that it also contains Parinari heteronetala

6. MARANTHES. All species occur together and form an isolated group except for the close proximity of Parinari macrophylla

7. EXELLODENDRON. Three species of Parinari, P. barbata, P. gardneri and P. coriacea, which we subsequently decided to assign to the genus Exellodendron, form an isolated group in close proximity to

P. canarioides

8. MAGNISTIPULA. All species (including the former Parinari tessmannii) form a compact group in close proximity to Couepia.

9. PARINARI BENNA occupies an isolated position.

Principal component analysis of characters 1-21 using the programme of Jeffers

In this analysis it was found that the first five components were necessary to display the taxonomic relationships effectively. The proportional weightings for these are given in the table below. Those characters which contribute most to a component have high values (ignoring the sign) and are underlined.

Proportional weighting for components:-

Character	I	II	III	IV	V
1	+0.0280	-0.0597	+0.0204	-0.0134	+0.0634
2	+0.0879	-0.0109	+0.0201	+0.0259	+0.0688
3	+0.0229	+0.0169	+0.0197	+0.0494	-0.0134
4	+0.1000	-0.0074	-0.0311	-0.0209	+0.0154
5	-0.0818	+0.0343	-0.0380	+0.0218	+0.0148
6	+0.0128	+0.0221	+0.7401	-0.0081	-0.0794
7	+0.0221	+0.0484	+0.0254	-0.0528	+0.0775
8	+0.0292	-0.0042	-0.0129	-0.0087	+0.0169
9	+0.0858	+0.0319	-0.0136	+0.0161	-0.0470
10	-0.0771	-0.0053	-0.0061	+0.0133	-0.0164
11	-0.0049	+0.0037	+0.0192	+0.0581	+0.1000
12	+0.0063	+0.0114	+0.1000	-0.0138	-0.0555
13	+0.0377	-0.0719	+0.0452	+0.0046	+0.0274

(Proportional weighting for components contd.)

Character	I	II	III	IV	V
14	-0.0874	-0.0459	+0.0273	+0.0202	-0.0095
15	+0.0492	+0.0829	-0.0012	+0.0015	+0.0497
16	-0.0517	-0.0193	+0.0729	+0.0537	+0.0499
17	+0.0046	+0.0808	+0.0591	-0.0198	+0.0096
18	-0.0812	+0.0131	+0.0063	-0.0799	+0.0486
19	-0.0440	+0.0103	-0.0153	+0.1000	-0.0113
20	-0.0202	+0.0956	-0.0181	+0.0195	+0.0061
21	-0.0229	+0.1000	+0.0147	+0.0186	+0.0062
Percentage of variation accounted for by components	24.3%	15.2%	9.4%	7.4%	6.7%

Component I

+2 hairs on inside of receptacle or not.

+3 bilocular or unilocular ovary.

+4 bracts enclosing flower buds or not.

+9 glands present on the petiole or not

+10 prominent lenticels present on young stems or not.

-5 stamens exserted or included.

-18 type of endocarp dehiscence.

Component II

+15 nature of the glands on the bracts.

+17 type of endocarp.

+20 number of stamens.

+21 flower size.

Component III

+12 shape of receptacle.

Component IV

+19 type of inflorescence.

-18 dehiscence of endocarp.

Component V

+7 level of ovary insertion.

+11 staminal ligule present or not.

-6 staminodes united or free.

Seventeen of the original twenty-one characters are given significant weightings in these five components, which account for 68.2% of the total variation (calculated from the latent roots). The table shows that the first component accounts for 24.3% of the variation. That a single component accounts for a comparatively small percentage and that the first five components account for only 68% of the total variation indicates the complexity of the variation within the Hirtelleae. This is partly due to the type of data used. Principal component analysis is more suitable for, and gives components accounting for, a greater fraction of the variation with exclusively or predominantly quantitative data. Only two quantitative characters were used in the case of Hirtelleae. The components beyond the fifth account for less than 5% each and are not worth considering (they represent less variation than any one of the original characters). The fact that there are five significant components means that each must be plotted against the other four. All ten diagrams were prepared for the original thesis but only one which illustrates the projections of the first and second components is reproduced here. The results from all ten diagrams are summarised in Table 1 Fig. 2 is an example of these projections of the components using their values for each individual species calculated in the same way as in the previous component analyses. It is also possible to prepare isometric diagrams to include three components in the same figure, but this was not necessary here.

Fig. 2 shows that a number of genera are clearly demarcated when these components are projected.

Since a number of components are involved, it is obvious that no single pair of components will isolate all the genera. There is always a residue of mixed clusters. Conversely genera which are not isolated when one pair of components are projected may be isolated when different

components are projected. The groups which were isolated by the ~~eleven~~ ten projections are summarised below.

Components	Pure groups	Mixed groups
I/II (Fig.2)	1. PARINARI sens. str. 2. P. BENNA 3. MARANTHES 4. P. MACROPHYLLA 5. P. CANARIOIDES 6. P. TESSMANNII 7. EXELLODENDRON	1. Couepia + Cyclandrophora + Magnistipula pro parte + P. myriandra 2. Hirtella + Magnistipula pro parte + Grangeria
I/III	1. MAGNISTIPULA (with P. TESSMANNII as a satellite) 2. PARINARI sens. str. 3. P. BENNA 4. P. MYRIANDRA	1. Maranthes + P. macrophylla + Cyclandrophora + P. 2. Hirtella + Couepia + Grangeria.
I/IV	1. PARINARI sens. str. 2. P. BENNA 3. EXELLODENDRON + P. CANARIOIDES 4. P. MYRIANDRA	1. Maranthes + Grangeria + Cyclandrophora 2. Hirtella + Couepia + Grangeria + Magnistipula
I/V	1. PARINARI sens. str. 2. P. BENNA 3. P. MYRIANDRA 4. EXELLODENDRON 5. CYCLANDROPHORA	1. Hirtella + Couepia + P. canarioides + P. tessmannii + Magnistipula + Grangeria
II/III	1. MAGNISTIPULA + P. TESSMANNII 2. COUEPIA MACROPHYLLA 3. P. BENNA	1. Cyclandrophora + Maranthes + Couepia + P. macrophylla + P. myriandra 2. Exellodendron + Parinari sens. str. + P. canarioides + Grangeria + Hirtella
II/IV	1. P. MYRIANDRA 2. COUEPIA MACROPHYLLA 3. CYCLANDROPHORA 4. MARANTHES 5. GRANGERIA 6. P. MACROPHYLLA	1. Couepia + Magnistipula pro parte 2. Hirtella + Parinari sens. str. + P. canarioides + Exellodendron + Magnistipula pro parte + P. tessmannii + P. benna
II/V	1. CYCLANDROPHORA 2. GRANGERIA 3. P. MYRIANDRA	1. Couepia + Maranthes + Magnistipula pro parte + P. macrophylla 2. Hirtella + Parinari sens. str. + Exellodendron + P. canarioides + P. tessmannii + P. benna + Magnistipula pro parte

Components	Pure groups	Mixed groups
III/IV	1. P. MYRIANDRA 2. CYCLANDROPHORA 3. MAGNISTIPULA 4. P. TESSMANNII 5. MARANTHES + P. MACROPHYLLA	1. Couepia + Hirtella + P. benna + Parinari sens. str. + P. canarioides + Exellodendron
III/V	1. P. MYRIANDRA 2. MAGNISTIPULA + P. TESSMANNII 3. CYCLANDROPHORA 4. P. BENNA 5. P. MACROPHYLLA 6. MARANTHES pro parte 7. GRANGERIA	1. Couepia + Hirtella + Parinari sens. str. + Maranthes pro parte + P. canarioides
IV/V	1. P. MYRIANDRA 2. CYCLANDROPHORA 3. GRANGERIA 4. MARANTHES pro parte	1. Maranthes pro parte + Hirtella pro parte + P. macrophylla 2. Couepia + Hirtella pro parte + Parinari sens. str. + Exellodendron + P. canarioides + Magnistipula

Confirmation of the tentative taxonomic conclusions by principal component analysis

1. Acioa. Because of the uncritical nature of this genus it was not included in the Jeffers PCA described above. However a second analysis was made in which 13 and an additional 19 species from the other genera. species of Acioa were included. The results were similar to those obtained by the first analysis except that Acioa was clearly separated as a distinct group by the two most significant components.

2. Magnistipula. The 13 species of Magnistipula (some of which have frequently been placed in Hirtella) are separated into a distinct cluster by the Williams PCA and in four projections of the Jeffers PCA. They are clearly separated from Hirtella in both analyses. Parinari tessmannii which is discussed later nearly always occurs with Magnistipula. The conclusion that Graham was correct in placing most African species of Hirtella in Magnistipula is abundantly confirmed by these analyses.

3. Hirtella and Couepia. The Williams PCA clearly separates these two genera (fig. 1), but indicates that they are closely related. The Jeffers PCA never sorted Couepia and Hirtella into pure groups, but they

are usually found in different mixed groups, indicating that these two genera are closely related but separable. It is clearly demonstrated that Couepia and Hirtella are much more closely related to one another than the various segregates of Parinari are to each other.

The Williams PCA indicates that even the borderline species fall clearly into one genus or the other, for example Couepia dodecandra and C. polyandra are placed with the other species of Couepia. As a result of this analysis, and in the light of the fruit structure it is possible to state more definitely that these species belong to Couepia.

No change in the circumscription of either Couepia or Hirtella is necessary and the characters used by Martius & Zuccarini (1832) are confirmed as the best. In several projections of both component-analyses Couepia macrophylla (59) is isolated from the rest of the genus. This is because of its much larger flowers and the large number of stamens (two hundred more than in any other species of Couepia). This species is only isolated from the rest of Couepia by the component including flower size and staminal number. I do not consider that differences in these characters merit the creation of a new taxon for C. macrophylla, which in every other respect is a typical member of Couepia.

4. Grangeria. Both the Williams and the Jeffers PCA confirm this genus is distinct from Hirtella and other genera. It is isolated by four different projections in the Jeffers PCA, and is placed far from Hirtella by the Williams PCA.

5. Parinari (segregates). The most obvious conclusion from the results of the PCAs is that Parinari is markedly heterogeneous. The PCAs show that subgenera and other groups of species within Parinari are as distinct from each other as are many long-established genera. If the tribe Hirtellene is to be divided into genera at all, Parinari must be split. In the Jeffers PCA, the pure groups listed in the table consist more frequently of segregates of Parinari than of the other genera.

6. Parinari sens. str. Both PCAs demonstrate that this is the most distinct group within the family. It is isolated by four projections of

those
the Jeffers PCA involving ~~the~~ components which account for the greatest percentage of the original variation. In the Williams PCA its isolated position can be seen in fig. 1. The component analysis also helps to determine the status of certain species formerly placed in Parinari subgenus Parinari about which we had been undecided.

a) P. henna : This species is isolated by both PCAs. Fig. 1 shows that it is isolated by the Williams PCA. In the Jeffers PCA it is separated by no less than six of the eleven projections. This clearly indicates that it cannot be kept in Parinari sens. str. It is therefore assigned to a new unispecific genus, Bafodeya.

b) P. argentea-sericea and P. canarioides are placed between Parinari sens. str. and Exellodendron by the Williams PCA. In the Jeffers PCA they are isolated into a pure group by a single projection, and are placed with various species of Parinari sens. lat. in the other projections. The PCA indicate that these two species do not merit recognition as separate genera but that they are, nevertheless, significantly different from the rest of Parinari sens. str. P. canarioides differs from true Parinari in the absence of stomatal cavities from the leaf under-surface, and in the bracts which are small and caducous. P. argentea-sericea differs only in the absence of stomatal cavities. These are important characters and this group clearly presents a borderline situation but in our opinion the differences between P. canarioides and its relatives and Parinari sens. str. are not sufficient to justify the recognition of the former as a genus. Kostermans (1965) has recently placed these two species in a distinct subgenus.

7. Neocarya. Parinari macrophylla is slightly isolated by the Williams PCA and is isolated by four projections of the Jeffers PCA. The decision to treat this species as a unispecific genus is confirmed.

8. Maranthes forms a distinct group in the Williams PCA and in three projections of the Jeffers PCA. Both analyses indicate that it is a distinct genus.

9. Cyclandrophora is very much isolated by the Williams PCA (fig. 1) and is isolated by six projections of the Jeffers PCA. The analyses indicate that it is also one of the most distinct genera in the tribe.

10. Uncertain species. The residual species about which we were undecided were also definitely placed by the PCAs.

a) Parinari tessmannii is placed with Magnistipula by both analyses. The Jeffers PCA indicates that it occupies a slightly isolated position in the genus. As a result of these analyses it was decided to transfer P. tessmannii to Magnistipula, and to create a separate subgenus for it.

b) The three species, Parinari barbata, P. coriacea and P. gardneri, were always grouped together by the PCAs. The Williams PCA places them in a fairly distinct group, and the Jeffers analysis groups them apart in three projections. They never form pure groups with P. tessmannii, indicating that Hauman's (1951) subgenus Pellegriniella contained two diverse groups. The three species mentioned above differ from Parinari in significant fruit and vegetative characters, and as a result of the analyses, it was decided to create a new genus, Exellodendron, for them.

c) Parinari heteropetala was only grouped with Cyclandrophora by the Williams PCA, probably because its most significant feature of a staminal ligule was not included in the data used. In the Jeffers PCA it is isolated by eight of the ten projections, that is more often than any other group. This implies that it is an extremely isolated species. It is placed, together with another species not used in the analyses, in the new genus Kostermanthus.

THE RUBIN MODEL OF CLUSTERING ANALYSIS

Full details are given by Rubin (1966 and in press). His method is of general application but was devised with biological problems in mind.

It is a Q-technique since the first stage is the computation of a similarity matrix in which all the objects are compared with each other. The purpose of the clustering procedure is to partition the objects into classes (or clusters) where the average similarity of the objects inside the classes exceeds a particular 'breaking' value of similarity' (S^*) and where the average similarity value between the members of different classes is as low as possible. These two factors are incorporated into a single criterion. Each partition is measured for the extent to which it satisfies this criterion, and the partitions which maximise this measure are sought. The model also provides a useful measurement of the stability of an object in a class. An object is considered stable if its average inside similarity with other members of the class in which it is a member exceeds S^*, and if this is not true with members of any other classes (i.e. the average similarity of its own cluster is greater than S^*, and its average similarity with any other cluster is less than S^*). Hence an object is stable if there is little question about its class membership.

Once an initial partition of the group has been made, the effect of moving each object into every other class is calculated. If the new partition is better (i.e. gives a lower average similarity value between classes) the object is moved. This process, termed the 'Hill-climbing Algorithm' by Rubin, is continually trying to find a partition which is better than the best one previously found according to the measurement of stability. If the value of S^* is low, the number of classes formed will be few and will tend to consist of a large number of objects, since many of the pairs of objects will share similarity measures greater than the breaking value (S^*); similarly for high values of S^* clusters will be small and many. Partitions obtained for increasing values of S^* are not necessarily hierarchical

because large clusters formed at low values of S^* need not wholly contain the smaller ones formed at higher values of S^*.

When this method was used for the Chrysobalanaceae data, the set was run at three values of S^*, 0.3, 0.4, 0.5, using 140 species (i.e. including Acioa). The clusters formed at the different levels of similarity are summarized in fig. 3.

Discussion of the clusters level by level

0.3 level

At this level only two clusters are formed. It is interesting that one of these consists of all of the species of Parinari subgenus Parinari plus two other species of Parinari, P. benna and P. chrysophylla. P. benna is in fact most closely related to Parinari/ subgenus Parinari, and might be expected to cluster with it. A valuable feature of the Rubin programme is the print out of the stability of each object in its cluster as well as the effect on the clusters of moving it to the nearest related cluster. Only P. chrysophylla is noted as being unstable at the 0.3 level. This is not surprising as this species is a member of Parinari subgenus Maranthea, and is not closely related to other members of the cluster. P. chrysophylla has an average inside similarity much smaller than any of the other species in the cluster. Thus the grouping at the 0.3 level is satisfactory considering the taxonomy of the group.

0.4 level

At this level eight clusters are formed. It is at once apparent that the clusters have been made by further division of the old genus Parinari sens. lat., and the segregation of some of the other genera traditionally accepted as distinct. Three clusters consist exclusively of species of Parinari sens. lat., while species of Parinari also form parts of three other clusters. All ~~of~~ the other longstanding genera of the tribe are separated into clusters at this level, i.e. Hirtella, Grangeria, Acioa and Couepia. The two latter genera are in clusters with no additional species from other genera, the other genera are all in clusters together with a few species of Parinari. Hirtella is in a cluster together with three species of Parinari (P. coriacea, etc.),

now segregated into the genus Exellodendron. In this cluster the three species of Parinari (Exellodendron) and one species of Hirtella are indicated as being unstable. Magnistipula is segregated into a cluster together with Parinari tessmannii. All objects are considered relatively stable in this cluster. It is interesting that P. tessmannii is placed here, since other evidence shows that it should be transferred to Magnistipula. Grangeria is in a cluster with Parinari macrophylla and P. canarioides. Other evidence shows that these two species are not closely related to Grangeria and they are shown to be unstable in this cluster. Parinari subgenus Maranthes forms a cluster on its own. The member of this subgenus that was placed with subgenus Parinari at the 0.3 level crosses over to the expected position. The fact that members can cross over between clusters indicates that the clusters are not necessarily hierarchical in nature. The crossing over of P. chrysophylla to the Maranthes cluster leaves Parinari subgenus Parinari in a pure cluster except for the closely related P. benna. All members of this cluster are stable. Parinari subgenus Cyclandrophora is also in a cluster of its own together with P. argenteo-sericea.

0.5 level

Ten clusters have been formed at this level. Three of these are unchanged from the previous level, i.e. the clusters formed by Couepia, Acioa and Parinari subgenus Maranthes. Furthermore, no unstable species occur in these clusters. The remaining seven clusters have all changed slightly between the two levels. P. benna has crossed over from the Parinari subgenus Parinari cluster to form part of a new cluster (group 2). Parinari subgenus Parinari is now in a cluster on its own with all members of the cluster stable. Fig. 3 shows that the main change between the 0.4 and 0.5 levels is that some species from five 0.4 level clusters cross over to form a new cluster (2). The other important change is that Parinari subgenus Neocarya is separated from Grangeria to form a cluster consisting of the single member of this subgenus. This segregation by

the model is pleasing since it had already been decided that subgenus Neocarya should be separated into a distinct genus.

The result of the formation of cluster 2 at the 0.5 level is to leave the other clusters as pure groups as proposed by Prance in the new classification i.e. Parinari sens. str. (Fig. 3 cluster 1), Macnistipula (3), Neocarya (4), Grangeria (5), Cyclandrophora (6), Maranthes (7), Hirtella (8), Acioa (9), Couepia (10).

Confirmation of tentative taxonomic conclusions using the Rubin clustering analysis

1. Acioa forms a pure cluster at the 0.4 and 0.5 levels.

2. Macnistipula together with Parinari tessmannii (see 10a, below) forms a pure cluster at the 0.4 level but M. albida is transferred to the dump cluster at the 0.5 level.

3. Couepia forms a pure cluster at the 0.4 and 0.5 levels and is never confused with Hirtella which forms a cluster with Exellodendron at the 0.4 level and a pure cluster at the 0.5 level.

4. Grangeria forms a cluster with two anomalous species of Parinari at the 0.4 level and a pure cluster at the 0.5 level.

5. Parinari (segregates). That Parinari sens. lat. should be split is clearly indicated by the fact that its species occur in 6 of the 8 clusters formed at the 0.4 level.

6. Parinari sens. strict. is the only cluster separated at the 0.3 level.

a) P. benna is associated with Parinari sens. strict. and one other unstable species at the 0.3 level, only with Parinari sens. strict. at the 0.4 level and belongs to the dump cluster at the 0.5 level.

b) P. argentea-sericea and P. canarioides are associated with Cyclandrophora and Grangeria respectively at the 0.4 level and are placed

in the dump cluster at the 0.5 level.

7. Neocarya is isolated as a pure cluster at the 0.5 level.

8. Maranthes forms a pure cluster at the 0.4 level and remains unchanged at the 0.5 level.

9. Cyclandrophora, together with P. sericeo-argentea form a cluster at the 0.4 level and a pure cluster at the 0.5 level.

10. Uncertain species.

a) Parinari tessmannii is associated with Magnistipula at the 0.4 level and joins the dump cluster at the 0.5 level.

b) P. barbata, P. coriacea and P. gardneri form a cluster with Hirtella at the 0.4 level and joins the dump cluster at the 0.5 level. These three species have the highest inside similarity in this cluster.

c) P. heteropetala is not separated from Cyclandrophora at the 0.4 and 0.5 levels because its diagnostic characters were not scored.

Discussion of the Rubin method

This method clearly demonstrated the main thesis for making changes at generic level. It showed that if Couepia, Hirtella, Acioa and Magnistipula are to be recognized as genera then Parinari must be split. Of the genera proposed in the revision (Prance, in press) all the large ones (Parinari sens. strict., Maranthes and the four just mentioned) and three of the small ones (Neocarya, Grangeria and Cyclandrophora) form clusters at the 0.5 level. The remaining cluster at this level is an extremely heterogeneous "dump" cluster containing Parinari henna which Prance has transferred to Bafodeya, P. tessmannii which I placed in Magnistipula and with which it is associated at the 0.4 level, P. congoicoides and P. argenteo-sericea which I retain in Parinari sens. strict. although they differ in some respects, the small segregate genus Exellodendron and Magnistipula albida which was correctly placed at the 0.4 level.

The distinctness of most genera is very clearly shown by this method but the formation of a dump cluster for the critical species is a serious disadvantage. It is for the correct placing of such species that the taxonomist may be tempted to use numerical techniques for guidance.

The formation of a dump cluster probably occurs because the method forms clusters by a measure of goodness of the whole partition rather than by maximising the goodness of the individual clusters in the light that each object must be meaningfully classified in the end. To change a good cluster such as number 2 weakens the others considerably. The taxonomist is more interested in optimal clusters than in optimal partitions, Optimal clusters do not necessarily provide optimal partitions.

The model also provides much information about the objects and their clusters. The print out, in addition to showing the stability of objects in their clusters, provides, 1. The effect of moving each object to the most closely related cluster, 2. the average similarity between clusters, 3. The average similarity of each object to each group, and 4. A variable frequency analysis for each cluster, which gives the occurrence of each attribute. These features add to the value of the model because each one of them is informative in the interpretation of the results. It gives a good picture of the relationship within and between clusters. In the present form where a dump cluster is formed the use of the similarity matrix between clusters is somewhat obscured, although in general, clusters which the author would expect to be most closely related, have the highest similarity values. The variable frequency analysis is useful for the definition of the clusters in the terms of the characters used. This feature is not shared by the other models used.

THE WIRTH, EASTABROOK and ROGERS MODEL OF CLUSTERING ANALYSIS

As in the Rubin method this is a Q. technique since the first stage is the computation of a similarity matrix in which all the objects are compared with each other. The method of clustering is fully described by its authors (1966). It is based on graph theory and the clustering technique is similar to the single linkage method described by ~~Sokal and~~ Sneath ~~(1963)~~. The similarity function which is determined for every pair of species varies between zero and 1. Zero indicates complete dissimilarity: 1 complete similarity.

Sneath '57

Clusters are then formed using this similarity function. For the purposes of this technique a cluster is defined as follows:

(1) a collection of OTUs is isolated for some fixed value of similarity if each OTU in the collection is less similar to every OTU outside the collection than that fixed similarity value.

(2) an acceptable cluster is a collection of species (OTU's) isolated for some fixed similarity value, but which contains no smaller cluster isolated for the same fixed value. In this way members of a cluster show a discontinuity with non-members, and clusters cannot be subdivided without the parts being less isolated than the whole.

All pairs of species at least as similar as some fixed similarity value are linked, and these aggregates of linked species constitute acceptable clusters. The result of this linking process is a partition of the OTUs into equivalent classes (clusters). The partitioning of the collection into a number of clusters can be carried out for all distinct* fixed linking similarity values; the clusters of a partition formed for a low linking similarity will be made up entirely of clusters of a partition formed for a higher linking similarity value for if two species are linked for a high similarity they must also be linked for a lower similarity. A hierarchy can be formed by arranging the collection of partitions in order of decreasing fixed similarity

Ice

* Two values are distinct if there are more than distinct partitions.

values. The disjoint partition is the first formed, in which all species are unlinked single objects. The next partition is formed by lowering the linking similarity until a different partition is found. This process continues until all species are in the same cluster, which is known as the conjoint partition.

In addition to the clustering procedure, this technique provides other valuable data.

(1) The "moat" value of a cluster is a numerical value of the degree of isolation of a cluster. If a cluster forms at a linking similarity of C, then the "moat" value of that cluster is found by subtracting from C the greatest similarity between any pair of species, one within the C-cluster and the other outside the C-cluster.

(2) The "internal connectedness" of a cluster, based on a comparison of the minimum number of linkages between the species of a cluster required to hold that cluster together and the total possible number of linkages between the species of a cluster.

(3) When the membership of a cluster changes, for example if two previously isolated clusters join, the identity of the species forming the new linkage is easily found. Any new linkages forming within a cluster are also pointed out by the computer as the clustering proceeds, so both the external and internal relationships of a cluster are known.

Although there can be as many different isolated clusters as one less than twice the number of OTUs in the study, there are usually far fewer in practice. If all possible partitions are made and arranged in succession and all the internal connections properly inserted, these levels can be thought of as single frames in a strip of motion picture. The film would show the OTUs initially as distinct points clustering together hierarchically through all the allowable partitions and terminating as a single cluster.

The results of the computer analysis can be shown in two ways, by means of subgraphs and by the skyline plot, each reflecting different properties of the suggested classification.

In the subgraphs for any chosen level of clustering every pair of specimens sharing a similarity at least as great as the given linking

similarity c_i is connected by a line (c_i is the linking similarity used to make the i^{th} partition). The connecting lines are shown in three degrees of boldness : the darkest lines are made for those similarities that changed the previous partition to the present one (i.e. connected previously unconnected groups); the intermediate lines represent the internal structure which existed previous to the present situation and the finest lines are those connections which will form for similarities less than c_i but greater than $c_i + 1$. Use of this custom makes it possible to visualise exactly how the clusters change as the similarity criterion drops gets less.

Figs. show how clusters were formed for the Hirtelleae. In this case the conjoint partition is the fourteenth. Since the number of values of c is small it has been possible to show all but two in the subgrephs. With more complex data a large number of partitions may occur.

The 'skyline' plot (figs.) which is prepared by the computer summarizes the results of clustering processes by showing the clusters which formed and the value of similarity at which they formed, their hierarchical relationships, and the measure of isolation associated with each cluster. The specimen number is placed across the bottom of the plot and a vertical similarity scale runs up the side. The skyline plot is a convenient method of summarizing the taxonomic relationships detected by the method, but since it does not indicate the degrees of relationship among cluster members nor the objects linking one cluster to another, nor the steps in cluster development, the subgraphs are necessary as well.

Confirmation of tentative taxonomic conclusions using the Wirth, Eastabrook and Rogers clustering analysis

1. Acioa was not included in this analysis.

2. Magnistipula forms a pure cluster at the eighth level. M. glaberrima (111) does not join until the tenth level when it joins strongly. M. albida (114) which was kept apart from the rest of the genus by the Rubin method is strongly associated with the rest of the genus by this method.

3. Couepia and Hirtella form distinct clusters which are not associated until the tenth level. H. guyanensis (95) does not join the Hirtella cluster until the ninth level, but then it joins strongly and obviously belongs there.

4. Grangeria forms a distinct cluster which is not associated with Hirtella until the tenth level.

5. Parinari. The fact that Parinari sens. strict., Maranthes, Parinari (Bafodeya) benna, Cyclandrophora, P. (Neocarya) macrophylla and Exellodendron form distinct clusters at the eighth level clearly indicates that this genus should be divided.

6. Parinari sens. strict. forms an isolated cluster with a high moat.

a) Parinari benna (31) occurs as an isolated species until the twelfth level when it joins Parinari sens. strict.

b) Parinari argenteo-sericea (124) and P. canarioides (33). The former joins the Parinari sens. strict. cluster at the eighth level confirming the author's opinion but the latter remains isolated until the tenth level when it joins Magnistipula. ← what is the A's opinion on this?

7. Neocarya remains isolated until the final level when it is linked to a species of Couepia. At this level it is the only species without at least three connections with other species.

8. Maranthes forms an isolated cluster at the eighth level.

9. Cyclandrophora forms a pure cluster at the eighth level except that Parinari (Kostermanthus) myriandra, which differs chiefly in characters not used in the analysis, is attached.

? isolated?

10. Uncertain species.

a) Parinari tessmannii remains isolated until the thirteenth level when it links with Magnistipula albida (114) but by this time Magnistipula itself has merged with several other genera.

b) Parinari barbata, P. coriacea, P. gardneri (Exellodendron) form an isolated cluster with a strong moat at the eighth level.

An effort must be made by the taxonomist to describe his objects adequately!

c) Parinari myriandra (Kostermanthus) is associated with Cyclandrophora at the eighth level because the characters used in the analysis did not take sufficient account of some of its most characteristic features.

The subgraphs and the skyline plot show how well this model has clustered the data. Every species has been placed in a cluster which corresponds to one of the genera tentatively proposed on the basis of the preliminary study except for Parinari myriandra whose characters were not adequately represented by the features scored, and Parinari canarioides which is thus singled out for further study and re-assessment. The fact that there is no dump cluster is a considerable advantage over the Rubin method.

As in the other numerical methods used it is clearly demonstrated that the segregates of Parinari proposed as genera are at least as worthy of generic rank as long-standing genera in the family such as Hirtella and Couepia.

Because this method displays the way in which clusters are formed and degrees of relationship, both internal and external are recorded, it is of great value in the study of relationships between clusters, a feature not shared by the other two methods used. The clusters formed at the seventh to ninth level correspond well to the genera proposed by the author. It is at the tenth to fourteenth levels that the links between clusters indicate the relationships of genera.

At the tenth level the strongest connection occurs between the Hirtella and Couepia clusters. This is interesting since the Jeffers principal component analysis was unable to separate these two genera. The graph theory model shows that they form two good independent clusters with a single most but are nevertheless closely related.

Another interesting union, that between Maranthes and Couepia occurs at the tenth level, demonstrating, as was suspected from the preliminary study, that Maranthes (formerly Parinari subgenus Sarcostegia) is more closely related to Couepia than to Parinari sens. strict.

At the same level Grangeria becomes linked to Hirtella. It is interesting that Magnistipula does not join up with Hirtella until the next, eleventh level, so confirming the views of those who have advocated their separation.

From the beginning of our study it was clear that Parinari subgenus Parinari not only occupies an isolated position within Parinari sens. lat. but in the subfamily as a whole. The fact that this is the last of the larger clusters to lose its separate identity abundantly confirms that this is so.

The measure of isolation of the clusters, the most value, also provides useful indications. The generic clusters formed at the eighth level have similar most values. These are as high as, and in some cases, higher than, the most values at any other level. In many of the clusters the most is in the range .030-.055. The clusters of Hirtella formed from the second to fifth c values have mosts of .002-.005. These clusters are not particularly significant as regards the taxonomy of the tribe. When, however, all the species of Hirtella have joined at the sixth c value the most value increases significantly to .035.