



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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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.

January 27, 1973
Box 447
Blanding, Utah

Dear Dr. Rogers:

Since I last wrote you in November progress on my dissertation has been very good. It now looks like it will be finished by spring. However, I do have one question. I have assumed but am not certain that average not single link clustering was used. Could you verify this for me?

Next week I'll be sending in the final chapters on the first draft. As soon as I get committee approval and everything finalized I will send it off for publication. I have two sources interested and so expect results. You will receive a copy of the publication when it is off the press. Again, may I say how useful the program has been. It has a lot of potential for archaeology.

Sincerely,

Dee F. Green
Dee F. Green

ARIZONA STATE
UNIVERSITY

DEPARTMENT OF ANTHROPOLOGY

TEMPE, ARIZONA 85281

Dr. David J. Rogers
University of Colorado
Department of Environmental, Population
and Organismic Biology
Boulder, Colorado 80302

January 23, 1973

Dear Dr. Rogers:

To start with, I would like to apologise for taking so long to answer your letter of Nov. 14th. Between semester breaks and attempting to complete a thesis as well as several other projects, my correspondence has fallen way behind.

Dr. Gains was correct in informing you that I had been working with your cluster program. The applications I have used have centered mostly around chronological and cultural ordering of various archaeological manifestations. Though I have used some ceramic data, the majority of my work has been restricted to using lithic assemblages (principally Epipaleolithic groups from Israel).

At the present time I have only working papers of the work I have done and there are a number of problems with interpretation because of the methods used to make the collections themselves. These problems center around the issue of whether the tool assemblage does actually represent a true picture of each group.

Whether these issues are ever resolved or not, my experience with this particular method of ordering as it applies to archaeological data has indicated to me that it offers a dependable and very convenient way of handling this material and does provide insights into the existing relationships within and between sites and levels within a site.

I hope that these comments serve to answer your questions and, if at any time in the future a paper or papers results from this work, I will see that you are sent a copy.

Sincerely,

Landon Douglas Smith

Landon Douglas Smith
LDS:hs

October 26, 1972

Carol Tidwell, Programmer
Computer Center
Arizona State University
Tempe, Arizona 85281

Dear Carol:

In response to your question about conversion of our Similarity-Clustering Program from CDC 6400 to UNIVAC 1100 series, I suggest you contact:

Mr. Henry Fleming
GURC
NASA/MTF
Bay St. Louis, Miss., 39520

I believe that he has recently put this program up on a UNIVAC 1108, but what the availability is, I cannot say. Nor can I say that the program Mr. Fleming has running is still the same as it was on the CDC. Anyway, I am sure that he is willing to tell you what is happening at the NASA Mississippi Test Facility with regard to this program.

I recently had a request for the Similarity-Clustering Program from Dr. John R. Sturgul of the Dept. of Geosciences at the University of Arizona, Tucson. I earlier responded that he could best get the program from people in your department of Zoology. If you don't mind, I will suggest to him that he could get the programs directly from you, rather than going through the Zoology department.

You noted in your letter that the program is now being used by the Department of Anthropology. Is there any way that you could list the applications there, or put me in touch with users in the Anthro Dept? I am always interested in applications of our systems, and would like to keep in touch with as many different applications as possible. Thank you for help.

Sincerely yours,

David J. Rogers
Professor of Biology

October 26, 1972

Carol Tidwell, Programmer
Computer Center
Arizona State University
Tempe, Arizona 85281

Dear Carol:

In response to your question about conversion of our Similarity-Clustering Program from CDC 6400 to UNIVAC 1100 series, I suggest you contact:

Mr. Henry Fleming
GURC
NASA/MTF
Box St. Louis, Miss., 39520

I believe that he has recently put this program up on a UNIVAC 1108, but what the availability is, I cannot say. Nor can I say that the program Mr. Fleming has running is still the same as it was on the CDC. Anyway, I am sure that he is willing to tell you what is happening at the NASA Mississippi Test Facility with regard to this program.

I recently had a request for the Similarity-Clustering Program from Dr. John R. Sturgul of the Dept. of Geosciences at the University of Arizona, Tucson. I earlier responded that he could best get the program from people in your department of Zoology. If you don't mind, I will suggest to him that he could get the program directly from you, rather than going through the Zoology department.

You noted in your letter that the program is now being used by the Department of Anthropology. Is there any way that you could list the applications there, or put me in touch with users in the Anthro Dept? I am always interested in applications of our systems, and would like to keep in touch with as many different applications as possible. Thank you for help.

Sincerely yours,

David J. Rogers
Professor of Biology

School of Botany
University of Melbourne
Parkville, Victoria 3052
Australia
26-5-71

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

Dear Professor Rogers,

I should greatly appreciate receiving a
listing of the program CHARANAL described by BISSY (1970)
New Phytol. 69:1149-1160. for use in my research in the
taxonomy of the genus Eucalyptus.

Yours faithfully,

Philip Smith

ARIZONA STATE
UNIVERSITY

COMPUTER CENTER

TEMPE, ARIZONA 85281

October 18, 1972

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

We have a version of your Similarity-Clustering Program which we obtained from Kenneth McWilliams when he was doing graduate work here in 1970. Since we are presently undergoing a conversion from a CDC 6400 to a UNIVAC 1106, I was wondering if you have a more recent version of your program and if you know where it is running on a UNIVAC 1100-series computer.

We are fortunate to have your program which is mainly used by the Department of Anthropology - although it has had some other curious applications.

Sincerely,

Carol Tidwell

Carol Tidwell
Programmer

CT:sd

Taximetries Lab.

March 31, 1971

Mr. William E. Harmon
203 Tucker Hall
Department of Botany
University of Missouri
Columbia, Mo., 65201

Dear Mr. Harmon:

In response to your letter of the 24th, I have the following comments and suggestions. The Graph Clustering Analysis for the IBM 7044 is a rather old version of the program we now have. If you want it, you may obtain it from the Program Information Department, IBM, 40 Saw Mill River Road, Hawthorne, New York 10532. Ask for a listing and documentation for Library #3501. The present version, however, is written for a CDC 6400 machine, and conversion to IBM equipment would be more troublesome because of differences in word length. I will be glad to send you a listing and documentation for the CDC program, should you be willing to get it converted for the IBM 360-65.

The character analysis program has never been fully programmed for the 7044, and is presently also running on the CDC 6400. Again, if you want to make the conversion, we would be happy to supply a listing and documentation (at cost, of course).

An alternative suggestion you may wish to consider is that we process your data here on our equipment, saving you the bother of making any conversion. Perhaps you could arrange for the costs of running it here. It may depend on your budget. If time is a problem for you, and money is not too tight (say \$200) I believe this alternative is the easier way out.

Whichever way you choose, please let me know, and I'll go along with your desires.

Sincerely,

David J. Rogers
Professor of Biology

Please note change of address.

University of Missouri

COLUMBIA
65201

COLLEGE OF ARTS AND SCIENCE
Department of Botany

XXXXXXXXXXXX
XXXXXXXXXXXX

March 24, 1971

Dr. David J. Rogers
Taxonomy Laboratory
Dept. of Botany & Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

I am a graduate student nearing the completion of my Ph.D. under the direction of Dr. D. B. Dunn at the University of Missouri. As you may know, Dr. Dunn and his students have been working on the systematics of the genus Lupinus. I have been working with a group of small flowered lupine species for several years, long enough to come to the realization that the problems of developing a monographic treatment of these species, or for that matter, the entire genus, will require the aid of statistical modeling techniques such as those that have been developed in your laboratories. Our problems in describing the taxa of Lupinus are not unlike those encountered by Dr. Irwin and yourself in your taximetric treatment of the Sect. Apoucoutia of the genus Cassia. For this reason, I would like to employ the same graph theory model and the information theory model for character analysis as techniques of analyzing the fantastic variability we are faced with in working on Lupinus.

I understand that your laboratory makes available programs designed to carry out these functions. The University of Missouri has an IBM 360-65 computer available for our use. I have been told by our computer technicians that conversion of programs written for the IBM 7044 to the IBM 360-65 is not a big problem. Because I have had minimal experience in computer programming, it would be less time consuming for me to obtain programs which require a minimum of conversion from one system to another.

I would appreciate any assistance you can give me in my efforts to obtain these programs. If there is any cost involved in getting them duplicated, etc. please let me know.

Thank you for your consideration and cooperation.

Sincerely yours,

William E. Harmon

William E. Harmon
203 Tucker Hall

4

TAXIR is a very powerful, completely generalized info and stor. retr. syst.,

based on a novel retrieval technique which involves direct Boolean calculations on the stored data rather than the traditional linear file search.

This technique results in significant savings in both machine storage

space and machine execution time. The user addresses the system in a

(English, French, German, etc.)

high level language which permits him complete control over building, ~~data~~

~~banks~~ ^{↓ and} querying ~~them~~ and updating ~~them~~ ^{data banks.}

The input data may be presented to the system in either fixed-field

or free-field format, and these formats may be changed by the user at any

time. Data records may be added to the data bank or deleted from at

desired intervals throughout the operative life of the data bank. A

powerful correction statement provides complete file maintenance capability.

Although the first pilot version of TAXIR operated only on small

data banks that were core-resident, the system has since been extended

~~to handle data banks~~ ^{resident} in auxiliary core storage. There

is no intrinsic obstacle to further extending this capability to data

banks resident on tapes or disks, which will permit the system to handle *infinitely large*

data banks, ~~of any size.~~ ~~in~~

In defining data to the system, the user need not be limited by conforming to any ~~xxxxx~~ predefined set of values or any predefined hierarchical structure. On requesting responses to ^{queries}~~queries~~, however, the user may instruct the system to organize the output in any hierarchical structure he desires. The system checks the data as it enters the machine for a wide variety of error conditions and prints messages to the user warning him of incorrect procedures or errors.

The query language of TAXIR is the language of Boolean algebra, fully generalized, so that the user may select subsets of the data bank for print out, using any possible Boolean combination of selection criteria. Thus, his queries may be very simple ^{ones}~~queries~~ based on a single criterion, or may be as complex as he wishes, involving ~~xxxx~~ combinations of many such criteria. The use of parentheses/in Boolean expressions ^{and the operators NOT, AND, OR} ~~are~~permitted. Boolean expressions, therefore, may be written by the user to select a single item from the data bank or any possible combination ^{of items}~~of items~~ from the data bank, or ^{indeed}~~the~~ entire data bank. This latter capability permits the printing of hierarchically-structured representations of the entire data bank, which in the case of museum applications, can be the catalogue. Since the hierarchical organization is under the user's control, various hierarchical permutations of the catalogue may be created. For example,

alphabetically

a catalogue ordered/on taxonomic name, or a catalogue ordered alphabetically

on geographical location or a catalogue ordered numerically on catalogue

number ^{all are} ~~is~~_λ within the capability of the system. This same Boolean power

permits the user to ^{employ} ~~use~~_λ the querying language as a simultaneous key for

the identification of unknowns.

1. Correction - complete file maintaining capability
2. Not true that compiled data base must be present at file creation.
Can add items at any time
Correction statement built on base
3. Core limitations not true.
~~Not~~ not inherent limitations (disk).

Predefine - not true.
for coding only.

Automated error messages present.

Merand

INTERNATIONAL COUNCIL OF SCIENTIFIC UNION
INTERNATIONAL BIOLOGICAL PROGRAMME

Section HA: Human Adaptability

J. S. WEINER, CONVENOR
ROYAL ANTHROPOLOGICAL INSTITUTE
21 BEDFORD SQUARE
LONDON W. C. 1.

HA 154

11th May, 1971

Dr. S.G. Appan,
Department of Biology,
University of Colorado,
Boulder,
Colorado 80302,
U.S.A.

Dear Dr. Appan,

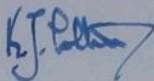
USA/HA/31

Thank you very much for your letter of 27th April concerning TAXIR and Professor Rogers' participation in the IBP/Human Adaptability programme.

I would much appreciate a copy of the TAXIR manual when it is available and I much appreciate the explanatory note on the Retrieval System and the reprints.

Thank you again for your co-operation.

Yours sincerely,



K.J. Collins,
(H.A. Scientific Co-ordinator)

INTERNATIONAL COUNCIL OF SCIENTIFIC UNIONS
INTERNATIONAL BIOLOGICAL PROGRAMME

Section HA: Human Adaptability

J. S. WEINER, CONVENOR
ROYAL ANTHROPOLOGICAL INSTITUTE
21 BEDFORD SQUARE
LONDON W.C.1.
TEL 01-636 8985

HA. 154

16th April, 1971

Dr. D.J. Rogers,
Department of Biology,
University of Colorado,
Boulder, Colorado 80302,
U.S.A.

Dear Dr. Rogers,

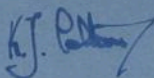
USA/HA/31 - (Biological information retrieval system.)

During the present research phase of the I.B.P. we are endeavouring to maintain contact with the progress of work in each of the Human Adaptability projects with a view to eventually preparing a catalogue summary of all H.A. research and findings. It would, therefore, be of great assistance if you could give me some indication of

- (a) the current status of the above project and an estimate, if possible, of the date of completion
- (b) papers and publications relevant to your investigation when they become available.

With kind regards,

Yours sincerely,



Dr. K.J. Collins
(H.A. Scientific Co-ordinator)

27 April, 1971

Dr. K.J.Collins
H.A. Scientific Co-ordinator
International Biological Programme
Human Adaptability Section
21 Bedford Square
London, W.C.1., England

Dear Dr. Collins:

With reference to your letter dated 16 April addressed to Professor David J. Rogers, I am sending herewith a brief note on TAXIR and some reprints.

Since the development of the pilot version of TAXIR in 1967-1968, the system has undergone continuous improvement. A users manual explaining the system as it stands today is under preparation. A copy can be sent to you when it is ready.

The TAXIR system will be undergoing further intensive and extensive development for the use in the Gulf Environmental Program (GEP) of the International Oceanographic Decade.

Sincerely,

Dr. S.G.Appan
Research Associate of Prof. D.J.Rogers

c/o Tropical Products Institute
56/62 Gray's Inn Road
London, W.C.1
England.

8th April 1971

Ans. to Conway
4/21/71

Prof. B.N. Okigbo
Dean, Faculty of Agriculture
University of Nigeria
NSUKKA
East-Central State
Nigeria.

Dear Professor Okigbo,

You will recall that when we were at the Seminar on Root and Tuber Crops in West Africa at Ibadan recently, we discussed with Professor Rogers and Patrick Chinwuba the question of the taxonomy of the West African yams of the Dioscorea rotundata/D. cayenensis group, and the need for some statistical taximetric work on the profusion of cultivated forms to be done. I indicated that I did not think that my own Institute would be likely to want to be involved, although I personally am exceedingly interested in the matter, and would like to be concerned, and that I would try to find some means of supporting the work.

At the time, I believe I mentioned to you my friend Dr. Ayensu, a Ghanaian gentleman who is currently Chairman of Botany at the Smithsonian in Washington. He is himself extremely interested in the yams: he did his Ph.D. on the genus, has published a number of papers of various aspects of the Dioscorea and their close relations, and currently has a book in press on 'Dioscoreales' as Volume VI of Metcalf's series on the Anatomy of the Monocotyledons. One of his most recent papers dealt specifically with this very question of the taxonomy of D. rotundata/D. cayenensis. Fortunately, he was passing through London only very shortly after my return from West Africa, and we were able to discuss the matter: these discussions were very encouraging, and I am now writing to you in the hope that we may be able to get something going. It will, of course, be necessary for you to initiate formal action with Dr. Ayensu: this letter is primarily to let you know that an application is likely to receive a sympathetic hearing from him, although at the same time I would like to sketch out a few ideas of my own. I am sending copies of this letter to Dr. Ayensu and also to Prof. Rogers, so that they know what is 'in the air'.

The work, as I see it, would be very much on the lines we discussed at Ibadan - the collection of as many as possible of the cultivated forms of yam ~~as possible~~, and their morphological description in a manner that permits the data produced to be handled by computer. We would, at least initially, confine our attention to the D. rotundata/D. cayenensis group, and perhaps also to the very similar D. alata, which, although not indigenous, exists in West Africa in many cultivated, and possibly feral, forms. The study would involve the preparation of a very large volume of data - based on measurement, observation, etc., - on each form in the collection, and also the preparation of herbarium specimens, photographs, etc. Initially, it would probably be convenient to cover only the area of the former Eastern Region, and immediately adjacent parts of the North (there is a lot of very good yam country in the southern parts of what used to be Benue Province, which is geographically easily accessible from a base in Nsukka), but as the work proceeded it would be desirable to cover as much of West Africa as is relevant (vide the map in the paper which I gave at Ibadan). The contacts from Dahomey, Togo and the Ivory Coast which we made at the Seminar should be especially useful in this context. (I visited the Ivory Coast on my way back - the collection at IRAT Bouake is really quite comprehensive).

I would think that the work would occupy a Graduate Research Assistant for three years (at least in the first instance), and it would be desirable to provide him with an assistant ('A' levels in science) to help with the routine work. The graduate would need to spend a period in the U.S.A. fairly early on in the three years, to work with Dr. Ayensu and with Prof. Rogers: he could spend some time with

me in London on his way out or back. Probably, his work would be quite suitable for a Ph.D. thesis, but that is a purely internal matter for you to decide.

So that the project would be independent of the availability of other transport facilities, I would like to suggest that we should cost the purchase and running of a Landrover type vehicle into the budget from the first. At the same time, we could include such items of equipment of which you may be short, which you will need - especially such expensive items as photographic equipment, as a good camera and accessories will be needed for record making.

Presumably, Dr. Ayensu, Prof. Rogers and ~~myself~~ myself will each want to pay a visit to West Africa while the work is in progress: in fact, it would seem to me desirable that we should all be able to have a meeting at Nsukka fairly soon after the man selected returns from his period overseas. Our fares and subsistence should therefore be costed into the project.

At a very rough provisional costing, we could perhaps think on lines somewhat resembling the following:-

Graduate Research Assistant's stipend for 3 years	£4000
Technical Assistant's salary	£1500
GRA's overseas travelling expenses	£1000
Landrover - purchase	£1500
- running costs	£500
G.R.A.'s local travelling expenses	£2000
Miscellaneous equipment	£1000
Expenses for visiting consultants	£1500
Labour and incidentals of maintaining living plant collection	£3000
	£1000
Total	£15000

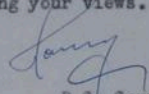
Beyond this, there are two other factors which I am not in a position to calculate, at present, but which should be mentioned. Firstly, Prof. Rogers will presumably need funds for the purchase of time on the computer, and possibly other expenses. Secondly, if I am to spend any appreciable amount of time working with the man in West Africa - which I would hope to do - I would very likely have to take unpaid leave of absence from my employers, and so would need some form of stipend for the period, quite apart from expenses, travelling, etc.

Obviously, with your knowledge of local conditions, you will be in a much better position than I to work out details, but I include these figures as some sort of basis from which to work. From my recent discussions with Dr. Ayensu, I would think that there would be a reasonable prospect of obtaining funds of this order of magnitude, although of course the essential decisions must be made between you and him - my role in this matter is very largely one of a catalyst.

I do hope that you will find these suggestions of interest. I think this could be a very worth-while operation indeed, and go a long way towards clarifying the ~~genetics~~ genetics of these important West African food crops. If you wish to go any further with the matter, I suggest you could write to Dr. Ayensu (address below) to approach him more formally about the idea. I look forward to hearing your views.

With kindest regards,

Yours sincerely,


D.G. Coursey

Copies to:-

Dr. E.S. Ayensu
Chairman of Botany
Smithsonian Institution
Washington D.C. 20560
U.S.A.

✓ Prof. D.J. Rogers
Taximetrics Laboratory
Armory 101
University of Colorado
Boulder
Colorado 80302
U.S.A.

Taximetrics Lab.

March 30, 1971

Dr. James Peters
Division of Reptiles and Amphibians
United States National Museum
Smithsonian Institution
Washington, D.C. 20560

Dear Dr. Peters:

Thank you for your letter of March 23. As you will no doubt know, I have taken the liberty to send you a statement we prepared on TAXIR two days ago. I agree with you that we should have a full description of TAXIR, and we are in the process of preparing a handbook of some 50 pages. In the meantime, I hope that we can survive, using something like the description which we have provided.

It is difficult to cover the waterfront for many different levels of requirements. I am certain that the wide range of taxonomists will find difficulty in reading our description and understanding what has been said. But I am equally certain that a computer programmer would not be fully enlightened by the same description.

Estabrook's comments are accurate, but unless they are placed in the right context, they will be as poorly understood as anything else. I believe he would agree with the statement we have made, but perhaps not in exactly the same words. Incidentally, I sent a carbon of our description to Harriet Meadow, asking her to review our statement to discover if there were something in it which is either not true or poorly described.

We have ourselves discovered one poorly worded parenthetical statement, and would like to correct it. I am including the change with this letter.

Thanks for your attention. Hopefully, all of us will sometime reach a sufficient level of understanding to be able to read and talk about computers as well as we can about taxonomy!

Sincerely,

David J. Rogers
Professor of Biology

Encl.

SMITHSONIAN INSTITUTION
UNITED STATES NATIONAL MUSEUM
WASHINGTON, D.C. 20560

March 23, 1971

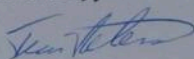
Dr. David J. Rogers
Taximetrics Lab.
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

I have your letter of March 8, and several copies of others that have been sent or received locally since the last MUDPIE (or should I say the most recent) appeared. I am sorry that you found the piece about TAXIR so bad from your point of view. If I recall correctly, you responded in a similar manner some years ago when I returned from having had an opportunity to see the system run, and wrote what I had seen and heard while in Boulder. Unless I am in error, Harriet Meadow also travelled to Boulder, and discussed the system at some length with your group. Her note includes what she found out, and what she concluded. The trouble seems to be coming from the fact that the only people writing about your system are outsiders who have a brief exposure to it and who do not come away from that exposure with as firm a grasp of the system as you would wish. Perhaps the time is ripe for you to put into print a summary of the system that says just what you think it ought to, so anyone interested could have your analysis available for detailed study and careful digestion. TAXIR was just one of many information systems Mrs. Meadow was charged with reviewing, and I am sure she felt the lack of published documentation as strongly as anyone could when she worked on the TAXIR statement.

In any event, I have had a point by point analysis of her note on TAXIR from George Estabrook, and enclose herewith a copy of it. If you will look it over, and tell me where George might also have gone wrong, I'd be glad to include this in a later MUDPIE.

Cordially,



James A. Peters
Curator
Division of Reptiles and
Amphibians

Enclosure

cc: Dr. Richard Cowan
Mrs. Harriet Meadow
Dr. George Estabrook

(The following comments were made by G. Estabrook concerning the recent statement about TAXIR in MUDPIE, and will probably appear, after editing, in a future issue. Your comments are welcome).

- 1) A catalogue of any subset whatsoever of the data base is available.
- 2) TAXIR is designed to handle data bases which grow and change frequently because no data records are needed for its creation and data (either new data or corrections and updates of old data) may be presented at any time afterward.
- 3) In addition, the size of the data matrix is limited only by the size of peripheral storage (e.g., tapes, data cell, etc.) and, in those cases where the data base is too large to fit in core, queries can be processed from external data base constructs. It is a very large data base that does not fit in core.
- 4) The input data structure is flexible and input format is described at each time of inputting data, for the block of data to be input at that time. It is never necessary to predefine the "values a field can take," as these are defined, after the fact, as the "values the field has actually taken" in the input data. Of course, if the user wished to predefine his set of permissible values he may-in this way TAXIR will generate data input diagnostics.
- 5) The data base is a "highly compacted matrix"-it is not augmented by it but replaced by it. It need not be entirely core resident at time of execution.
- 6) To the extent that the user wishes to limit the range of possibilities for his input data, TAXIR provides automatic diagnostics.
- 7) File maintenance is straight-forward and entirely under user control through the English-like TAXIR language.
- 8) I do not recognize the form 'data item=data value'--
I do not know what parentheticals are.

Taximetrics Lab.

March 26, 1971

Dear Miss Meadow:

I enclose a copy of a description of TAXIR. The original has been sent to Dr. Peters. If you find it completely out of question, please let me know, and we can work over any disputed area. Thanks for your prompt reply to my complaint.

Sincerely,

David J. Rogers
Professor of Biology

Taximetrix Lab.

March 26, 1971

Dear Dr. Peters:

Enclosed is a reworked description of TAXIE which I trust you will put in a forthcoming issue of Mudpie. It follows the same general format as that used by Harriet Meadow, and will fill but one page.

Thank you for your cooperation.

Sincerely,

David J. Rogers
Professor of Biology

Taximetrics Lab.

March 8, 1971

Dr. Richard W. Gowan, Director
Natural History Museum
Smithsonian Institution
Washington, D.C. 20560

Dear Dick:

Please refer to the enclosed carbons of letters to Miss Meadow and Dr. Peters. I am disturbed to think that this sort of thing is happening, and hope that it can be straightened out. I had no idea that TAXIR was to be so described, nor given the opportunity to check the work for accuracy before it appeared. All that she needed to do was send me a draft of her synopsis, and ask me for comments. This she did not do, nor did Dr. Peters who circulates Mudpie.

This matter is disturbing for more than just the inaccuracies about TAXIR, but leads me to question the competence of Miss Meadow with respect to other work that you have assigned to her, such as the survey of resources and requirements for computer processing of specimen data, for the Directors of Systematic Collections. If the level of competence to judge other systems is equivalent to that which she made of TAXIR, then you can have no confidence in her report.

But even more important is the Flora North American Project itself. If the computing systems chosen for this complex job are to be under the direction of Miss Meadow, then the whole project is doomed. I am sorry to be so pessimistic about this, but if her method of operating is typified by her lack of knowledge of TAXIR, and her willingness to make the errors in print, then I, for one, will withdraw any support so far given to the ideas, and actively oppose the present staff of FNA.

Sincerely yours,

David J. Rogers
Professor of Biology

Encl.: cc. of letters to Miss Meadow and Dr. Peters.

Taximetrics Lab.

March 8, 1971

Dr. Richard ~~W~~owan, Director
Natural History Museum
Smithsonian Institution
Washington, D.C. 20560

Dear Dick:

Please refer to the enclosed carbons of letters to Miss Meadow and Dr. Peters. I am disturbed to think that this sort of thing is happening, and hope that it can be straightened out. I had no idea that TAXIR was to be so described, nor given the opportunity to check the work for accuracy before it appeared. All that she needed to do was send me a draft of her synopsis, and ask me for comments. This she did not do, nor did Dr. Peters who circulates Mudpie.

This matter is disturbing for more than just the inaccuracies about TAXIR, but leads me to question the competence of Miss Meadow with respect to other work that you have assigned to her, such as the survey of resources and requirements for computer processing of specimen data, for the Directors of Systematic Collections. If the level of competence to judge other systems is equivalent to that which she made of TAXIR, then you can have no confidence in her report.

But even more important is the Flora North American Project itself. If the computing systems chosen for this complex job are to be under the direction of Miss Meadow, then the whole project is doomed. I am sorry to be so pessimistic about this, but if her method of operating is typified by her lack of knowledge of TAXIR, and her willingness to make the errors in print, then I, for one, will withdraw any support so far given to the ideas, and actively oppose the present staff of FNA.

Sincerely yours,

David J. Rogers
Professor of Biology

Encl.: cc. of letters to Miss Meadow and Dr. Peters.

Taximetrics Lab.

March 8, 1971

Miss Harriet R. Meadow
Flora North American Project
Smithsonian Institution
Washington, D.C. 20560

Dear Miss Meadow:

I have just seen "mudpie #11" which contains your description of TAXIR. You have made several errors in fact and interpretation about TAXIR, which must be corrected. I do not understand why you did not let me have the opportunity to review your paper before submitting it to Dr. Peters, because we could have given you the benefit of our knowledge of the system.

Under the circumstances I request that a complete new statement be made in Mudpie, so that the readers will not be mislead, and that we have the opportunity to examine the statement to be assured that it is correct.

Sincerely yours,

David J. Rogers
Professor of Biology

cc.: Dr. Richard Cowan
Dr. James Peters
Mr. Stan Shetler

Taximetrics Lab.

March 8, 1971

Miss Harriet R. Meadow
Flora North American Project
Smithsonian Institution
Washington, D.C. 20560

Dear Miss Meadow:

I have just seen "mudpie #11" which contains your description of TAXIR. You have made several errors in fact and interpretation about TAXIR, which must be corrected. I do not understand why you did not let me have the opportunity to review your paper before submitting it to Dr. Peters, because we could have given you the benefit of our knowledge of the system.

Under the circumstances I request that a complete new statement be made in Mudpie, so that the readers will not be mislead, and that we have the opportunity to examine the statement to be assured that it is correct.

Sincerely yours,

David J. Rogers
Professor of Biology

cc.: Dr. Richard Cowan
Dr. James Peters
Mr. Stan Shetler

Taximetrics Lab.

March 8, 1971

Dr. James Peters
Division of Reptiles and Amphibians
National Museum of Natural History
Smithsonian Institution
Washington, D.C. 20560

Dear Dr. Peters:

On p. 3, Mudpie 14, Harriet Meadow gave a synopsis of TAXIR. The synopsis contains several factual and interpretive errors, which should not be allowed to pass. I was not given the opportunity to review the statement before it was submitted to you. This is particularly disturbing since it is misleading to your readers, and misrepresents TAXIR.

I recognize the service which you are providing through the pages of Mudpie, but I am disturbed by the fact that no review process occurs to be certain of accuracy. I enclose a copy of my letter to Miss Meadow, which asks that TAXIR be given a complete new description in the pages of Mudpie.

Sincerely yours,

David J. Rogers
Professor of Biology

cc.: Dr. Richard Cowan
Miss Harriet Meadow
Mr. Stan Shetler

Taximetrics Lab.

March 8, 1971

Dr. James Peters
Division of Reptiles and Amphibians
National Museum of Natural History
Smithsonian Institution
Washington, D.C. 20560

Dear Dr. Peters:

On p. 3, Mudpie 11, Harriet Meadow gave a synopsis of TAXIR. The synopsis contains several factual and interpretive errors, which should not be allowed to pass. I was not given the opportunity to review the statement before it was submitted to you. This is particularly disturbing since it is misleading to your readers, and misrepresents TAXIR.

I recognize the service which you are providing through the pages of Mudpie, but I am disturbed by the fact that no review process occurs to be certain of accuracy. I enclose a copy of my letter to Miss Meadow, which asks that TAXIR be given a complete new description in the pages of Mudpie.

Sincerely yours,

David J. Rogers
Professor of Biology

cc.: Dr. Richard Cowan
Miss Harriet Meadow
Mr. Stan Shetler

FLORA NORTH AMERICA

Secretariat: Smithsonian Institution/Washington, D.C. 20560/Telephone (202) 381-5801



Editorial Committee

Peter H. Raven, Chairman
and Program Director
Smithsonian Institution

Stanwyn G. Shellen, Secretary
Smithsonian Institution

John H. Beaman
Michigan State University

Walter H. Lewis
Massachusetts Botanical Garden
and Washington University

John T. Mickel
New York Botanical Garden

Roy L. Taylor
University British Columbia

John H. Thomas
Stanford University and
California Academy of Sciences

Systems Development Manager
Harriet R. Meadow

10 March 1971

Dr. David J. Rogers
Professor of Biology
Taximetris Laboratory
University of Colorado
Boulder, CO 80302

Dear Dr. Rogers:

I regret that you found discrepancies in the description of TAXIR carried in MUDPIE #14. The write-up was written shortly after I visited you in January 1970 as an internal working document.

I would certainly appreciate information from you clarifying those points which you feel are incorrect or represent misinterpretations.

Sincerely yours,

Harriet R. Meadow

HRM/lme



UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE
CROPS RESEARCH DIVISION
NORTH CENTRAL REGIONAL PLANT INTRODUCTION STATION
AMES, IOWA 50010

February 17, 1971

Dr. David J. Rogers
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dave:

We are presently putting seed list information on IBM cards and using them to obtain print-outs of our seed list. These print-outs are then printed and distributed.

I am interested in having a look at the TAXIR system and to see if it is compatible with our system. We received considerable information recently from Sam Dietz which we have not yet had an opportunity to examine. I am unable to say at this time whether or not we will change over, but all I can say is that we are interested in it and are taking a look. If we do change to your system, we will get word to you either direct or through Sam Dietz.

Sincerely yours,



Willis H. Skrdla
Coordinator
Regional Project NC-7

WHS:vh

DONALD G. BROTZMAN
2ND DISTRICT, COLORADO

COMMITTEES:
INTERSTATE AND FOREIGN
COMMERCE
SUBCOMMITTEE ON COMMUNICATIONS
AND POWER

Congress of the United States
House of Representatives
Washington, D.C. 20515

WASHINGTON OFFICE:
413 CANNON BUILDING
WASHINGTON, D.C. 20515
TELEPHONE: (202) 225-2161

DISTRICT OFFICE:
BUILDING 40, ROOM 202
DENVER FEDERAL CENTER
DENVER, COLORADO 80225
TELEPHONE: (303) 233-3611
EX. 6709

January 13, 1971

David J. Rogers
Professor of Biology
Department of Biology
University of Colorado
Boulder, Colorado 80302

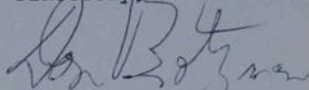
Dear Mr. Rogers:

Thank you for your recent letter.

In order to be of assistance to you, I consulted the Catalogue of Federal Domestic Assistance which reports to contain a catagorical listing of all Federal funding programs. After reviewing this I was not able to find any programs covering informational data retrieval other than those under the National Science Foundation, of which you are already aware. Therefore, it would appear the answer lies in the budget of the National Science Foundation.

I appreciate your taking the time to apprise me of this situation. An understanding of your problem will help me when the National Science Foundation's budget comes before the 92nd Congress.

Sincerely,



Donald G. Brotzman
Member of Congress

DGB/vb

Dr. A. Ashri

THE HEBREW UNIVERSITY OF JERUSALEM **האוניברסיטה העברית בירושלים**
FACULTY OF AGRICULTURE . REMOVOT ISRAEL . P.O.B.: 12 . TELEPHONE: 951981-4



DATE: Nov. 10, 1970
OUR REF:

Dear Dr. Rogers:

I was delighted to have your letter of Sept. 10, although it came belatedly. It arrived by regular mail instead of by air. This happened to several of my colleagues recently and it may ^{have} had something to do with plane security problems.

Your offer to help is very generous. I wish I knew before of your work and your programs. It would have saved us a lot of trials and errors. Your offer to make the programs available to us is very kind, although for some of the tasks we have evolved or modified programs already.

I am writing to you by hand so as not to delay my answer. In a couple of days I shall fly to Buenos Aires, through the courtesy of the IAEA in Vienna to participate in a Study Group meeting on Mutations and Plant Improvement. Dr. Konzak will be there and I shall be able to learn from him more about the programs

and their utilization. Unfortunately it will
be impossible for me to go to Colorado now.
We are at mid-semester and I shall have
to rush back - a pity. Still, with the
information I'll get from Dr. Kozak and the
literature which you sent I shall be able
to see where we can cooperate on this
topic. I would like to do this work together.

Thank you again for telling me about
your work and your offer to help. I shall
write to you again in December.

Best wishes

Sincerely yours

D. D. Shi

TAXIMETRICS LAB.
ARMORY 101

September 10, 1970

Dr. Amram Ashri
Faculty of Agriculture
Rehovot, Israel

Dear Dr. Ashri:

Dr. Konzak and his colleague Dr. Dietz were kind enough to forward to me a letter you had written on August 23. I note with interest your desire to use various computing machine programs for information retrieval and numerical taxonomy, and further, that you have a CDC 6400 computer at your disposal.

Perhaps Dr. Konzak mentioned that he and Dr. Dietz were using our program, TAXIR, with good results there in Pullman. He may have told you that the original programs were written for the CDC 6400, and that one of their problems had been the difficult task of transferring them to their machine (an IBM 360/65). But since you are using the same computing facilities that I have, perhaps you would be interested in the use of our programs. We have programs in information retrieval (TAXIR), and two other powerful programs for data correlation and clustering, all running on the CDC 6400. They are all designed for the specific tasks you mentioned in your letter.

We attempted earlier to work with Dr. Knowles on safflower, but for several small reasons, we never did get a cooperative program going. If you are interested, we would be pleased to have one of your assistants or yourself come here to learn to use our programs, and have him take them back to Israel with you or him.

Unfortunately we do not have any effective descriptive literature on the programs, but I am sending a few which may indicate how the two programs in data correlation and clustering work. These are being sent separately.

I look forward to hearing from you.

Sincerely yours,

David J. Rogers
Professor of Biology

*Dave, Cal gave me a copy of this & I thought you might be interested.
 Dan Hudson recently visited Beltsville & showed them the map drawn by computer
 of the sites where were collected in Turkey. Done by coordinates. Dr. Gershen
 was very pleased. Our Technical Committee meeting is next. 29 x 30
 & we will have a TAXIR demonstration then.*

החיברות הטכני של אוניברסיטת בירושלים
 THE HEBREW UNIVERSITY OF JERUSALEM
 FACULTY OF AGRICULTURE · REHOVOT, ISRAEL · P.O. BOX 12 · TEL. 951981-4

Sam Ditz

DATE August 23, 1970

RECEIVED

AUG 31 1970

Prof. C. F. Konzak
 Department of Agronomy
 Washington State University
 Pullman, Washington 99163
 U. S. A.

By _____

Dear Prof. Konzak:

Thank you for your letter of July 30. I have followed your publications on documentation and heard about your wheat program from Dan Ezhary. I share your feelings on the need for large germ plasm banks and for standardized adequate records that will be of general benefit. The loss of information, even in small countries like ours is great and it is two-fold: First, simply loss of records, lines etc., and second, lost opportunities. With a whole collection grown once for a given purpose, say screening for resistance to a disease, with very little extra effort and marginal cost increments, it is possible to collect much more data. This is rarely done. Standardized data collection procedures which will make such work easier will help in this regard too.

First some background would be in order. The project grew out of a USDA P.L. 480 program on safflower which I conducted from 1966 to 1970. Very briefly, the goals were: 1) To evaluate the world collection of safflower for resistance to the safflower fly (Acanthophilus helianthi R.) and several diseases which are endemic in Israel and are not present in the USA or represented by different races (e.g. rust). 2) To identify geographical centers of various desirable traits. 3) To discern and define regional types. 4) To evaluate evolutionary divergence within the cultivated species. 5) To evaluate and study the wild Carthamus species. The world collection contains now about 2,000 accessions from 36 countries and much of it was collected by Dr. P.F. Knowles (U.C. Davis) in extensive trips in 1958 and 1965.

It was clear to me that without resorting to the computer I will not be able to do much more than the first goal listed above. I started exploring various possibilities and your publications and reports (IAEA, FAO) were very helpful. In 1968 I spent 3 weeks in the USA in connection with the safflower

research and discussed the plans with all leading safflower investigators. (I wanted to go to Pullman too but could not fit it ~~into~~ into my itinerary).

At the time I made a draft proposal and envisioned a "live" file which will be a permanent and expanding repository of information on safflower lines, wild species, etc. (see enclosed summary). This file would be used for information retrieval, analysis, sorting, etc. It was planned with 3 major sections as shown in the enclosure. Primarily because of problems with section II.B. I concluded that the plan was too far-reaching at this stage. We had no way of making intelligent estimates on format dimensions etc. Such a program and its maintenance would have also been very expensive and special financing would have been required. This was another factor in deciding that we were not ready yet for a comprehensive computer setup.

The important need was to evaluate the world collection. Since we have relatively few commercial varieties in the world, their evaluation could be handled separately at a later date. While in the USA I secured the cooperation of Dr. Kmwles who had data on hundreds of accessions from both Davis and Pullman (yes!) and of Dr. D.E. Zimmer and Mr. L. Urie of Logan, Utah, who had data on about 1,000 lines from their tests. Thus, we have 4 sets of data—the above 3 plus Israel. I then planned it so that it will be possible to analyze the data from the locations separately or jointly. We may add a fifth soon—the collection is now tested near Bombay in India and they are interested in cooperating. From Israel I have data on some 25 traits for each line, from the other locations we have less entries on each line. If a "live" file is undertaken later on all this information will be in a readily transferable form.

The information from each location was recorded using codes for alternatives (e.g., 1 = white flower, 2 = pale yellow... etc.) and giving in full all metric data. For statistical analysis we prefer not to use "0" as code but start with 1. Each line in every location has one IBM card with the identification data and other entries and we transferred all data to magnetic tape for convenience. I am aided by a programmer who is on our Faculty staff and we are using a CDC 6400 computer of the Hebrew University. Much of the analysis is still ahead of us and more ideas arise as results are obtained. We have used Israel's data as pilot so far. We calculated means, SD, SE and CV for the different traits by the countries of origin and other categories; contingency tables for traits which do not follow normal distribution, correlations and regressions (including stepwise and multiple regressions), range simultaneous

Taxir written
for CDC 6400 so can
use program
now.

THE

FOR

tests for regions, and various other findings, e.g. taking the lower and upper 5% of each trait (or the immune vs. very susceptible) and comparing the means for all other traits between these groups. For the contingency tables we use the BMD program (U.C. Press, 1967, p. 341-356); for regression our programmer has a locally modified program (Rehovot 1970 version), for the range test we use the Gabriel method (see Sokal & Rohlf, 1969, Biometry, Freeman & Co. 235-246, 685-687). All other programs, subroutines etc. were prepared by our programmer following discussion. We are now at the stage where we are beginning to analyze jointly data from different locations, e.g., correlation of oil % in Davis and Logan, predictability of growing season length in Logan from performance in Israel, and so on. Of course, we use the computer to print assorted classified lists, and this is also a very big help.

We had very little past experience to draw upon and are making progress by trial and error. Before long I hope to tackle the problem of regional types using discriminant functions and also to try some of Sokal's "Numerical Taxonomy". If you can help with ideas and suggestions on any aspect, it will be a great help. If you are interested in certain aspects and want more details, please let me know. The system for safflower can be adapted to other crops with minor modifications. I shall be glad to contribute from our experience and to be a bit of a "missionary" in this subject.

Before closing I would like also to say that I am going on with mutagenesis work in peanuts. Actually it is expanding now. I received a 3-year grant to investigate natural divergence for plasmon types affecting growth habit in peanuts and for the induction of plasmon mutations. We are planning to treat mature seeds, developing embryos in their pods while still attached to the plant and embryos in culture. So far we settled on EMS, Ethidium Bromide, Chloramphenicol, Erythromycin, Acriflavine and 5-BU, and perhaps the ICR chemicals. If you can help with ideas on pre- or post-treatments, mutagen suggestions, etc., I shall be very grateful. At a later date I shall write more on this aspect of my research.

Looking forward to hearing from you soon

Dr. Anram Ashri
Senior Lecturer in Genetics
and Plant Breeding

P.S. Please give my warm regards to Dr. Nyman. I am sending you some reprints separately; safflower ones are still in preparation.

Copies: Dr. E.N. Brooks, USDA, Beltsville; Dr. L. Zimmerman, USDA, Tucson; Dr. P.F. Knowles, U.C., Davis, Calif.; Dr. D.E. Zimmer, USDA, Fargo, N.D.; Dr. A.L. Urie, USDA, Shafter, Calif.

APPENDIX 1. CRITERIA FOR A STANDARDIZED CLASSIFICATION OF
TRAITS IN SAFFLOWER, Carthamus L.

A draft proposal prepared for discussions
and consultations with scientists in the
U.S.A. in 1968.

The availability of a very large world collection of
safflower, and its large scale testing in the U.S.A. and
other countries make a standardized classification system
most desirable. The advantages of such a system are very
obvious and need not be listed here.

In preparing this draft I had several considerations:

1. That it should be suitable for computer work.
2. That it conform as much as possible to other
standardization schemes now in development
(e.g. FAO/IAEA).
3. That it will fit both cultivated entries and wild
species.
4. That it can be used both as a repository with rapid
information retrieval and as an analytical tool.
5. That the traits selected will be important either
economically or diagnostically.

The draft is divided into 3 sections:

- I. Identification of the material - this will be
recorded once and will need no revision or updating.
- II. Variety description and properties
 - A. Traits unaffected by environment. The over 50
traits assembled will be useful in describing
varieties or distinguishing between species.
No doubt other traits can be added.

One problem is the inclusion of information on the
inheritance of various characters: Should it be included as
another entry with each trait, and then have a uniform code
for one gene, 2 genes, complementary, additive, etc.? Or,
should it be given as another entry (it may also be entries)

but giving more details such as who found, in which cross, what was found? Or, have a separate "chapter" on inheritance and in it give the information. I think that the second alternative may be the most suitable. The problem will be of allotting format dimensions and careful consideration should be given it.

Similarly, physiological response data may pose a problem. Should information on response of specific traits to external applications of hormones (e.g. Gibberellic acid A_3) be included and if so how? Here also multiple entries may be expected. The alternatives are much the same as above and should be weighed carefully.

How active this section will be (in the sense of adding, revising, etc.) will depend on how the above problems are resolved.

II.B. Traits affected by the environment:

Actually this section will contain all or most of the economically important traits. A partial list, rather general, is given here. It should be envisioned that this section will be very active, with many multiple entries. The planning of format dimensions here should be very careful.

I. Identification of Materials

These should include:

- 1) International Accession Number; 2) Country and State (or province) of origin; 3) Species; 4) Subspecies; 5) Variety Name; 6) Pedigree; 7) Comments of source or collector on purity; 8) Nature of Material.

II. Variety Description

It is deemed desirable to separate the traits into two categories:

A. Those not affected by environment, e.g. color of spines, fresh color of corolla. In this file new information should be recorded as it is received, but it will not be necessary to identify the source since it will be irrelevant. Also, here there will be a single entry per trait.

B. Those affected by environment, e.g. days to flowering, yield, oil contents. Information in this category should be identified by source, location and year. Duplicate data and even contradictory data from different sources should be recorded.

It is suggested that all traits not clearly belonging to group A be included in group B.

Traits in group A

1) Chromosome number; 2) duration of life; 3) growth habit. 4-6) Branches: location, angle, length; 7) stem lodging; 8) stem color; 9) pubescence; 10-11) Cotyledons: shape, midvein color; 12-13) Lower stem leaves: shape, margin; 14-16) Upper stem leaves: shape, margin, color; 17-18) Primary head: shape, closure; 19-25) Outer involucre bracts: shape, cross section, length, No. of veins, location of spines, attitude, color; 26-28) Middle involucre bracts: tip, shape, tip color; 29-30) Inner involucre bracts: tip color, tip margin; 31-34) Spines: number, length, strength, color; 35) Spininess, score; 36) Corolla color - (following Knowles' 1958 classification); 37) Style, color; 38) Anthers, color; 39) Pollen, color; 40-43) Achenes: shape, color of crenulate (wide) end, color of hilum (pointed) end, surface texture; 44) Hull thickness; 45-48) Pappus: presence, length, longest scales, color; 49-51) Reaction to Puccinia carthami: age, race, relative resistance of host plants; 52)... and on: Entries on reaction to other diseases: minor diseases should be included too. Entries on other traits (isozymes, fatty acids, etc.).

Traits in Group BTraits which are affected by environment

Many of these traits will vary with environment. Therefore multiple entries should be envisioned and every entry should be accompanied by data on source, location, planting time, soil fertility level, soil moisture, damages from various pests, etc. Here much thought on the format will be required. Following is a list of traits which belong in this group.

1. Yield (specify moisture %?, Exp. Sta. or commercial?)
2. Oil contents of whole seeds (specify moisture %, oven dry basis or air dry)
3. % hull
4. Proportion of various fatty acids (should perhaps be included in Group A?)
5. Iodine No. (to Group A?)
6. Other properties of oil? (e.g. color, viscosity, etc.)
7. Days to first flower
8. Height of plants
9. Branches -- few, normal or profuse?
10. Mean diameter of primary head
11. Mean No. of seeds in primary head
12. Achenes -- weight/1000
13. Salinity tolerance
14. No. of heads/plant.

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE

REGIONAL PLANT INTRODUCTION STATION
WASHINGTON STATE UNIVERSITY
JOHNSON HALL 59
PULLMAN, WASHINGTON 99163

December 11, 1969

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dave:

Plant Introduction recently received a letter from Dr. Fred A. Bliss of the University of Wisconsin requesting your name and address as the developer of the computer program we are using for our bean inventory. I'm sure that he will be checking with you later.

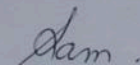
Write

We also just had a visit from Mr. Gary V. Richardson who is the Biometrician in Charge of the Biometrical Services Staff for the ARS Statistical Laboratory newly set up at Ft. Collins. He presently is attempting to set up some sort of data retrieval system for the National Seed Storage Lab. with Ed James there at Ft. Collins. Konzak, Lee Hudson, and I all informed him of the TAXIR system and suggested that he see you at Boulder. I'm sure Gary will also be checking in soon. If we can get Gary interested in using your program I'm sure that the National Seed Storage Facility will fall in line.

I am also attempting at this early date to set up a National Meeting of the four Coordinators and Crops personnel together with our W-6 Technical Committee next September. One of our main purposes is to put on a demonstration of what the TAXIR system is doing for us at the present time. Although this is a long ways away I wanted you to know that we are still plugging at salesmanship.

Hope you have a happy holiday season and a prosperous new year.

Sincerely,


S. M. Dietz
Coordinator

ARIZONA STATE
UNIVERSITY

DEPARTMENT OF ZOOLOGY

TEMPE, ARIZONA 85281

9 July 1969

Dr. David J. Rogers & Associates
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

Behold — a letter of Praise from an elated Chaldean (deep down I feel computers are Black Magic).

At this sitting I am on the last lap of writing thesis, having surmounted (with assistance) many meddlesome obstacles. In short, the computer speaks beautiful things to me, and machine is man's friend. Not that I had any doubts, but the forces of derision were gathering on all sides. However, truth and beauty (= computer taxonomy) prevailed, and I see a daily burgeoning of converts. It was helpful that the members of my committee were among the front runners.

I am most pleased with the results and have you and George and Bob and others to thank. The concerted effort has resulted in a vastly more cohesive thesis. When it is finished, I'll try to get a copy up to you; we're xeroxing it on Minckley's grant and I'll shoot for some extras.

In the meantime, I send grateful thanks to you and yours for opening the way and sharing.

Most sincerely,

Glady
Glady L. Lytle

UNIVERSITY OF THE WEST INDIES

Department of Biological Sciences

CABLES: "STOMATA" PORT-OF-SPAIN

TELEPHONE: 662-5511



ZOOLOGY SECTION
ST. AUGUSTINE,
TRINIDAD, W.I.

25 July 1969

Prof. D. J. Rogers,
Dept. of Biology,
University of Colorado,
Boulder, Colo. 80302.

Ans. 8/1

Dear Dave,

I have been meaning to drop you a line ever since the Root Crop Conference in 1967 but tempus fugit. I am taking my triennial leave this summer and am hoping to spend most of it (about two months) in Canada. I would also like to return to Trinidad by way of the U.S. in order to look you up and renew the acquaintance. If you happen to have any tame mathematicians and/or computerniks kicking around I would also like to have a session with them, as I have some ideas that are beyond my capabilities to work out mathematically, also beyond the abilities of our computer.

I have made enquiries this end and find that I can get a flight from Seattle to Denver. This would be sometime around the middle of September. If you will be there at that time I can let you know more details later. I am leaving Trinidad on August 4th, after which my address will be as follows:-

1665 Hollywood Crescent,
Victoria, B.C.,
CANADA

Hoping to hear from you soon,

Yours sincerely,

Brian

Brian D. Ainscough

*P.S. Roger Barnes sends his regard.
B.*

17 June 1969

Dr. Nicholas Jardine
King's College Research Centre
Cambridge CB2 1ST
England

Dear Dr. Jardine:

Thank you very much for the invitation to your seminar in September, entitled "Mathematical and Theoretical Aspects of Numerical Taxonomy. Unfortunately the date of the seminar coincides with the beginning of our Fall Term, and I cannot personally attend.

May I, however, suggest as a substitute Mr. Frank Bisby, whose address after July 15 will be St. John's College, Oxford OX1 3JP. Mr. Bisby has one year to complete his doctoral work in Oxford under Mr. Frank White at the Forest Herbarium. He has for the past seven months been working at our laboratories to learn our methods and philosophy. I find him a very able, indeed brilliant student, and believe that he would contribute nicely to the seminar. In addition to his work with us, he has had a good experience in normal non-numerical methods in taxonomy and therefore brings a balance to discussions. I trust that you will see fit to offer him an invitation.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm
CC to F. Bisby

Frank: Are you interested?
KING'S COLLEGE RESEARCH CENTRE

KING'S COLLEGE CAMBRIDGE CB2 1ST ENGLAND
TELEPHONE CAMBRIDGE (0223) 50411

17 May 1969

Let me
know -
Dave.

Dear

Dr Rogers.

On 15 and 16 September 1969 there will be a seminar entitled *Mathematical and Theoretical Aspects of Numerical Taxonomy* in King's College Research Centre. Special emphasis will be placed on the relation between theoretical aspects of numerical taxonomy and the particular problems which arise in the practical application of the methods. The sessions of the seminar will consist of semi-formal or informal discussions, but it is also intended as an opportunity for those working in this subject to meet one another. The discussions will be chaired by Professor P.H.A. Sneath; and in order to retain as informal an atmosphere as possible, membership of the seminar will be limited to about twenty invited participants. We enclose a tentative programme.

Meals will be provided during the seminar for all those taking part, and accommodation will be provided for those not resident in Cambridge. If your sponsoring body cannot pay travelling expenses, we may be able to assist with expenses for travel within the United Kingdom; we regret that we cannot pay any travelling expenses other than this, but we hope that if it is necessary for you to travel from abroad, then this letter will enable you to obtain funds from your sponsoring body.

We very much hope that you will be able to come to this seminar, and look forward to hearing from you at an early date whether it will be possible for you to attend.

Yours sincerely,

Nicholas Jardine

Nicholas Jardine

Robin Sibson

Robin Sibson

St. Johns College
Oxford
OX1 3JP
England.

10 January 1969

Dr. Carl L. Withner
Department of Biology
Brooklyn College of CUNY
Brooklyn, New York 11210

Dear Dr. Withner:

Responding to yours of January 3, we are pleased to hear of your work and your desire to use our program on *Oncidium* flower data. I would like to propose that you follow an alternative route to satisfy your needs, in as much as the conversion from one machine to another is very complex, costly and time consuming. We would be happy to run your data on our machine here and send you the printout. To accomplish this we would ask you to prepare the data following the description given in the two enclosed papers. The mimeographed paper tells you the actual physical details of setting up the run, and the printed paper gives you details of the method for using the characters.

The cost on the computer use will probably be so small that we can absorb it ourselves. You will have the information back much more rapidly this way than if you have to bother with mounting the program on your own machine. I trust that this will be a satisfactory procedure for you and hope to hear from you soon.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Postcard request for clarification: Haemalia or Haemaria?

ARIZONA STATE
UNIVERSITY

TEMPE, ARIZONA 85281

DEPARTMENT OF ZOOLOGY

April 22, 1969

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101A
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

At last, satisfaction! This Spring proved to be far more traumatic than I had anticipated, and, unfortunately, I received no help from Judy and Sally in the preparation of my data for your computer. I found it best to condense the original 52,000+ pieces of data into 3200+ workable bits, the result is coming to you under separate cover. In their own time, no doubt, they will have the program up and running for ASU, but they cost me a June graduation.

I decided to use the mean values (as you suggested) for each collection, resulting in a total of 60 cards with 54 columns of characters, rather than 971 individuals with the same number of characters. In the final analysis, the former promises less obscure results. Also, this shortened my load so that I was faced with only the reworking of the qualitative data and pulling the means together and placing everything into character-states. I am no doubt at fault by sending the cards with no one to check them but myself, but I have reached the point of desperation.

O.K.! Questions

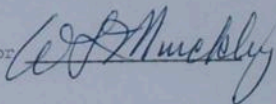
1. The "number of studies" cards puzzled me. I have no reference point for it; run all of the studies. Can you punch the card for me?

2. Are the character parameter cards correct? All of my characters are simple, therefore the parameters should be zeros, correct?

End of questions....

Minckley asks that the bill (not to exceed \$100.00) be charged to NSF-GB-6477X, Account # 5272-128, care of Dr. W. L. Minckley, Department of Zoology, Arizona State University. He is countersigning this letter as a commitment of funds. Will our cost be within the \$100.00?

Approval for Commitment of Funds - Principal Investigator



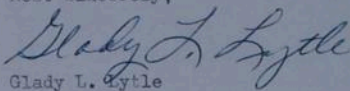
Dr. David J. Rogers
April 22, 1969
Page 2

And.... did you hear anything from Dr. Justus of this department about seminarng? I went to see him with your letter and he said that he had asked another member of the faculty, who knows you, to write a letter inviting you to speak to us (he didn't identify the person). However, his final statement was that the schedule was full this semester, but that he (Justus) would ask you to come next Fall and present a seminar. Did he or didn't he....? I suspect not.

I close for now. Give my regards to George and Bob Brill. I hope none of this is inconveniencing you or them. I truly appreciate all of the help you have given me, I hope the final thesis reflects intelligent decisions. Minckley and I have had a paper accepted by the Washington (D.C.) Academy of Science, it will probably be out this Fall. I'll send you a reprint of it ASAP. Part of the data in it was utilized in my thesis. Thesis will be published and reprints of it also sent to you. Also, you will be acknowledged in it. Hardly large enough praise.

I go. Thank you again. I light candles that all is well. Write and/or return it, if not well.

Most sincerely,


Gladys L. Lytle

P.S. There are duplicates of the data cards here, in case of mishap.

TAXIMETRICS LABORATORY
DEPARTMENT OF BIOLOGY
APMORY 101
UNIVERSITY OF COLORADO
BOULDER, COLORADO 80302

To: Gladys Lytle

Computing Expenses, May 1 - 29, 1969

CPU time, 49 seconds at \$5/minute	\$4.20
PPU time, 107 seconds at \$1/minute	1.80
Paper and printing costs	3.00
Skyline Graph	10.00
Operating and Consulting	25.00
	\$44.00—



*paid -
check arrived 16 June 1969
forwarded to IRS
to telephone acct*

29 May 1969

Miss Gladys Lytle
Department of Zoology
Arizona State University
Tempe, Arizona 85281

Dear Miss Lytle:

At last! Your deck has been run, printout printed and Skyline graphed. They are being sent off to you in various packages.

Expenses run to \$44.00, itemized on the attached sheet.

We hope that your research progresses successfully to a thesis.

Sincerely,

John W. Wong, Jr.
Programmer

JWW:gm

To: Gladys Lytle

Computing Expenses, May 1 - 29, 1969

CPU time, 49 seconds at \$5/minute	\$4.20
PPU time, 107 seconds at \$1/minute	1.80
Paper and printing costs	3.00
Skyline Graph	10.00
Operating and Consulting	25.00
	\$44.00

20 January 1968⁹

Miss Gladys L. Lytton
Department of Zoology
Arizona State University
Tempe, Arizona 85281

Dear Gladys:

Thanks for your letter. I suppose the only thing that needs an answer immediately is that I have not yet heard from Dr. Justus. I hope he does not want me for the week of March 17. I would prefer the week of the 3rd.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

ARIZONA STATE
UNIVERSITY

TEMPE, ARIZONA 85281

DEPARTMENT OF ZOOLOGY

January 14, 1969

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101A
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

At long last news from the south country. The flu has taken its toll here, with even the most able-bodied of us falling before it. However, a semblance of normalcy has been established and I have returned to the timetable set up before the flu hit.

It has been a fairly frantic month and a half here; in the midst of which Mother and I find ourselves moving to a new domicile in lieu of the landlady's going commercial and selling her properties.

We most gratefully received the cards for the cluster program about two weeks ago and Judy and Sally spent the holidays going over them. They are busy working on a sub-routine that will convert my already punched data cards into input cards appropriate for Charanal.

Minckley and I finished our first paper this week, and within the next two weeks I should have data up to you. By hand I went through my data and pulled ranges (a manipulation we hadn't included in our previous programming) and by fortune, found mistakes that probably would have gone unnoticed. Barring any further unforeseen inconveniences, work should progress rather rapidly here on out.

The most important piece of news I have is that we have \$100.00 to spend on computer time, etc. So let me know what the billing procedure is and I can get on it from this end. I hope that will be sufficient to handle our needs.

Also, have you received any correspondence from Dr. J. T. Justus about giving a seminar here in March? I saw him the first day I was back and put him onto your offer.

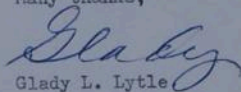
And finally, the Bureau of Commercial Fisheries is going to support research by Minckley on telemetry collection of aquatic environmental data. Right now he is interested in primary data retrieval systems and how they function. I know George has developed a system for this and Minckley would appreciate some literature references to build from.

Dr. David J. Rogers
January 14, 1969
Page 2

I hope we're not imposing by asking George for assistance on this subject.
(Thank you George....)

Judy, Sally and I send you and company our warmest regards. We very much appreciated your help and will have the data together and runnable without snags. And, we will see you in March, if you are coming.

Many thanks,



Gladys L. Lytle

JAN 9 1969

DAVE!
DEL PROMISES TO
HAVE THE CHROMATOGRAPHIC
DATA ANY MINUTE - BUT
WE'VE HEARD THAT BEFORE!
- Frank

Botany Dept
BRITISH MUSEUM (NATURAL HISTORY)
Cromwell Road,
London, S.W.7
Telegrams: NATHISMUS, LONDON, S.W.7

2nd January 1969

Dr F. G. Hawksworth,
U. S. Forest Service,
Rocky Mountain Forest and Range Experiment Station,
Colorado State University,
Fort Collins,
Colorado 80521
U. S. A.

Dear Dr Hawksworth,

I feel I must let you know how excited I was to read your article in Taxon. Although I am not informed fully on information theory, I can see that this approach will make a great deal of difference to the understanding of "characters".

I have long felt that the analysis of the way in which characters as entities affect classifications is what we need to investigate. With best wishes for your continued research on this topic.

I am,

Yours sincerely,

John Lewis

P.S. I am abstracting your article for our
internal newsletter to scientists in all Depts.

Rocky Mountain Forest & Range Experiment Station
240 WEST PROSPECT STREET
FORT COLLINS, COLORADO 80521
OFFICIAL BUSINESS

Taximetrics Laboratory
Biology Department
Armory 101
University of Colorado
Boulder, Colo. 80302
10 January 1968⁷

Mr. John Lewis
Botany Department
British Museum (Natural History)
Cromwell Road
LONDON, S.W.7, England

Dear Mr. Lewis:

Dr. Frank Hawksworth forwarded to me a copy of your letter of 2 January with your kind thoughts on the character study we did on Arceuthobium. We thought you might like to have the companion paper published earlier which sets out the thought processes we employed to study characters. This paper gives more of the conceptual basis for the process. It is being mailed separately.

Sincerely yours,

David J. Rogers
Professor of Biology

DJR:gm

UNIVERSITY OF COLORADO

BOULDER, COLORADO 80302

January 23, 1969

PHONE: 303-443-2211

Ext. 6712

TAXIMETRICS LABORATORY
DEPARTMENT OF BIOLOGY
ARMORY 101

Dear

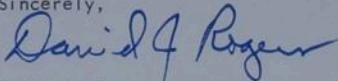
In a recent conversation with Dr. Edward Weiss of OSIS we came to the conclusion that many information retrieval systems developers in the country are in need of closer contact with other IR developers. We have concluded that there are probably many system designs which are more similar than different.

It was also concluded that it would be advantageous for these systems designers to become acquainted and to discuss their mutual concerns. I agreed to act as a focal point to discover whether you might be interested to join in some informal gathering with no intent towards publication of any sort of report, but merely with the desire to know "What's cooking?" Ed Weiss indicated that it would be very beneficial to him to sit with such an informal gathering and to trade shop talk.

I propose therefore that some informal meeting be arranged here in Boulder on the campus of the University of Colorado for some period of time (two days to a week). There is no proposal with any agency to sponsor this gathering; it would be a "pay your own way" affair.

To allow some more definite planning to be made I enclose a very informal questionnaire which I hope you will return to me. We'll keep you informed if you are interested.

Sincerely,



David J. Rogers
Professor of Biology
Project Director

Return To: Dr. David J. Rogers
Taximetrics Lab.
Armory 101
University of Colorado
Boulder, Colorado 80302

_____ I am (am not) interested in an informal
get-together in Boulder.

I can spare a week _____ 2 days _____ 3 days _____
for such a meeting.

I prefer: June 2-6 _____
June 9-13 _____
June 16-20 _____
June 23-27 _____

July 7-11 _____
July 14-18 _____
July 21-25 _____
July 28-Aug. 1 _____

Please Score
1--1st Choice
2--2nd Choice
3--Impossible

Your Name _____

Address _____

Prof. Allen Kent
Knowl. Avail. System Center
University of Pittsburgh
Pittsburgh, Penna. 15213

Dr. Edwin B. Parker
Inst. for Communication Res.
Cypress Hall
Stanford University
Stanford, Calif. 94305

Prof. Donald J. Hillman
Center for the Inf. Sciences
Lehigh University
Bethlehem
Penna. 18015

Dr. James W. Sweeney
Computer Center
University of Oklahoma
1610 Newton Drive
Norman, Okla. 73069

Prof. Peter H. Rohn
Department of Political Sci.
University of Washington
Seattle
Washington 98105

Dr. Atmaram Soni
Oklahoma State University
Stillwater
Okla. 74074

Mr. Kent K. Curtis, Head
Institutional Comp. Services
Office of Computing Activities
National Science Foundation
Washington, D. C. 20550

Dr. James L. Carmon
Director of Computer Center
University of Georgia
Athens
Georgia 30106

Mr. Eugene S. Schwartz
Computer Sciences
IIT Research Institute
10 West 35 St.
Chicago, Ill. 60616

Dr. David Lefkowitz
University of Pennsylvania
3625 Walnut Street
Philadelphia
Penna. 19104

Dr. Carlos M. Bowman, Dir.
Computation Research Lab.
1707 Building
Dow Chemical Co.
Midland, Mich. 48640

Dr. Peter Lykos
Illinois Institute of Tech.
Chicago
Ill. 60616

Mr. Thomas W. Quigley, Jr.
Chemical Information Unit
OSIS - NSF
Washington,
D. C. 20550

Mr. Robert C. Verderber
Special Projects Program
OSIS - NSF
Washington
D. C. 20550

Mr. Paul J. Wolfe
Computer Center
University of Iowa
Iowa City
Iowa 52240

Dr. Gilbert Daniels
Hunt Botanical Library
Carnegie-Mellon Univ.
Schenley Park
Pittsburgh, Penna. 15213

Dr. Richard Dougherty
Norlin Library
University of Colorado

Dr. Kenneth Zabriskie
BioSciences Info. Services
2100 Arch Street
Philadelphia
Penna. 19013

DEC 9 1968



JOINT FAO/IAEA DIVISION OF ATOMIC ENERGY
IN FOOD AND AGRICULTURE



INTERNATIONAL ATOMIC ENERGY AGENCY - FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

IAEA, KÄRNTNER RING 11, A-1010 VIENNA, AUSTRIA - TEL: 52 45 11 - TELEX: 01-2645 - CABLE: INATOM VIENNA

DAT/411-1

2 December 1968

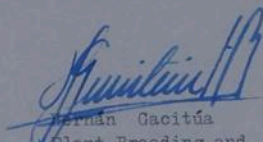
Dear Dr. Rogers,

Thank you for your letter of 5 November 1968 and for the description of TAXIR.

I appreciate very much your information and suggestions and I am writing to Dr. Arthur S. Boughey in connection with botanical keys.

I thank you also for adding my name to your mailing list.

Sincerely yours,


Germán Cacitúa
Plant Breeding and
Genetics Section

Dr. D. G. Rogers
Professor of Biology
University of Colorado
Boulder, Colorado 80302
U.S. A.

5 November 1968

Dr. Hernan Gacitua
Plant Breeding and Genetics Section
Joint FAO/IAEA Division
Karntner Ring 11
A-1010 Vienna, Austria

Dear Dr. Gacitua:

In response to your letter of October 11, I am enclosing a short description of TAXIR, our information retrieval system. This description does not, of course, go into any detail and we are in the process at the moment of putting together a User's Manual for the system.

In your letter you also asked to have a copy of "the operating taxonomic keys for plants." I am not just sure what is intended by this, but let me say that the TAXIR system has a part of its computer program that we call a "Key Facility." The Key Facility allows an individual to make an identification in a simultaneous manner describing an unknown organism in terms of the descriptors (characters) used by the expert in describing the data bank. As far as there being a whole data bank of keys, none exists. In this connection you may wish to write to Dr. Arthur S. Boughey
Department of Population
Biology
University of California
Irving, California 92650

who is also working on botanical keys.

We have added your name to our mailing list for our publications as they come out.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

NOV 13 1968



CENTRAL STATE COLLEGE

Edmond, Oklahoma 73034

DIVISION OF SCIENCE
Biology

November 11, 1968

Dr. David J. Rogers
Taximetrics Lab
University of Colorado
Boulder, Colo.

Dear Dave:

I certainly appreciated all the many kindness you extended to us at the Second Annual Meeting of the Numerical Taxonomy. I hope next time to be able to participate more fully. Please express my thanks to the other members of your department.

The next time I visit Boulder, I wish to spend considerable time visiting with you and studying the Lepidoptera (especially the Papilionidae) in your museum.

I wish to bring up again the idea of having a Numerical Taxonomy workshop some August lasting from two to three weeks. The National Science Foundation seem interested in workshops so it should not be too difficult to obtain necessary funds to cover the expenses of the institute and the participants. I think it would be ideal to have it at your laboratory as you have the personnel, know how and facilities. I would love to spend several weeks studying at your laboratory and working out Numerical Taximetrics problems.

Sincerely,

A handwritten signature in dark ink, appearing to read "Kent", written over the typed name.

Kent H. Wilson
Department of Biology

KHW:jo

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

November 7, 1968

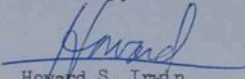
Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dave:

Killing two birds with one stone, I'm sending the family lists you asked for in your letter of October 29. The more important matter, it seems to me, is Brieger's letter, which, beyond his brief ruminations about your papers, is an implicit question about the possibility of his getting to the International Congress in Seattle. Is there going to be a session on numerical taxonomy/taximetrics, and if so, how about inviting him to be a participant? Although you may not subscribe to his views, I think a great deal is to be gained by including a thinker from Latin America in any taximetric session that may be in formulation. The paper he proposes to write on his return to Brazil might well be suitable.

Please keep me informed of any specific or general reply you may have to Brieger's letter. I will miss his Germanic hand at Brasilia; although he has a propensity for going after mosquito-sized issues with club-like solutions, he does get things done, and in Brazil that's saying something!

Cordially,


Howard S. Irwin
Head Curator

HSI:mi

Enc.

CHARLES B. HARDING, *President* • FREDERICK S. MOSELEY, JR., *Vice President* • MRS. HERMANN G. PLACE, *Vice President* • DALE E. SHARP, *Treasurer* • OAKLEIGH L. THORNE, *Secretary* • WILLIAM C. STEERE, *Director*

RECEIVED

Oct. 28 1968

11-4-68

Dear Dr. Irwin:

THE NEW YORK BOTANICAL GARDEN

Thank you for the xerox copy you have sent me. Here at Kew, I had not much time for improving my paper on Taxometrics or Numerical Taxonomy, since I used every possibility in making use of the ample material which exists here. Firstly I collected material, in the form of photographs of herbarium specimens, xerox-copies of diagnoses and notes on all Maxillaria species, a genus where everybody since Liddley said that a monograph is urgently needed, but nobody seems to have had the courage to wade through some fivehundred diagnoses. Now with this material I will at least be able to propose a reasonable definition for the genus and eliminate a number of elements, and then give a subdivision of the genus. Then I did the same with some 25 smaller genera of the Epidendrinae/Ponerinae. I also continued work with Epidendrum, with the object of eliminating species which do not belong there and which complicate seriously the formulation of a clear genus definition. Mrs. Brieger helped me very much, as usually on such trips, in the bibliographical work.

However I did find the time to look up the major part of publications on Numerical Taxonomy and can now conclude that, what I wrote in the Brazilian paper, is essentially correct. All authors seem to me rather weak on the basic biological aspects. From the mathematical-statistical side everything seems to be correct, except that all authors seem to overlook that the final results of computation are only estimates, and thus subject to chance variation resulting from the type of sampling used in obtaining the data, which are fed into the computer. For some, such as the coefficient of linear correlation or the Mahalanobis "distance", the error and chance distributions are known and thus tests for significance are possible, but never carried out. For others even this may not be the case. Thus an enormous amount of work is carried out, but the results are necessarily vague, without any measure of their degree of precision or perhaps lack of precision.

Since the object is to obtain well founded conclusions about biological questions, the basic biological postulates must be as clearly defined to begin with. The situation would be different if, instead of accepting the theory of evolution as basic to Taxonomy, one may return to that of "Idealistic Morphology", assuming the existence of some metaphysical "Ordnungs-prinzip". But with the former, characters which have frequently independent origin, through

parallel or even converging evolution, should have much less importance than those which are in a way unique and thus may serve to distinguish taxa.

In the paper you sent me, I found one mathematical formulation with which I cannot agree. Under the three basic assumptions it is stated, that the similarity measure within two species should always be equal, which would exclude the comparison of a very uniform species, and thus with a high within similarity measure, with another very variable one, where the within similarity measure must be smaller. I think it is sufficient to postulate that the between species similarity should always be smaller, and significantly smaller, than any ^{between} within species similarity.

I plan to write now, after my return, a paper in English on these subjects, as soon as I find the time. As you may know, I do not return to Brasilia but accepted an invitation to go to the new University of Sao Paulo State, in Campinas, as Coordenador Geral, which corresponds somewhat, it seems, to the post of a Vicechancellor here in England. This will allow ^{me} to live in Piracicaba, and thus to remain in close contact with my orchid material. If you should again come to Brasil, I hope we can meet in Piracicaba or in Campinas, while it is rather problematic if I may be able to come to the States. The Internat. Congress could be an attraction, but it is the rule in Brasil to help only for such trips if one participates actively, giving a paper. I feel however that giving a short contribution is not enough justification, and for a longer one one must be invited, which in my case is highly improbable, since I have so far published too few results of my work in orchids. I hope that the work in Campinas will be less wasteful as that in Brasilia, and that it will leave me thus still some time, and especially more time, for my own research.

With best greetings from both of us,

Sincerely yours,

F.G. Brieger

For future correspondence I prefer so far my private one: Piracicaba SP., Avenida Franca 163, Brasil.

FILE
UNIVERSITY OF COLORADO

BOULDER, COLORADO 80302

TAXIMETRICS LABORATORY
DEPARTMENT OF BIOLOGY
ARMORY 101

PHONE: 303-443-2211
Ext. 6712

12 November 1968

Dr. Howard S. Irwin
New York Botanical Garden
Bronx Park
Bronx, New York 10458

Dear Howard:

Thanks a lot for sending the family lists. Sometime when I'm there I will ask you how that list got established, but won't bother you with it now because I am sure it will take a lot of digging around to find where the numbers came from.

With respect to Brieger I am sorry to report that the door clanked shut on symposia at such an early stage that nobody had a chance to propose a session on *Taximetrics*. I wrote way last spring and discovered that the people who make up the program for symposia (Billie Turner is in charge of systematics) got their own interests down and left no room for anything else. I have seen a list of symposia and find them to be the same unimaginative efforts that have damned such meetings for a very long time. All it does, apparently, is give certain people guaranteed tickets. We are left with the decision as to whether we wish to present a series of contributed papers and try to make an *ad hoc* symposium out of them or to put on a demonstration. These are the only two routes open to us, so I am sorry that you have absolutely nothing to offer Brieger.

There is not much I can say in reply to his two pages written to you from Kew. Unfortunately his comments are of such a buckshot nature that I get little from reading what he has said. Much as I would like to support Latin American efforts in this action I simply have no way to communicate with Dr. Brieger unless he be willing some way or other to pass through Colorado; whereupon we can perhaps in a week give him a rundown on the present status in taximetrics and perhaps get through the pseudo-statistical wording which unfortunately damns this field.

Best regards,

David J. Rogers
Professor of Biology

DJR:gm

P.S. You can, if you like, take what I have said about Brieger and pass it on to him. This seems to be the way he chooses to work, so we will continue to allow you to be the middle man.

November 4, 1968

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

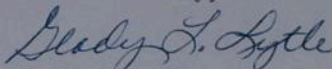
Dear Dr. Rogers :

You can expect to see Judi and I on the morning of November 21. We well understand your reluctance to cut into the weekends of your staff and are most appreciative of your efforts on our behalf.

I will bring as much data as I can carry with the hope that it will not be insufficient nor too bulky.

At any rate, thank you for your many assistances and I look forward to seeing you on the 21st.

Yours sincerely,



Gladys L. Lytle
Department of Zoology
Snow Hall
University of Kansas
Lawrence, Kansas
66044

29 October 1968

Miss Gladys L. Lytle
Department of Zoology - Snow Hall
The University of Kansas
Lawrence, Kansas 66044

Dear Miss Lytle:

We would be pleased to see you sometime in November. The earliest date that would be convenient to us are Thursday and Friday, November 21 and 22. I hope that you can come at that time. You can see that we do not care for weekends. I hope it will not be too rough for you to come at that time. I hate to break into the weekends of my staff members.

When you do come, please bring along some of your data and other descriptive information about your problem. It is usually better if we can talk in terms of your own efforts rather than some hypothetical problem.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

SMITHSONIAN INSTITUTION
UNITED STATES NATIONAL MUSEUM
WASHINGTON, D.C. 20560

28 October 1968

AIR MAIL

Dr. David J. Rogers
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dave:

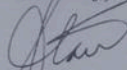
Now that I've had a little time to catch ~~my~~ breath since returning from Boulder, let me tell you how much I appreciated the TAXIR demonstration and enjoyed your hospitality. From the standpoint of FNA it was an especially important visit, and I was glad we had so much time to chat and to clear the air a bit. Once there are some resources I hope that we can maintain regular liaison. Needless to say, we are still busy digesting what we saw in the light of our own developments here at the Smithsonian, but I think we all agree that your system has tremendous potential for the individual monographer.

After we left Boulder, Jim Peters, Dan Piacesi, and I went on to Albuquerque, as we planned, and we were impressed to see the extent of development that Eugene Rypka and his team have achieved at the Lovelace Foundation with respect to applying the computer to bacterial identification.

Sometime this winter or early spring Jim and I hope to convene a small work-study conference here specifically on the use of the computer for plant and animal identification. We hope to keep it small, perhaps no larger than a dozen persons. What I am wondering is whether you would care to participate. If so, when would be most convenient, and could you manage your own expenses? The most difficult problem will be finding enough money to sponsor such a meeting.

Thanks again for the great hospitality. As you know, you were a very welcome sight when we deplaned in Denver upon arrival! Come see us soon.

Sincerely,



Stegwyn G. Shetler
Associate Curator and
Secretary, Flora North
America Project

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

October 18, 1966

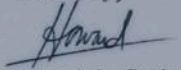
Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado

Dear Dave:

In the course of talking with Dr. Armen Takhtajan, who has been visiting here the last six weeks or so, the subject naturally got around to numerical taxonomy in the USSR. Takhtajan himself seems not too much interested, but did mention a Dr. Yuri Menitsky of the Komarov Botanical Institute, Prof. Popov Street 2, Leningrad P-22, as the leading practitioner and theorist in that country.

Although Menitsky can handle English in letters, his publications will be another matter. Thought you would be interested in the lead.

Cordially,


Howard S. Irwin
Head Curator

HSI:mi

CHARLES B. HARDING, *President* • FREDERICK S. MOSELEY, JR., *Vice President* • MRS. HERMANN G. PLACE,
Vice President • DALE E. SHARP, *Treasurer* • OAKLEIGH L. THORNE, *Secretary* • WILLIAM C. STEERE, *Director*

October 17, 1968

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101A
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

Firstly, thank you very much for sending the program, etc. Judi Turner, the graduate student at ASU who is doing my programming has long since received them, and I have delayed in thanking you because her letter was quite some time in getting to me.

Secondly, the pace of my life has slowed to almost being tolerable. And, if you can forgive the erratic nature of a woman's mind, I would like to make plans to visit your laboratory as you first suggested. Judi would also like to visit at this time. I think we both separately arrived at the conclusion that it would be a disservice to the program, the data, you and ourselves if we didn't take advantage of the chance to visit you and discuss the program.

Dr. Jerry Smith here at KU gives me word that your immediate future includes plans for a taximetric symposium, so I leave the decision of a date in your hands. I would prefer a weekend since I am taking classes and working, and Judi is doing the same. Otherwise, we are open to any suggestion you make.

I am looking forward to coming to Boulder and wait anxiously to hear from you. Thank you for your most thoughtful and considerate attention.

Yours sincerely,

Glady L. Lytle

Glady L. Lytle
Department of Zoology
Snow Hall
The University of Kansas
Lawrence, Kansas
66044

CC: Judi Turner

BARSS, REITZEL & ASSOCIATES, INC.

134 MT. AUBURN STREET
CAMBRIDGE, MASSACHUSETTS 02138

(617) 864-4320

September 6, 1968

Professor David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Professor Rogers:

We were considerably relieved to receive your good letter of 26 August. Thank you for it. I may say that we had been pretty disheartened here by the initial misunderstanding.

Whenever we propose usage of your similarity-clustering program, we identify it as the "Rogers-Barnes Flowers Program" and specify that it was initially developed at Colorado State and the University of Colorado by you and Mr. Brill. Following is an up-to-date list of sponsors of projects in which we have made/are making application of the program:

1. Advanced Research Projects Agency. Our first employment of your program, this project had as one requirement the classification of various types of social phenomenon in cities experiencing rapid growth.
2. Center for International Studies, MIT. The program was employed in analyzing attitudinal data of a panel of senior European officials on supra national questions. For this study, we found that a weighted single linkage approach yielded particularly interesting results.
3. Harvard-MIT Joint Center for Urban Studies. Here, our purpose was to classify sub-groups within Boston with respect to their position on certain local political issues. Single linkage here proved as useful as a weighted approach. We are about to begin testing our average linkage program (a modification of the Kansas University NTSYS program) on these same data. (A parenthetical footnote -- do you feel a piece on our evaluation of the relative utility of single-linkage, a weighted approach, and average linkage, would be of general interest? We suspect it might.)

September 6, 1968

4. Office of Economic Opportunity. Our full battery of taxonomic programs will be used here as part of a major effort to identify the attributes of successful Community Action Programs. We may also use the Rogers-Barnes program in classifying the attitudes of ghetto poor regarding themselves and their environment.
5. Organization of American States. The Rogers-Barnes program will be included within the management information system we are beginning to install in Bogota. The system will be used by the Inter-American Center for the Promotion of Exports (CIPEX) to assist local manufacturers in identifying potential markets for their products, and for determining detailed product specifications which they must satisfy in order to successfully "penetrate" target markets.

We have several other applications in prospect as soon as survey data have been generated for analysis. This suggests to me that we should perhaps send you an updated statement of applications shortly before each of your returns to NSF for new funding. If this would be helpful, do let us know the date on which such information would be needed -- and also any suggestions on useful descriptive information to supplement the types of statement made above.

We are greatly interested in both CHARANAL and TAXIR and would welcome descriptions of each program. Since our programming staff is small and presently overextended, I suggest that it might be premature for us to request full program listings at this time. But your brief description suggests that both programs would fill gaps and otherwise increase the power of our total systems capability as soon as they can be integrated. At present, our system is built up around Rogers-Barnes, Kansas' NTSYS, Harvard's DATA-TEXT, California's BIOMED, Michigan's AID and MIT's SLIP. This gives us a pretty broad base for analytic work.

Sincerely yours,

Lawrence W. Barss
Lawrence W. Barss

LWB:jt

SEP 5 1968

0
68

ARIZONA STATE
UNIVERSITY

DEPARTMENT OF ZOOLOGY

TEMPE, ARIZONA 85281

September 4, 1968

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

I was very gratified by your letter of August 30th and would be pleased to visit your laboratory. However, my commitments at this time are such that I will not be able to do so. I am in the process of leaving Tempe for the University of Kansas. It will be quite difficult for me to get out of Lawrence to visit anywhere for the next year as I will be finishing the work that is left on my thesis and beginning my doctoral work.

Consequently, I would appreciate it very much if you would send the documents I requested. I have 3 friends here who are graduates in computer programming and they have handled the programming for me to date. If we encounter any difficulties, may we make inquiries of you? With their help, any questions I may have about the program and its methodology will be sent to you in a more coherent state than if I were to attempt the task alone. It is unfortunate that the process will be lengthier in time, but, hopefully, it will be as successful as if I were to visit you. If you require information about my data, please write me at KU. If not, then forward the materials to Dr. Minckley at ASU.

I am very grateful for your thoughtful consideration of my problems and thank you for the time and effort you have expended in my behalf.

Yours sincerely,

Gladys L. Lytle
Gladys L. Lytle

9 Sept

2 reprints
Writeup for Similarity
How to Read the Printout

30 August 1968

Miss Gladys L. Lytle
c/o Dr. W. L. Minckley
Department of Zoology
Arizona State University
Tempe, Arizona 85281

Dear Miss Lytle:

We will be pleased to send you the information you have requested concerning our graph-theory clustering program. However, I must suggest that we have an alternate proposal which I believe you should listen to.. In our work with other users of our program we have discovered that a considerable amount of time, effort and energy can be saved if it is at all possible for you, as the investigator, to come to our laboratories for a day or two to discover the processes necessary to make this program do the best work for you. My staff and I will be glad to discuss your problems and advise you on the procedures which we have found to be most useful. Primarily we are concerned with the structuring of your data. There are many hidden processes which unfortunately are extremely difficult to express in published works, which if known about can give you results that are of considerable more benefit than otherwise.

Let me suggest therefore that if at all possible you visit our laboratory, bringing with you your data as you presently have it constructed. We will then be able to give you all the necessary advice for the continuing operation. I suggest that perhaps one full day will be required, and if you can spend another half day with us this will be an even better arrangement for you to understand the intricacies of the methodology. Thank you for your inquiry and I do hope that you will find it convenient to visit us here. If, however, it is inconvenient and you cannot find ways to take off time to come, then we will send you the documents that you have requested.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

AUG 20 1968

ARIZONA STATE
UNIVERSITY

TEMPE, ARIZONA 85281

DEPARTMENT OF ZOOLOGY

August 26, 1968

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101A
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

On the suggestion of Dr. Clair Stalnaker I am writing you. I am interested in running the program that Paul Holden ran on his data on Gila. I wrote Dr. Stalnaker, as I had no information as to the whereabouts of Holden, and asked him to fill me in on the details surrounding the IBM catalog number of the program, where Holden ran it, and any other information Dr. Stalnaker deemed helpful.

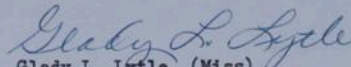
My own research centers on a genus of fish that is in the same family as Gila. My Advisor here at Arizona State, Dr. W. L. Minckley, put me on to the program as he was familiar with Holden's work and his finished thesis, of which we have a copy. I have been running my data on Analysis of Variance, which though it has been helpful, has not been just right. I am very interested in Holden's work and find that the program he ran is the closest to what I want than anything I have found.

I would appreciate any information you can send, and I would be most grateful if you would send it at your earliest convenience. I would appreciate the catalog number, the listing and documentation, the system on which it was run, including the compiler level, and the language. I understand that it may be difficult for you to supply me with the information, and I would appreciate anything you can return.

Please forward any correspondence to Dr. W. L. Minckley, Department of Zoology, Arizona State University, Tempe, Arizona, 85281

Thank you very much for your attention and consideration.

Very sincerely,


Gladys L. Lytle (Miss)
Department of Zoology

26 August 1968

Dr. Lawrence W. Barss
Barss, Reitzel and Associates, Inc.
134 Mt. Auburn Street
Cambridge, Mass. 02138

Dear Dr. Barss:

I am (finally) in a position to respond to your letter of July 22, and to acknowledge, with thanks, the various documents and tapes which Mr. Barnes and Miss Aiello have sent.

It was kind of you to explain your position in response to what was unintentionally, and unjustified, a fit of pique on our part. May the whole matter rest now that we understand your position, so well displayed in your letter. We are aware of the problems of soft-ware protection, and of the difficulty of making its sale a profitable (or even break-even) venture. While we in this shop are not able to sell our product for dollars, we do get our "pay-off" by use of our programs, and in this sense, we are grateful to you for taking up our similarity-clustering program. We hope we have a satisfied customer in Barss, Reitzel et al. and that you may have luck in getting users of the program. We recognize that your input to the program development was nontrivial, both in converting to IBM hardware, and in the new sorting routine which is much more efficient than the one we had.

Since we are a "satellite" to the University of Colorado, we sympathize with your organization's desire to more easily cut across departmental lines for interesting work. We are interdisciplinary, with biologists, mathematicians, economists and programmers, yet we have not made a complete break with the University, and since we gain all our support from grants from NSF, NIH, etc., we must produce sufficiently satisfactory work to keep these agencies willing to continue our grants. It is therefore important that we get recognition from people for our efforts. We would be pleased if you could give us that recognition when you find customers for our methods, and further, let us know the type of application made of the program. Also, if this isn't too privy information, the names of the users. We use such information only for purposes of gaining additional grant support, for impact-studies etc., and would under no circumstances, infringe upon your territory.

Since you have, for some reason, found the single-linkage program interesting, I wonder if you would also be interested in companion programs to it? For example, our program CHARANAL, intended (in an information-theoretic sense) to give measures of the values of the input data in a study, where, given the value of one piece of information, another information^{per} is found to cover some part of the first, is distance from the first, or is identical to the first. This program has been very useful to us and to other users. Our big effort now is TAXIR, an information retrieval system for biology, and we have nearly completed version one. There are some quite distinct methodologies in this system, both in terms of efficiency of storage of information, and in terms of user facility. An important feature of the system is that the data input technique is nearly identical to data structuring of the two other major programs, thus allowing data massaging in several interesting ways.

If either of these more recent developments are interesting to you, we would be happy to explain them in more detail.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

AUG 12 1968

BARSS, REITZEL & ASSOCIATES, INC.

134 MT. AUBURN STREET
CAMBRIDGE, MASSACHUSETTS 02138

(617) 864-4320

August 8, 1968

Professor David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Professor Rogers:

Enclosed, as promised is a copy of the listings of the similarity-clustering program as modified by myself for the IBM 360/75 and the documentation of the changes made. Under separate cover, you will find a 9-track tape containing the card images, one physical record containing 40 80-byte logical records.

I am sorry for the delay in response. If I can be of further assistance, please do not hesitate to contact me.

Sincerely yours,

Richard A. Barnes

Richard A. Barnes

Enclosure

RAB:jt

19 July 1968

Miss Aiello
Barss, Reitzel & Associates, Inc.
134 Mt. Auburn Street
Cambridge, Mass. 02138

Dear Miss Aiello:

My associate, Mr. Brill, did indeed receive from you on the day you mentioned, a listing of our program which you converted to the IBM 360.

However we would be very happy to have a source deck and any documentation which you have done. The reason for this request is that many of our collaborators want to know if a conversion has been done to the 360 system and we would like to be able to supply them with listings. In as much as we gave the program to Barss, Reitzel and Associates in the first place, we thought it would not be too much to ask for a source deck. If you find this unreasonable, however, we would pay the postage for the cards.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

BARSS, REITZEL & ASSOCIATES, INC.

134 MT. AUBURN STREET
CAMBRIDGE, MASSACHUSETTS 02138

(617) 864-4320

July 22, 1968

Professor David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Professor Rogers:

Your 16 July letter to Mr. Barnes, and 19 July letter to Miss Aiello, have found their way to me with a marginal comment that you appear distressed by some action on our part. I hope a note may vaporize the teapot, if in fact we have a tempest therein.

We are delighted to be able to submit the materials you requested and should have been happy to do so at any time subsequent to our transmittal of the revised listings to you on 11 October 1967. Rather than a source deck, however, we are taking the liberty of having a mailer tape prepared. This lessens the possibility of damage in shipment.

Since there seems no cause for your group's distress with our actions, my inclination is to suspect that you may have heard that we failed to respond to the University of Illinois in a fashion satisfactory to them. My hope is that the enclosed copy of our correspondence with ^{Urbana} Champaign, plus the remainder of this letter, will suggest that such has not been the case.

There is for us a critical -- possibly irresolvable -- conflict between the need for free access to knowledge, and the equal need to remain solvent. This latter is not easy, since we are trying to accomplish academic-type research without an endowed base.

B/R is almost entirely made up of professors *maquette*, or graduate students. I used to be a member of the MIT faculty; our Vice President is the former Chairman of Social Sciences at Haverford; et cetera. Our academic credentials are, I think, attested to not only by the fact that both Harvard and MIT have contracted with us to perform data analyses, but also by the fact that certain of our staff have been given formal approval by Harvard to work with us parttime while

July 22, 1968

holding Harvard Prize, Woodrow Wilson or NSF fellowships. The Harvard position has been that B/R's work contributes usefully to scholarship and to the educational process per se.

We are in business for two reasons. First, we feel we can organize interdisciplinary research efforts more easily than a university, since we are not bound by departmental, or other administrative regulations. Second, we feel that transliteration of basic research advances to immediately useful purposes can more simply be done outside the university than within. We believe, then, that we have a metier, and we are sufficiently certain of our objectives so that the stockholders have risked substantial capital and time to the venture.

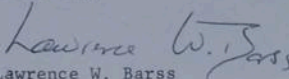
The effort is exciting, fraught with uncertainty, and -- so far -- totally unprofitable. This has relevance in the systems area since we have invested heavily in the extension and modification of programmes such as your single linkage one. While we are anxious to share our results broadly, we feel we are justified in trying to limit usage of our programmes to groups which would not be competitive with us in the social change areas which are our primary interest. Concern with unrestricted competition lay behind our suggested restrictions to ~~Champaign~~^{Urbana} which, you will note, were almost identical with the terms Messrs. McCormick and Flowerdew suggested to us.

Our problem is made more complex by the fact that faculty members of two local universities have bluntly tried to "pirate" some of our methods specifically in order to compete with us vis à vis certain potential sponsors. We are sure that ~~Champaign~~^{Urbana} would conform to the letter of its commitment to us. But once we have released our work to members of one department in one university, how can we justify not giving it to others?

This begins to sound like complaints in the Double Helix. I have written more than planned because your apparent distress concerns all of us -- and we think you may have misunderstood our position throughout.

Possibly our 11 October letter somehow missed your desk. I am enclosing a copy of that letter to make the record complete. Would you, incidentally, be interested in seeing the results of our work on development of an average linkage system after it has been debugged?

Cordially,


Lawrence W. Barss

LWB:jt

Enclosures



DEPARTMENT OF COMPUTER SCIENCE UNIVERSITY OF ILLINOIS URBANA, ILLINOIS 61801 U.S.A.

July 11, 1968

Lawrence Barsa
Barsa, Reitzel and Associates, Inc.
134 Mt. Auburn Street
Cambridge, Massachusetts 02138

Dear Dr. Barsa:

I am writing in reference to my telephone call of July 9th in which I requested a deck, listing and relevant documentation of your 360/75 program of D. J. Rogers' graph theory clustering model.

1. Our investigations are purely exploratory and in the nature of feasibility studies to find whether methods of numerical taxonomy might be applicable to the classification of documents and file structure organization of an automated library.
2. Usage of any program and documentation which you might send us would be internal to this group and company confidential. Also, no high volume usage would be made of any such program.
3. We are particularly interested in obtaining a listing so that we can make suitable modifications if they become necessary.
4. Our reports are public and we will be pleased to keep you informed of future developments.

For your interest, please find enclosed a classification bibliography. Please mail any correspondence to S. J. Flowerdew

Very truly yours,

B. H. McCormick (Principal Investigator)

S. J. Flowerdew (Research Assistant)

SJT:djs

July 16, 1968

Mr. S. J. Flowerdew
Research Assistant
Department of Computer Science
University of Illinois
Urbana, Illinois 61801

Dear Mr. Flowerdew:

We should be happy to send you a tape, listing and relevant documentation of our Rogers-Barnes 360/75 taxonomic programme on the following conditions:

1. That the University of Illinois absorb the modest costs for tape and listing reproduction;
2. That the programme in its entirety be used only for documents classification and file structure of an automated library, as specified in your 11 July letter;
3. That programme usage be held internal to your project group and handled as company confidential;
4. That any references made to the programme list Rogers and Brill of Colorado, and Richard D. Barnes of Barss, Reitzel as major contributors to programme design.

If these conditions are satisfactory to you, please have a cognizant official sign the duplicate copy of this letter and return it to me. We should then be able to forward materials to you within a fortnight.

Very truly yours,

BARSS, REITZEL AND ASSOCIATES, INC.

Carol-Ann Aiello
Administrative Officer

Signature & Title

Date

October 11, 1961

Mr. David J. Rogers
Mr. Robert C. Brill
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers and Mr. Brill:

Earlier this year you were kind enough to send me listings of the similarity clustering program. I revised it for use on an IBM 360/75. The program successfully fulfilled our requirements, and for this I am most grateful. As promised, I am sending you a copy of the revised listings along with my own commentary. The comments cards were not changed so the following will be necessary to explain the changes.

Because of the tremendous size of the 360/75 (262, 144 words), it proved possible to keep the entire program and data set in core at all times. This eliminated the need to use secondary storage devices and speeded up program execution immensely. Coupled with the small cycle time of the machine (200 nanoseconds), this allowed us to cluster some 400 objects, with 27 attributes each, in 6 minutes.

MAIN has been expanded somewhat to print out information as to the execution times of the 4 phases. TIMERS is a systems subroutine provided by AVCO which automatically prints out the elapsed time between each pair of calls. The loop ending at statement number 15 is used to clear the unused part of core in case a dump should be necessary. The dump routine does not print explicitly identical rows in storage, thus saving paper and print time. All data is now stored in various COMMON blocks to facilitate program communication and to permit overlaying of old data once it is no longer needed.

SIMLAR required few changes beyond the elimination of the tape writing. The input format statements were changed to allow for the fact that the 360 word holds 4 characters. I am now in the process of writing a routine to speedup elimination of identical objects as the present method takes longer than necessary thanks to the existence of variable-length compare instructions.

PKEDIT is the name I have given to the deck containing the editing and unpacking routines. PREPAK now edits all the characters at once using the TRANSLATE instruction (TR) which by means of a translate table automatically changes characters from one coding system to another.

UNPACK unpacks the characters, storing them in KOMP1 or KOMP2 and checks for illegal characters. Due to the advantages the 360 has in handling characters PKEDIT was made more efficient.

The SORT program was completely rewritten and uses a different algorithm than originally. Due to the large data pool it was written in assembly language to allow tight coding. The algorithm picked for SORT is the shell sort, probably the fastest comparative internal sort now available. Its running time goes up by a factor of $n \log_2 n$, where n is the number of words in the table to be sorted. During each pass the table is divided into subsets whose members are an equal number of words apart, and these subsets are completely sorted. During the first pass each subset contains only 2 or 3 members roughly $n/2$ words apart. For each successive pass the interval between members is divided by 4 if the interval is greater than 15 or 2 otherwise. The interval is adjusted to make it odd to avoid subsets made up of exactly 2 or 4 previous subsets. Due to the large amount of ordering produced by the first passes the sort is amazingly fast taking about 16 seconds to sort ORANGE and LEMON with 84255 words apiece in the largest case. The sort is completed when the last subset, whose members are adjacent, is sorted. You will note that I have continued the tradition of using picturesque variable names.

The only changes in GRAPH and FINDK result from the elimination of tape operations and the packing of object pairs, 1 object per half-word. Use of 360 FORTRAN IV-A allows direct reference of halfwords, thus making the handling of these object pairs easy. GRAPH also overlays data used in SIMLAR which is no longer needed.

From your printout which showed the nodal distance arrays, it was easy to write NODIST, a simple routine to do this.

All in all we were quite happy with the program you sent us and are happy to send you the listings of our revisions. We thank you for your cooperation and would be interested in hearing of any further work done by your group in this field.

Sincerely,

Richard A. Barnes

RAB:cpa
Enclosure

16 July 1968

Mr. Richard A. Barnes
Barss, Reitzel and Associates, Inc.
134 Mt. Auburn Street
Cambridge, Mass. 02138

Dear Mr. Barnes:

The Taximetrics Laboratory would appreciate receiving from you a source deck and program write-up, including flow charts and other explanatory documents which you may have for your recent transcription to IBM 360 of our program GRAPH, which we made available to you last year. These documents and decks are not only of ~~extreme~~ interest to us in our own research, but may also prove of use to the many scientists who cooperate with us.

Please note our change of address. Thank you.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

28 August 1968

Senator Jennings Randolph
Senate Office Building
Washington, D. C.

Dear Senator Randolph:

May I please have a copy of "The prospect for Technical Transfer", the report of your subcommittee on Science and Technology. I am interested to know whether or not my research has anything to offer somewhere within the framework of the recommendations of that report. I have a small team that has developed a computerized information retrieval system useful in science and other areas and would like to be able to contribute this system if it is possible to do so. In as much as the information system was developed with federal funds from the Office of Information Services of the N.S.F. I should think that we should make an effort to "get the system off the shelf" and into use in such areas as your subcommittee's efforts have been investigating, particularly on a transfer of technology.

Thank you for your efforts.

Sincerely yours,

David J. Rogers
Professor of Biology

DJR:gm

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

6
July 22, 1968

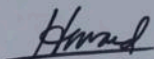
Dr. David J. Rogers
Professor of Biology
Taximetrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dave:

Thanks a lot for your efforts toward the education of Friedrich Gustav Brieger. I hope that by the time I get back to Brasília he will have had time to cogitate over your remarks and the series of papers sent along (xeroxes were made here of the two missing papers). Whether we can get him to venture so far west is another matter. You can be sure I'll push for it.

The Manihot roots that emerged from the confusion are on their way to you this week.

Cordially,


Howard S. Irwin
Head Curator

HSI:mi

1 July 1968

Dr. M. T. Hall
Morton Arboretum
Lisle, Ill. 60532

Dear Truf,

Misty Fogg is due to show up here on the 15th of July. He has sent me a few pages of numbers which appear somewhat "manipulated," apparently for an ecological analysis of *Juniperus*. He seems to have the idea that I am aware of the project that the two of you have, and that you have communicated with me some of the work. Somewhere or other I have not caught up with the project, though I am interested to know if there is something I can do to be useful. I'd appreciate it if you would lay on some of the objectives and background, questions you are asking, etc. to bring me up to date. I really want to work in on this project if there is anything useful we can do.

There are several interesting straws in the wind that we are doing that I think you would find helpful in your own operation of the Arboretum, and I would love to have a chance to tell you in detail about these. Is there any possibility that you could pop out here during the time that Fogg is here, so that we may go over them? We would be happy to have you. Let me hear from you as soon as possible and I can make arrangements for putting you up.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

12 July 1968

Dr. Alfred Etter
R. Rte. Box 117
Aspen, Colorado 81611

Dear Al:

My thanks for your and Helen's kindnesses during our visit.
We enjoyed seeing both of you very much.

I do hope that we can pick up some cooperative effort on the
Indian paintbrushes. If you care to do so, it would be well for
you to come over sometime for a couple of days (we'll put you up)
and bring with you your data sheets as well as some of the specimens.
Then we can go ahead with the work. What we usually do is to give
you some runs on the computer for free. In the meantime I am send-
ing along some of our reprints which bear on the computer work.

Thanks again for everything.

Sincerely,

David J. Rogers

DJR:gm

30 July 1968

Dr. Robert A. Davidson
Department of Biology
The Catholic University of America
Washington, D. C. 20017

Dear Bob:

My immediate response to yours of the 23rd with your request as to other granting agencies than NSF is to say "You are a lot closer to the horse's mouth than I am." I suggest that you use some of your "intuitive" taxonomic skill to identify the most likely source of funding. You are not the only one who has been shot down this year.

We have found out that it pays to put me in Washington about once a year for a week going around to the major agencies where I get an informal reading of what will and will not grab the fancy of the agencies. I take careful note of the words that are "in" at the moment, then try to come home and use these words to say exactly what I want to do. Following this I make an informal presentation (and you are in a better position than I to do this) and find out whether I have struck a responsive note or not. This kind of feed back seems to pay off.

When you are going after "scratch" forget all your desires to be a scientist and remember you are in the political arena which right now has many lions and tigers in it. Other than these pearly words of wisdom I have no real solid things to say.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

The Catholic University of America
Washington, D. C. 20017

DEPARTMENT OF BIOLOGY

July 23, 1968

Dr. David D. Rogers
Taximetrics Lab
Department of Biology
University of Colorado
Boulder, Colorado

Dear Dave,

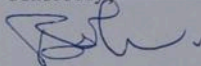
Hope you survived the Mexican heat and intestinal bugs.

Two topics:

- 1) Thanks a lot for the program, instructions, etc. I don't think it has been converted successfully yet but presume it will be straightened out with time.
- 2) In a spasm of mental arthritis NSF (Systematic Biology) shot down my proposal for another two years of support. This, after hundreds of reprint requests for several papers (mostly not from taxonomists) and just when things are really getting interesting. Can you suggest appropriate grant sources other than NSF to whom I might apply.

I am damn disgusted with the powers currently controlling NSF Systematic Biology but I guess nothing can be done. It sounds hokey to talk of the need for men with powerful intellects, foresight, and imagination--but we sure as hell need them.

Sincerely



Robert A. Davidson
Professor

Some gaps! NSF has been
very good to me & actually am
very grateful ————— Bat — — —

2 July 1968

To Drs. Brieger and Irwin:

I think it is evident from Dr. Brieger's letter to Dr. Irwin on the criticisms of the Cassia, section Apoucouita, paper that the philosophy behind the work is what needs discussion. In this connection, let me say that our efforts are to reinforce the decisions of the taxonomist, not to replace him with some "objective procedure. Let me say that I find nothing wrong with the subjective approach to taxonomy which has proved itself in numerous applications before the recent rise of models in mathematical framework which are intended to reflect the procedures of the taxonomist. The most important thing to realize is that there are many different possible types of classifications which are equally valid. It behooves the individual to select that type which is important to him. In most classifications of species, genera, families, etc., of plants, we rely very heavily on the phenotype, and only in very rare cases can we bolster our classifications with chemical, cytological, ecological and evolutionary types of information. From our classifications of the phenotype (phenetic classification of Sokal et al.) we may be able to derive some hints about the phylogeny of the group (phylogenetic classification).

With regard to our derivation of similarities, it is very important to understand about characters. All you can say about a character is that it is some basis for comparison of the organisms under study. Which characters you employ are your responsibility, and no statistical model can dictate to the taxonomist how many, or what kind can be employed. How you structure your characters is the business of the professional taxonomist, and is your prerogative. Again, no statistical model can or should be allowed to tell the taxonomist what he must do in establishing his attributes within the character. This philosophy throws the burden onto the shoulders of the taxonomist, where the burden should rest, to do the very best job possible to structure his characters in such a way that they do the job they are intended to do - make a classification. But there is nothing sacred about the way characters are structured, and there should be some system which gives the taxonomist the most aid in establishing the attributes of a character in such a way that the character does what the taxonomist knows the character must do - to make a satisfactory classification. Our approaches are not intended to dictate, but to be hints to the professional about how he can best use the information which he gathers.

This philosophy is back of my earlier statements, in the first letter, which said that it behooves the taxonomist to know the rules of classification, precisely defining his problem, and then seek out the most powerful mathematical expression of those rules. To apply a statistical technique established for an entirely different purpose is to do damage to the taxonomist's desires. If Brieger thinks otherwise, I am sorry for him. We

have tested enough of these "ad hoc" approaches to know what we're talking about, and I haven't got time to rebut any statistical arguments. But to get stuck with only one small branch of applied mathematics, not knowing even the mathematical base of the statistical approach is to do damage to oneself, and to be laughed at by professional mathematicians.

Now that is out of the way, I would suggest the following list of readings, some of which you will have to provide because we are completely out of reprints.

1. For an understanding of the graph theory model of classification, read:

Wirth, Estabrook and Rogers, Syst. Zool.

Estabrook, J. Theor. Biol., 12: 297-310. 1966. (We're out of this)

Irwin and Rogers, Mem. NYBG, Cassia.

2. For an understanding of characters, read:

Estabrook and Rogers, BioScience.

Estabrook, Taxon.

3. For an overall look at computers in taxonomy and evolution, read:

Rogers, Fleming and Estabrook, Chapt. 6 in Evolutionary Biology, Vol. 1, 1967. (eds. Bobzhanski, Hecht, and Steere; Appleton, Century, Crofts, N.Y.) (We're out of reprints)

Since our mutual work on Cassia, we have come quite some distance in understanding ways of supplying the taxonomist much more flexibility in using the computing machine to back up his work. For example, the BioScience article by Estabrook and Rogers (under 2 above) indicates how you can structure characters to reflect the various situations which arise with regard to qualitative and quantitative characters. How to estimate the value of a character as a classificatory structure is given by Estabrook in Taxon.

More papers are in press which I will forward as they come out. It will be difficult to see how our approach differs from the Kansas school (Sokal et al.) from reading of the individual papers, but perhaps something of our philosophy can be read in the Evolutionary Biology chapter.

I know that I have not answered any of Brieger's individual questions, but I think it more important for him to know that our philosophy is quite different from the other approaches, which might be termed "heuristic" for want of a better description of the work done in Kansas. Knowing this difference, he may read the reprints with more understanding. He should also know that we continue to refine our techniques, and that we are not entirely satisfied with any of them. We're still in the experimental stage, and to damn the experiments is to damn the usual approach to scientific developments.

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

JUN 24 1968

June 21, 1968

Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dave:

The miscellaneous specimens you returned are on the scene and I have John Grear trying to sort things out. Thanks for taking the trouble.

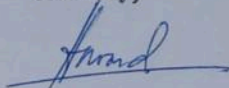
I enclose herewith a copy of a letter I received today from Friedrich Gustav Brieger, this, as you will see, in response to your letter to me about his paper on Numerical Taxonomy and also a commentary on our joint paper on Cassia Sect. Apoucouita.

As the meat of his critique deals more with your part of the operation, I hope you will find time to (1) provide the rebuttal, counter-information, or whatever in response to the points he raises, and (2) indicate which references will supplement the comments you make with background data. I can have xeroxes of these made from our library. If you send your letter here, I will include one of my own together with the references you suggest.

As I have indicated earlier, I believe Brieger to be the outstanding thinker on numerical taxonomy (or whatever he chooses to call it) in Brazil if not Latin America, and deem it very worth our while to not let the matter simply drop.

Hope your Mexican trip went well.

Cordially,



Howard S. Irwin
Head Curator

HSI:dz
Enc.

722 Schwennungen a.H.
Oberdorfstr. 21
Germany

June 8, 1968

Dear Dr. Irwin:

I thank you for your letter of May 14th and the copy of Dr. Rogers' letter of April 22. I must also inform you that most unfortunately I had to change my plans, delaying departure and cancelling my visit to the USA, at least at the beginning. I must also inform you, that my proposal to my successor, to draft a new agreement with you jointly, was treated with the now usual technique of delaying such decisions until after my departure. Now some observations regarding Numerical taxonomy:

Regarding the criticism of Rogers, that I should do more reading, I fully agree with him, and this is one reason why I am beginning the discussion with him through you. Apparently Rogers has no idea about the consequences of scientific isolation in Brazil.

I do not see any difficulties in applying to Taxonomy statistical methods set up for other problems, as long as the nature of the model or procedure is adequate for application to other problems.

I was of the impression that the term "Numerical Taxonomy" was invented by Sokal and thus my criticism was formulated on this basis. However, I do not do did not include under this term any form of application of statistical methods to Taxonomy. I believe it is a very poor term anyway!

Unfortunately I do not think it will be possible to spend some time with Dr. Rogers, but expect that the reprints or xerox copies which you will send me contain all the necessary informations, allowing to form my own opinion on the procedure. Please send these materials to our daughter, where they will reach me after June 15th, and let me know how much I owe you then.

I read with care your second Cassia-paper, and must say the following. I believe I understand your method as described until item 8 incl. (on pg. 81), but not the later part. What are "partitions and why are there 26 possible ones, and what is a "value of threshold"? Is this value determined a priori to determine similarity or is the determination of what should be called similarity established subjectively first?

Evidently, there seem to exist two steps also in your method: (a) of calculating some kind of statistical parameter, which in accordance to my understanding item 8, is the sum of all common attributes for a total of n characters, for each pair of specimens, or the corresponding fraction in the total of characters. Thus maximum similarity will be indicated by the value n (for the sum) or 1 for the fractions, and the minimum would be zero; (b) evaluation of significance of all intermediate values of this parameter.

Regarding 2, the statistical method of confounding data referring to totally different characters should avoid the subjective bias, involved in the usual taxonomic technique. I doubt if this is achieved. The information furnished by each of the n characters depends upon the number m_i of attributes of each character. Let one assume, that the distribution of attributes in specimens should be fully at random, then the probability of similarity in each one would be $1/m_i$. If characters are alternative with $m_i=2$, the probability of similarity would be $1/2$, and decreases with increasing numbers of attributes per character to $1/3$, $1/4$ etc. Thus the number of attributes per character diminishes the amount of information contributed to the final similarity-value, and this does not only repre-

sent bias, but a serious one, when the classes are established arbitrarily for quantitative characters, where we may choose classes of any size and number. Since the objective is the clustering of specimens into taxons, the inherent degree of variability per taxon may also cause bias. Thus for instance, regarding leaflet length, species nr. 5 with a range from 10 to 20cm would furnish representatives which may show similarity with a number of other species, in a much larger degree than representatives of more constant species. Thus the degree of variability of characters may cause another degree of bias. Thus the obtaining of the final parameter for similarity included subjective bias and one should take into consideration whether this bias is less in the "taximetric" procedure than in the ordinary taxonomic one. I suppose these questions are fully and much better discussed in the general papers on the method, which I do not know yet.

Regarding item b, I hardly can formulate questions without understanding fully this item. However, it seems to me that everything depends upon comparisons to be made between similarity values. These values as calculated represent estimates only, and thus one ought to know their standard error and chance distribution. One point seems to me evident. Since the similarity parameter can only vary within finite limits (of zeroth one for fractions as stated above), for estimates nearing these limits some kind of mathematical transformation will be required to transform these limits into values plus and minus infinite, and thus at least making the ideal distributions symmetrical. If however in your method, the values for the similarity parameters are taken as precise solutions and not as estimates, variable with regards to an unknown constant inherent to the material, since then I am afraid one returns to fully subjective judgement and the taximetric method would gain little over the old taxonomic method. However, I expect also these questions are fully discussed and resolved in the literature I shall receive.

Hoping that you will find the time to continue this discussion which for me is very interesting, I remain

Sincerely yours

F.G. Brieger

Dr. H. Irwin
New York Botanical Garden
Bronx
New York 10458
USA

22 April 1968

Dr. Howard Irwin
New York Botanical Garden
Bronx Park, Bronx, N. Y. 10458

Dear Howard,

The copy of F. G. Brieger's paper just came. I read the English summary but did not spend any time on the text in Portuguese. If we can take some of the English summary statements as to what he believes in, then I think that Dr. Brieger should spend some more time reading before he does anything else. The first thing I object to is the fact that he calls numerical taxonomy a new branch of the biological sciences. This it is not. I am sure that Sokal would attempt to have us believe that it is, but under no circumstances can we claim anything really unique, so far.

Some of Brieger's criticisms of the statistical models that he finds displeasing are valid but I can't see a hell of a lot of reason for them. The more damning problem is the fact that the statistical models were established to deal with the kinds of work other than taxonomy. This seems to be the focal point of the difficulty; that the attempt to shove taxonomic data through the statistical models of which Brieger speaks is an incorrect approach. It behooves the taxonomist to establish the rules he wants and then to get a mathematical model which will be a reflection of those rules.

If Brieger is going to put on a seminar for you, I sure hope he does more reading in the literature and includes some of our works as well. I didn't see our reference at all, and I think we have just claim to be disturbed that he has not. Your own work on *Cassia* is a very good example of some of the things that should be mentioned.

Of course my comments are made on the basis of what I read in his English summary. There may be some other nuggets which go into these sorts of details that I have not read in the Portuguese section of the paper. If Brieger is travelling in this country at all, you might tell him that we would be happy to see him here for a while to let him know what is doing in this shop. He might be interested and might be able to arrange it on a return trip basis to his own country. Thanks for sending me this paper; I am always interested.

Sincerely,

David J. Rogers
Professor of Biology

DJR:am

Com referência a bibliografia, quero reiterar que desisti de citar publicações, nas quais se aplicam os métodos da Taxonomia Numérica, pois não achei nas mesmas, discussões das justificativas da metodologia, sendo a questão central sempre a da taxonomia de um determinado grupo. O tratado de SOKAL, de outro lado, trata destas justificativas, de certa forma, isto é, mais em forma de postulados e não na forma de demonstrações, sem citar, por sua vez a literatura especializada.

re RESUME

The present paper represents an analysis of the basic requirements, and the scientific foundation in general, of the methods of the new branch of biological sciences, called Numerical Taxonomy. Since all papers, which I know, using such methods do not enter into a discussion of the method itself, but refer to special taxonomical problems, special literature references are not given. The book by SOKAL can be considered as an exponent of the new methodology and thus is the only one cited.

The basic claim of scientists, regarding the new method as superior to the classical method of taxonomy consists in that the new procedure is supposed to represent an objective approach and substitutes the old one, considered as excessively very subjective. It must however be remembered that the use of scientific intuition is considered perfectly justified in many branches of science, and it can be described, in terms of more modern terminology, as the result of computations by the "cerebral computer". Such intuition has been and certainly still remains the main basis of scientific work in nearly all branches, requiring that a demonstration should be given "a posteriori", showing that the conclusions reached are in fact correct. In biological sciences, such a proof consists, in general, in showing that the hypothesis formulated is the simplest one which explains uniformly and without requiring correcting additional hypothesis, the largest number of facts of a similar nature. This represents a difference, at least to some extent, to the situation in the so called "exact" science, where exact demonstrations "a posteriori" of a conclusive nature are possible.

The claim of increased objectivity of the methods of Numerical Taxonomy refer to two phases of the work.

Firstly two principles are established regarding the obtaining of data, which should be (a) of a quantitative nature, free of any bias, and which (b) should refer to all characters of a taxon, without any type of selection. Regarding the OBTAINING OF DATA IT IS IMPOSSIBLE TO DECIDE OBJECTIVELY, IN THE CASE OF MEASUREMENTS OF SIMILAR DATA of a quantitative nature, in which scale they should be taken. Variation may follow any scale, arithmetic, geometric, logarithmical etc., and any objective proof would require "a priori" knowledge of the mathematical functions involved. Any taking of quantitative data, in biology, includes an admitted or unknown bias, except when followed by a very detailed analysis of the nature of variation in each case. The situation becomes worse still, when one is dealing with qualitative data, which must be transformed into quantitative ones. If variation is discontinuous, what values should be attributed to each case? If characters are mutually exclusive, the only way would be to apply for each manifestation the alternative of "presence/absence", as for instance in the case of various pigments, the contrast: such a pigment/no pigment, which transforms in single character with multiple variations in multiple characters with two alternatives each. If qualitative character variations are of a continuous nature, such as color intensity, it would be necessary to prove that the size of classes are strictly equal. Thus, in general, and in the majority of cases the

treatment of any kind of quantitative data includes a personal choice of a definite scale, and not an objective one.

The requirement of not choosing ~~the~~ characters, but of taking into consideration "all characters without selection" cannot be fulfilled in practice, since there are no means of deciding what may be a "unit character". If two taxons differ in size of all organs, any one character may be sufficient to measure such a difference, and the inclusion of measurements in all organs would introduce an extreme bias in favor of the size difference. In measuring one special organ should the unit character be its total size or the additive size of all discernible parts? In special cases, the use of "all characters" may be a totally wrong approach. If we want to decide, if a number of taxons (species) belong to one or another higher taxon (genus), the inclusion of species-differences may obscure the value of characters characteristic for each genus. On the other hand, if emphasis lies on differences of taxons (species within a genus), the inclusion of all common characters of the genus or of higher taxon will obscure the effects of specific differences. We thus must conclude that the acceptance of any data referring to any character must include some kind of judgement, objective or intuitive.

Secondly, the following must be stated. The purpose of procedure consists to unite all data into single values, which represent then statistical parameters, and two such parameters have been in use: the coefficient r of association and the coefficient D of MAHALANOBIS. It is hard to understand, why Numerical Taxonomy, after calculating such parameters, abandons all objective procedure and uses them to draw wholly subjective conclusions!

These parameters, calculated from experimentally obtained data, are estimates of some hypothetical ideal values, and thus subject to chance variations. Statistical analysis offers methods to estimate the precision of such values, allowing to indicate, which values are probably different from zero and thus in fact existent, and which may be different from others.

The coefficient r of association, similar to the one of correlation, suffers from the fact that its distribution is abnormal, and normalcy may be reestablished by R.A. FISHERS $z(r)$ transformation. But even so, the precision is extremely low, and any difference between two estimates of r , even when based on one hundred character pairs, below the value 0.2734, within the general limites of zero and plus one, may be solely due to chance.

The coefficient D of MAHALANOBIS gives more precision, even when there are only some ten character differences, and allows a number of tests. These are based on either of two interpretations of D : (a) as a quotient between two errors, one a balanced mean estimate of character differences and the other of all residual variances and covariances within taxons, or (b) one mean balanced estimate of standardized character differences. Tests for distinguishing an estimate from the value zero or for judging the contribution to the total D -value, by any individual character difference, have been mentioned in literature. A test to compare D -values for different pairs of taxons is proposed in the present paper, and is based on the assumption, commonly accepted in analysis of variances, that all differences, which may exist, refer to the "between" error, while all errors "within" or covariances are considered as uniform and of no significance, an assumption which, though acceptable, may be right or wrong.

Dir:
 As. Gen.
 [Handwritten signature]

EVOLUÇÃO FILOGENÉTICA E OS PRINCÍPIOS DA TAXONOMIA NUMÉRICA MODERNA

F. G. BRIEGER

1. Considerações Gerais

O aspecto mais importante do trabalho, realizado por LINNÉ ao idealizar a taxonomia como ciência moderna, era o reconhecimento do seguinte princípio: a existência de unidades taxonômicas, que hoje denominamos de forma generalizada taxons, e que se caracterizam por duas propriedades: os seus membros são idênticos ou pelo menos muito semelhantes num certo número de caracteres, os quais assim caracterizam e definem o conceito do taxon, e divergem, com parte dos seus caracteres, dos membros de todos outros taxons da mesma categoria, e esta constatação inclui uma afirmativa negativa, a do taxon não possuir caracteres típicos de outros taxons, o qual completa a definição do conceito de taxon em consideração. Com esta formulação torna-se evidente que trata-se de questões de variabilidade: a existência de uma variação contínua e de amplitude restrita para os caracteres dentro do taxon, e uma variação descontínua a respeito dos caracteres de outros taxons.

Desde que os dados, referentes a caracteres, podem ser expressos na forma de valores quantitativos, em números, toda esta questão, de uma variação ou contínua ou descontínua, poderá ser submetida a análise apropriada estatística. Este raciocínio é um dos fundamentos básicos da Taxonomia Numérica, que justamente pretende chegar a uma avaliação, pelo menos, aparentemente, mais objetiva e exata de que aquela usualmente empregada na Taxonomia Clássica. Com isso a questão do método reveste-se de grande importância também para o geneticista que deseja estudar a questão da evolução filogenética de taxons e assim precisa de uma definição clara dos mesmos.

LINNÉ reconhece a existência de um taxon básico, dentro do qual não ocorre mais variação descontínua, e ele denominou este nível como o da "espécie", postulando que esta unidade básica deveria ter uma existência por tratar-se de uma unidade, criada por uma ordem ou inteligência superior, que ele definiu como Deus Todo Poderoso. Existem numerosíssimos casos, onde esta tese da uniformidade dentro de uma categoria taxonômica básica está comprovada e onde não podem existir dúvidas sobre a definição do conceito da espécie, mas igualmente existem numerosos casos, onde há indicações que, mesmo dentro desta categoria então aenas aparentemente básica, existem de novo variações descontínuas tornando-se necessário acrescentar categorias adicionais como sub-espécies, variedade, forma etc., na aplicação prática da taxonomia.

LINNÉ também já percebeu que existem outras categorias, acima do nível da espécie, e assim postulou a possibilidade de definir o que ele chamou "um sistema taxonômico natural", indicando pelo adjetivo "natural" que toda esta ordem, assim definida, teria sido criada ou instituída por um ente de inteligência superior. Hoje, depois de dois séculos de trabalho podemos definir, de forma mais concreta, esta ordem natural dos taxons em várias categorias. Reconhecemos a subordinação hierárquica das categorias em níveis sucessivos. Uma explicação melhor desta situação, pode ser conseguida, explicando o resultado da análise sucessiva dos caracteres de um grupo de espécimes ou outros taxons. Começamos com os caracteres mais comuns, para verificar que todos possuem

espécimes em estudo

um ou mais caracteres em comum num nível A. Passando, dentro do grupo agora denominado A_i , para um outro característico, já não comum a todos os elementos comparados, podemos distinguir os grupos AB_1 , AB_2 , AB_3 ... Passando a analisar mais outros característicos de um tipo C, constatamos³ novos sub-grupamentos $AB_1 C_{11}$, $AB_1 C_{12}$..., $AB_2 C_{21}$, $AB_2 C_{22}$..., e assim em diante até que

chegamos ao último nível dentro do qual não podemos mais distinguir nenhuma variação descontínua em qualquer característico. Cada elemento, neste último nível, terá então os característicos $AB_i C_{ij} D_{ijk}$. Todos os elementos desta última categoria tem assim em comum os caracteres $AB_i C_{ij}$, que podem apenas possuir uma certa variação contínua, quando os caracteres D_{ijk} mostram uma variação nitidamente descontínua..

Surge a possibilidade da aplicação de testes estatísticos objetivos, desde que exista a possibilidade de avaliar, de forma quantitativa, os valores que correspondem aos sufixos i, j, k... e desde que podemos achar um parâmetro comum que mede todas estes valores simultaneamente.

Ficou assim demonstrado que uma taxonomia numérica é perfeitamente admissível, desde que se considera como comprovada a existência de uma ordem denominada "Sistema Natural" e dentro do qual há, em qualquer nível tanto uma certa variação contínua, de amplitude limitada, subordinada a um sistema de variação descontínua, com subordinação hierárquica de todas as descontinuidades. Não pode

Não pode haver dúvida que esta ordem sistemática das variações contínuas e descontínuas de fato existe e está comprovada pela ampla literatura de monografias sobre este "Sistema Natural". Para a metodologia da Taxonomia, e especialmente da Taxonomia Numérica, não importa assim como se explica a origem desta ordem e esta subordinação hierárquica de variações descontínuas. Existem ainda cientistas que defendem os princípios de uma "Morfologia Idealística" e que filosoficamente se afastaram muito pouco do ponto de vista de LINNÉ, pois aceitam a existência de um princípio superior filosófico da coordenação. Pouco importa se, se da este princípio o nome de "Entelechia", conforme a filosofia de DRIESCH, ou de DEUS Todo Poderoso, conforme procedeu LINNÉ. Todavia, a maioria dos cientistas contemporâneos aceita e considera como comprovado, que esta ordem surgiu em consequência da marcha da evolução filogenética, pela evolução e fixação filogenética sequencial dos caracteres.

Para o cientista meramente interessado em estabelecer, da melhor forma possível, o "Sistema Natural" dos taxons, base teórica de fato pouco interessa mas as suas conclusões terão interesse limitado em geral. O geneticista de outro lado interpreta uma maior semelhança de caracteres de um taxon como resultante de um certo período de evolução em comum, e para ele as descontinuidades da variação de caracteres indicam de um modo geral, o início de uma evolução divergente. Ele assim sempre empregará como material de base os taxons, conforme estabelecido dentro do "Sistema Natural", sempre perguntando-se, se este Sistema está plenamente comprovado ou precisa de ajustes. Assim, uma melhoria de métodos da classificação taxonômica tem igual interesse e importância, tanto para o taxônomo como para o geneticista.

Nos últimos tempos, o interesse do geneticista concentrou-se no estudo de que se considera como o último nível na série de taxons dentro do "Sistema Natural", isto é, no estudo da espécie, bem como do processo evolutivo da especiação. Isto é perfeitamente compreensível, por ser o único nível, onde se podem aplicar métodos experimentais de análise. Todavia, me parece totalmente errado querer atribuir a este nível, uma natureza diferente daquela atribuída

a outros taxons. Devemos lembrar que a evolução é um processo contínuo e o estudo de uma determinada espécie recente ou fóssil representa um corte arbitrário dentro deste processo, isto é a fixação de um momento, em si passageiro, mas tomado como fixo para fins do estudo. Este corte, num processo contínuo, apanha diferentes séries de evolução filogenética em diferentes fases da sua evolução. Assim, em alguns casos o que caracterizamos como espécie pode representar de fato um taxon completamente homogêneo, quando em outros casos este taxon já está iniciando uma nova fase evolutiva e assim força o cientista a aceitar a introdução de categorias sub-específicas, caracterizadas por uma certa descontinuidade de variação dentro da espécie propriamente dita. Seguindo principalmente DOBZHANSKY, muitos geneticistas mostraram-se inclinados a modificar a definição do conceito do taxon "espécie", para assim destacá-lo de outros níveis de taxons, desprezando a sua caracterização e pondo em destaque apenas um ponto: que se trata de um grupo de indivíduos, separados de todos os demais por mecanismo de isolamento reprodutivo. Contra esta mudança de definição podem ser levantados dois argumentos. Em primeiro lugar, uma alteração "a posteriori" não pode ser considerada como lícita quando se deseja explicar um fenômeno constatado "a priori". Em segundo lugar, deve-se destacar, que os componentes de um mecanismo de isolamento biológico em nada são diferentes de outras quaisquer diferenças de caracteres. Este é especialmente evidente com referência a um dos mecanismos mais comuns em plantas superiores tropicais: as diferenças da época de florescimento, que representam ao mesmo tempo, um característico distintivo de espécies e uma nítida adaptação ao ambiente, surgida como resultado da seleção natural. Deve-se lembrar que a exigência, da existência de um tal mecanismo de isolamento, já faz parte, da definição da categoria "espécie", desde que MINNÉ postula que devem existir tantas espécies como foram criados pelo Deus Todo Poderoso, isto é, desde que se exigiu não somente a existência de diferenças de caracteres, mas também a sua constância.

Convém lembrar que os mecanismos de isolamento reprodutivo se comportam da mesma forma como outros característicos distintivos, no sentido de que não são absolutamente fixos no decorrer do processo evolutivo. Assim não é de se admirar que espécies mantidas em separado, provavelmente, durante períodos geológicos de certa duração, podem sofrer uma hibridização introgressiva e misturar-se, quando mudem os fatores de seleção, como ocorre em escala crescente em regiões onde o homem ou outros agentes alteram a situação do ambiente ou destroem o equilíbrio biológico da fauna ou da flora.

A afirmativa de que a categoria "espécie" pode ser identificada objetivamente, quando de outro lado a identificação de taxons de outros níveis resulta de uma decisão meramente subjetiva e assim contestável, parece-me também muito falho. Em primeiro lugar, o que tem sido constatado de fato é a existência de descontinuidades de variação, subordinadas. É uma questão totalmente secundária, quais destas descontinuidades o taxônomo considera como mais importantes, atribuindo a elas uma denominação especial em Latim, e reconhecendo um número mais ou menos fixo de descontinuidades que assim são denominados com termos especiais. O número de descontinuidades, que apareceram e se mantiveram em cada série filogenética, é característica somente para esta série, e não tem nenhuma relação com as descontinuidades em outras séries. Assim representa apenas uma convenção conveniente, a distinção de famílias, tribos, gêneros e outras categorias intermediárias. A espécie se destaca, as vezes, como a última descontinuidade, que existe por enquanto numa série filogenética. Isto é, no momento.

Em resumo, a existência de taxons de vários níveis, separados uns dos outros pela variação descontínua, com ordem hierárquica é uma fato constatado objetivamente, quando o destaque dado a certos taxons, identificados por nomes em latim é resultado de um julgamento subjetivo e, as vezes, até arbitrário.

Achamos esta discussão necessária para pôr em evidência de que consideremos o ponto mais importante para a metodologia da taxonomia: a existência de descontinuidades da variação e a sua subordinação, em geral hierárquica.

Partindo desta base, a taxonomia clássica e a taxonomia numérica diferem apenas na metodologia. Na primeira, a natureza das descontinuidades é avaliada subjetivamente, e por um processo assim incontrolável, que podemos denominar o trabalho do computador cerebral e que costuma ser chamado a "intuição científica". O uso da "intuição científica" é considerada como um processo legítimo, mesmo nas ciências mais exatas como a Matemática, exigindo-se em geral que o resultado, estabelecido intuitivamente, seja confirmado por métodos exatos e objetivos *mais* "a posteriori". Na taxonomia esta comprovação "a posteriori" de fato é somente raramente possível, o que representa um defeito.

A Taxonomia Numérica emprega máquinas de calcular ou ^{um} Computador Eletrônico e assim o trabalho de cálculo ou de avaliação dos parâmetros numéricos de descontinuidade torna-se controlável. A metodologia antiga é relativamente rápida e envolve menores despesas em tempo e equipamento, quando a nova metodologia envolve um ônus considerável financeiro, o que pode restringir a sua aplicabilidade generalizada. Assim torna-se necessário analisar todos os passos da metodologia da Taxonomia Numérica, para verificar se ela traz de fato um aumento suficiente da objetividade científica, e que justifica plenamente o aumento de trabalho e de despesas. Para não sobrecarregar esta discussão com citações de autores, endossando este ou aquele ponto de vista, deixarei de citar a bibliografia especializada em detalhes, deixando aos interessados o recurso aos trabalhos especializados ou tratados gerais como o de SORAL (196).

Discutirei, em seguida, todos os passos do trabalho, seguido pela Taxonomia Numérica, sempre analisando o grau de objetividade de fato atingido por este processo. De outro lado, somente farei referências ao método da Taxonomia Clássica, quando se torna necessário fazer comparações, julgando do resto que estes métodos clássicos são de conhecimento geral.

2. Metodologia da Taxonomia Numérica

2-1 Desde que se exige a tomada de dados de todos os caracteres sem seleção e sem atribuir a eles peso algum, surge a questão, o que deve ser considerado como caráter unitário, para depois poder definir o que representará um peso em relação ao caráter unidade. Explicarei a questão por meio de uns exemplos.

Supomos que queremos analisar a flor de uma espécie de lírio com completa simetria radial e igualdade entre sépalos e pétalos. ^{As} Medição em um órgão, entre os seis, define a situação. A exigência de que todos os seis sépalos deveriam ser medidos atribuiria a uma só variável o peso seis. Em orquídeas, de outro lado com flores de simetria bilateral e gradiente dorso-ventral, os órgãos presentes em pares, os sépalos e pétalos laterais, são iguais. Assim, se fazemos apenas medições nos órgãos desiguais, teremos quatro séries de dados, como comprimento, largura etc. do sépalo dorsal e de um dos sépalos laterais, do labelo e um dos pétalos laterais, mas se medimos "todos os órgãos", daríamos um peso duplo aos caracteres dos sépalos e pétalos laterais. Em *Lilium* temos seis órgãos, que julgamos são todos iguais e na orquídea temos os mesmos seis órgãos, dos quais, porém, apenas quatro são desiguais. Qual será a solução que corresponderia a exigência de levar em conta todos os caracteres de todos os órgãos, sem seleção alguma: a tomada de seis, ^{de} quatro ou de uma série de dados?

Passamos para um outro exemplo. O investigador tem à sua frente diversos taxons de insetos, numa com com variação do comprimento de pernas, sem variação das proporções entre as suas articulações e outros sem variação essencial do comprimento total, mas com grande variação entre as proporções das articulações. Ele deve ou não levar em conta "tódos os caracteres", ou seja, tanto os do órgão total como os de cada articulação, apesar de que ele sabe que no primeiro caso a proporção entre as articulações é fixa e o comprimento total é a única variável, ou quando no segundo caso o oposto ocorre.

Seja como o investigador proceda, se ele age como simples registrador da "tódos os caracteres", sem aparente seleção, ou se ele age depois de uma ponderação científica a respeito de que são os caracteres unidade realmente variáveis, ele atribuirá, subconscientemente no primeiro caso e conscientemente no segundo caso, "pesos" e fará uma seleção.

2-2 Uma segunda complicação, que aparece em todos os trabalhos biológicos que envolvem medições ou outras determinações quantitativas de objetos, é representada pela questão da escala, em que as determinações quantitativas devem ser tomadas. Devemos, por exemplo, medir as dimensões de um órgão, registrando em separado, comprimento, largura e grossura, ou devemos levar em conta, em seguida, o volume, lembrando que o volume é uma função de terceiro grau em relação a dimensão dos valores dos diâmetros. Quando devemos então aplicar uma escala simples aritmética e quando uma escala logarítmica? Se o volume tem o valor $8=2^3$, variação seguirá a escala de 1 até 8, com intervalos iguais, ou os intervalos iguais somente aparecerão na escala logarítmica de zero até três vezes o logaritmo de dois?

Caso o investigador age simplesmente, sem maiores preocupações, a escala que lhe parece a mais fácil, ele corre o risco de errar, da mesma forma ou talvez mais ainda do que quando ele estuda primeiramente, qual a escala apropriada. Errando a escala em alguns caracteres, ele atribue automaticamente a eles, pesos desiguais.

Quando os caracteres são de ordem qualitativa e então têm que ser transformados em dados quantitativos, a situação torna-se ainda mais complexa, pois agora devemos estabelecer uma escala, sem base objetiva alguma. Quando há gradações, por exemplo, de intensidade da coloração, torna-se em geral mais fácil distinguir classes intermediárias do que nas colorações mais extremas. Atribuindo valores iguais numéricos a várias classes, subjetivamente separados, não se pode ter certeza se estes valores representam pesos iguais ou desiguais. A situação piorará mais ainda, quando há cores ou qualidades diversas, que de nenhum modo objetivo podem ser incluídas numa escala numérica. Aqui ^{na prática} como única solução, ^{a transformação de} dar um valor de presença ou ausência, em separado para ^{cada} cor ou qualidade, por exemplo, de zero, para ausência, e de um, para presença de cada coloração. Desta forma se dará ao caráter "coloração" um peso muito grande, igual ao número de qualidades de cores cada uma sempre com exclusão das outras.

Torna-se, desta forma, evidente, que a transformação de dados, de natureza qualitativa, em dados quantitativos, é perfeitamente possível, mas nunca de uma forma automática, mas sim de uma forma que inclui um julgamento subjetivo e implicitamente uma atribuição de pesos.

2-3 A exigência de levar sempre em consideração "tódos os caracteres sem seleção" pode as vezes ter consequências totalmente contraproducentes, conforme a finalidade do estudo científico em andamento.

Supomos que desejamos analisar as diferenças entre elementos de um determinado taxon, por exemplo, entre as espécies de um gênero, então a

inclusão dos dados referentes aos característicos do gênero, diluirão extremamente o peso dos dados referentes aos caracteres diferenciais entre as espécies e assim tornarão qualquer teste menos eficiente. Ou inversamente em casos nos quais o pesquisador quer verificar se certos elementos pertencem a um agrupamento, os valores dos caracteres que devem indicar a homogeneidade deveriam ser escolhidos para um teste estatístico com exclusão dos caracteres distintivos.

2.4 Deve-se lembrar ainda que existem numerosos caracteres sujeitos a uma grande variação fenotípica, e assim de muito pouco interesse e mesmo sem valor para qualquer tipo de classificação, como, por exemplo, frequentemente, ^{de órgãos} vegetativos em plantas superiores. A inclusão de tais caracteres pouca importância terá a ~~representada~~ inclusão do valor médio do caráter apesar da pouca precisão da sua determinação. A inclusão da sua variância, forçosamente muito grande reduzirá sensivelmente a precisão geral da análise estatística. A inclusão de caracteres extremamente variáveis seria prejudicial, sem trazer qualquer vantagem.

2.5 Em resumo, consideramos a exigência da incluir sempre todos os caracteres, sem seleção e assim sem pesos diferenciados, como inaceitável e inaplicável. Ao contrário, deve ser feita uma seleção dos caracteres a serem incluídos, eliminando-se aqueles demasiadamente variáveis, e dando preferência aqueles que podem indicar ou maior homogeneidade do material para uns casos e maior divergência em outros, conforme a natureza da hipótese de trabalho a ser testado.

3. Parâmetros estatísticos e o seu cálculo.

Sempre quando estão disponíveis um número razoavelmente grande de dados, a serem usados para comparações de qualquer natureza, torna-se necessário definir certos parâmetros estatísticos, que reúnem em um ou poucos valores todos os dados disponíveis, permitindo assim a sua avaliação global, em conjunto. Conforme me consta, foram dois os parâmetros, por enquanto propostos para uso na Taxonomia Numérica. Um é representado pelo coeficiente de correlação linear, de BRAVAIS, r , que está sendo usado, neste caso, como coeficiente de associação. O outro parâmetro é geralmente denominado a Distância de Mahalanobis, D , e darei preferência a denominação, que julgo mais correta, de coeficiente de Mahalanobis.

3-1. O coeficiente de associação. A definição mais simples deste parâmetro é: o quociente da Covariância dividida pela raiz quadrada do produto das variâncias, ou, designando os valores de m caracteres ^{indon} a serem comparados com os números 1 e 2 ~~se aplicam~~ dar a seguinte fórmula:

$$r = \frac{S^m (x_{1i} - \bar{x}_{1i}) (x_{2i} - \bar{x}_{2i})}{\sqrt{S^m (x_{1i} - \bar{x}_{1i})^2 \cdot S^m (x_{2i} - \bar{x}_{2i})^2}}$$

$$= \frac{S^m (x_{1i} \cdot x_{2i}) - (S^m x_{1i}) (S^m x_{2i}) / m}{\sqrt{S^m (x_{1i}^2) - (S^m x_{1i})^2 / m \cdot S^m (x_{2i}^2) - (S^m x_{2i})^2 / m}}$$

Para efetuar o cálculo, organizamos os dados em uma série, com-sequência rigorosamente igual dos valores dos caracteres para cada um dos taxons a serem comparados em pares. Do resto, o cálculo não oferece nenhuma complicação, e pode ser feito com o uso de máquinas de cálculo ou, com mais rapidez, de computador eletrônico.

3-2. O coeficiente de Mahalanobis. Este coeficiente difere totalmente do anterior, pois leva em conta os valores das diferenças dos valores dos caracteres, em pares, entre os dois taxons a serem comparados. O teste apropriado estatístico, para avaliar a significância de diferenças e o teste de STUDENT, pelo qual dividimos uma diferença entre estimativas entre si, ou de uma delas relativa a um valor ideal, pelo erro padrão da diferença. Este mesmo termo algébrico também pode ser interpretado de outra forma estatística. A divisão pelo erro padrão ou erros "standard", causa uma estandardização das diferenças, que perdem com isso a sua dimensão específica de centímetros, gramas etc., e representam estimativas de parâmetros, que devem variar, de acaso, em volta de um valor ideal qualquer, com erro padrão unida-de.

Para avaliar, englobadamente, numerosas diferenças, o que será objetivo da análise em conjunto de todas as diferenças entre dois taxons, podemos proceder de duas maneiras: somar os quadrados de todas as diferenças e dividir a soma por uma única estimativa da sua variância, calculadas na forma de uma variância residual, e em seguida tirar a raiz quadrada, ou podemos antecipar as divisões e somar os quadrados de todas as diferenças já estandardi-zadas e dividi-las pelo seu número, extraída depois a raiz quadrada.

$$\begin{aligned}
 & \sqrt{S_1 \frac{m (\bar{x}_1 - \bar{x}_2)^2}{s^2 \text{ dif } \bar{x}}} \\
 & s_{\text{dif}}^2 = s_x^2 \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \\
 & = \sqrt{S_1 \frac{n_1 + n_2}{n_1 + n_2 - 2} (x_{ji} - \bar{x}_j)^2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \\
 & = \sqrt{S_1 \frac{n_1 + n_2}{n_1 + n_2 - 2} (x_{ji} - \bar{x}_j)^2} \cdot \sqrt{\frac{n_1 + n_2}{(n_1 + n_2 - 2) \cdot n_1 \cdot n_2}} \\
 & = \sqrt{S_1 \frac{n_1 + n_2}{n_1 + n_2 - 2} (x_{ji} - \bar{x}_j)^2} \cdot K
 \end{aligned}$$

Os valores $n_1, n_2, \dots$, que representam os números de indivíduos analisados nos diferentes taxons podem ser diferentes, mas devem ser constantes dentro de cada taxon, isto é, todos os caracteres devem ser levados em conta de cada indivíduo, uniformemente. Assim, se comparamos o taxon 1, com os taxons 2, 3, ..., o valor n_1 sempre é o mesmo, embora que pode ser diferente de n_2, n_3 . Por isso, podemos deixar ao lado no cálculo de cada valor D , o conjunto que denominamos o fator constante K na fórmula acima. Esta omissão deve ser corrigida, sempre quando queremos submeter valores D a qualquer teste de significância estatística.

Uma outra correção, no cálculo dos valores D , é importante. Deve ser levado em conta a possibilidade de que os valores de caracteres, dentro dos indivíduos, em cada taxon, sejam correlacionados entre si. Se levamos em consideração as diferenças de três caracteres a, b e c de dois taxons 1 e 2, então estão incluídas, nos valores das diferenças $(\bar{x}_{1a} - \bar{x}_{2a})$ e $(\bar{x}_{1b} - \bar{x}_{2b})$, os efeitos

os efeitos da correlação r_{ab} . Passando para análise da terceira diferença $(\bar{x}_{1c} - \bar{x}_{2c})$, devemos levar em conta não mais as correlações internas r_{ac} e r_{bc} , mas sem as correlações parciais, eliminando os efeitos da correlação r_{ab} , isto é, $r_{ac.b}$ e $r_{bc.a}$.

O cálculo destas correlações torna o trabalho de computação muito demorado mesmo para um computador eletrônico especialmente quando se executar os cálculos individualmente para cada um dos $a, (a-1)/2$ diferenças entre um número a de taxons. Recomenda-se então calcular sumariamente os valores entre caracteres das covariâncias $(ab), (ac.b), (bc.a)$, etc., usando para isso uma única matriz média de todos os dados. Este procedimento inclui certos perigos, pois aceitamos, sem prova, a hipótese de que as únicas diferenças existentes entre os taxons, a serem comparados, residem nas diferenças entre as suas médias, caráter por caráter, mas que não há diferenças nem das variâncias por taxon e caráter, nem das covariâncias entre caracteres. Somente a execução de dois cálculos, um efetuado para cada par de taxons individualmente e outro com o uso de todos eles em conjunto, através de uma matriz média, pode demonstrar, se a hipótese mencionada era certa ou não. Mas, de novo, o trabalho da computação seria tão grande, de modo que geralmente ela não é exequível.

2-3. Devemos agora comparar as vantagens ou desvantagens de dois métodos que acabamos de explicar.

O emprego do coeficiente de associação tem grande vantagem ^{uma vez} que o seu cálculo, mesmo com máquinas calculadoras elétricas, é possível, sem o gasto de um tempo excessivo, e torna-se rápido pelo uso de um computador eletrônico. Mas as propriedades estatísticas deste coeficiente tornam necessário, que o número de dados, isto é, de caracteres que devem ser incluídos, tem que ser muito grande, mais ou menos e de preferência, da ordem de cem caracteres, para que o valor obtido de r possa ser aceito como possuindo uma razoável precisão. Outro defeito consiste no desprezo completo do atual valor das diferenças entre as médias por caráter e par de taxons. Isto pode até conduzir a situações e resultados absurdos. Supomos que comparamos espécimes, não de taxons diferentes, mas que pertencem a um único taxon. Então os valores de todos os caracteres deveriam ser iguais as respectivas médias gerais do taxon e as diferenças, que aparecem seriam casuais e insignificantes. Mas desvios acidentais, por definição, não tem direção e assim não podem estar correlacionados. Consequentemente, os coeficientes de associação dos caracteres dentro do mesmo taxon terão aproximadamente valor zero e são insignificantes. Aparece assim entre indivíduos do mesmo taxon o mesmo valor extremo zero que, por definição, indica a existência de taxons totalmente diferentes e não aparentados entre si.

No caso do coeficiente de Mahalanobis, a situação difere num ponto importante: ele pode ser aplicado com qualquer número de caracteres, pequeno ou grande, sendo que cerca de dez caracteres já representa um número razoavelmente satisfatório, em contraste com a recomendação supra, de incluir até 100 caracteres no cálculo do coeficiente de associação. Assim, o coeficiente de Mahalanobis tem uma muito maior aplicabilidade, pois são relativamente raros, pelo menos na taxonomia vegetal, os casos nos quais torna-se possível encontrar tão numerosos caracteres como exigido pelo outro método. De outro lado, o trabalho de cálculo do coeficiente de Mahalanobis é mais complexo, e exige muito mais trabalho e tempo do computador eletrônico.

4. Análise estatística dos resultados. A principal finalidade da Taxonomia Numérica consiste em substituir o método científico subjetivo, da Taxonomia Clássica, por um processo mais objetivo. Parece que a simples obtenção dos números impressiona muitos biólogos, dando-lhes a impressão de maior objetividade. Não se pode entender, porém, porque se larga o processamento dos dados a meio caminho, isto é, sem submetê-los a testes estatísticos. Parece que se esquece, que os dados, nos quais se baseia o cálculo dos parâmetros, são sujeitos a variação do acaso, e que os valores calculados representam apenas estimativas dos parâmetros, baseados num número limitado de dados. Desde o teste t de STUDENT (1917) e depois dos numerosos trabalhos de R. A. FISCHER, é possível indicar, por meio de testes, se estimativas obtidas na análise estatística, são com certa probabilidade, diferentes ou não, ou de zero ou entre si. Todos os testes visam a separação de pelo menos dois tipos de componentes de variação, uma que pode ser atribuída ao simples jogo do acaso e que em geral não tem interesse para o problema a ser resolvido, e outras atribuíveis a fontes sistemáticas de variação. Os resultados destes testes estatísticos não são de absoluta precisão, permitindo afirmações definitivas, mas indicando a probabilidade de certas hipóteses estarem certas ou erradas. A omissão destes testes é ainda mais inexplicável, levando em conta que o trabalho da sua execução é insignificante, em comparação com o que se gasta no cálculo dos parâmetros em questão.

4-1. Testes do coeficiente de associação. A análise do coeficiente de correlação r era praticamente impossível, devido as suas propriedades de variação anômala em consequência da sua inclusão entre $(\pm)$ limites absolutos com valores de (-1) e de $(+1)$. A transformação $z(r)$ de R. A. Fisher, igual a metade do logaritmo natural do quociente $(1+r)/(1-r)$ remove estes limites, que ficam agora iguais a menos infinito e a mais infinito, permitindo ainda de considerar a distribuição do termo $z(r)$ como normal, com um erro standard igual a raiz quadrada de $1/(N-3)$, em volta de qualquer média entre mais e menos infinito. Esta mesma transformação poderá ser usada na análise do valor r no caso presente quando ele é usado como coeficiente de associação. A única diferença reside no fato que os limites iguais a zero e mais um sem transformação e a zero e mais infinito depois da transformação $z(r)$. O erro padrão continua ser o mesmo, igual a raiz quadrada de $1/(N-3)$, sendo N o número de pares de valores usados no cálculo do valor de r .

A falta pronunciada de precisão das estimativas do coeficiente r pode ser ilustrada pelos seguintes números: Se o valor ideal for igual a zero, podem resultar, dentro do limite de 1% de probabilidade e com um número de aproximadamente cem caracteres analisados, estimativas entre os limites de mais e menos 0,25. Se os valores ideais de r foram diferentes de zero, os seguintes limites de significância podem ser previstos: para o valor de $r = 0,25$ os limites zero e 0,48; para $r = 0,50$ os limites de 0,27 e 0,67, e para $r = 0,75$ os limites de 0,63 e 0,84. Uma estimativa do coeficiente de associação com valor

aproximado da ordem de $\hat{r} = 0,50$, pode então representar uma estimativa com valor ideal igual a $r = 0,25$, $r = 0,50$ ou de um valor pouco menor do que $r = 0,75$. Calculando não os limites das estimativas em volta de valores ideais mas sim as diferenças mínimas significativas entre duas estimativas de $\hat{r}$, com 1% de probabilidade e aproximadamente com caracteres, encontramos um valor igual a 0,37, isto é, quaisquer diferenças menores do que este valor, podem ser insignificantes. Estes poucos dados mostram que, depois do grande trabalho de registrar os dados referentes a cem caracteres, para contrastar dois taxons e depois do cálculo do coeficiente de associação $\hat{r}^0$, chegamos a resultados muito pouco precisos e assim, pelo menos ao meu ver, pouco compensadores.

Conforme conheço, pelo menos a literatura, nenhum autor, empregando este método na Taxonomia Numérica, aplicou qualquer teste, e recorrendo a interpretações totalmente subjetivas dos valores $\hat{r}$ obtidos, sem levar em conta de modo algum que estes fatos são estimativas, caracterizadas assim por uma variação de acaso inevitável com a sua amplitude, de variação casual considerável, como explicamos em cima. Em vez de admitir um julgamento subjetivo ou intuitivo desde o início, baseado na apreciação direta do material e dos dados, gasta-se muito trabalho desnecessariamente, pois no fim emprega-se simplesmente de novo a intuição científica.

4-2 - Testes do coeficiente de Mahalanobis. Conforme me consta, também neste caso os autores que aplicaram este método na Taxonomia Numérica, deixaram de aplicar testes, apesar de que eles já foram expostos na literatura, no que diz respeito aos dois primeiros, abaixo apresentados, sendo que o terceiro representa uma sugestão nova.

A possibilidade de aplicar testes de significância resulta da interpretação que podemos dar ao coeficiente de MAHALANOBIS. Deixando de lado os diferentes ajustes feitos durante o seu cálculo e explicado acima, como a omissão do fator e as exceções pelas correlações, ele representa a raiz quadrada de uma média balanceada das diferenças entre valores de pares de característicos entre dois taxons, divididos por uma estimativa do erro residual da sua variação dentro de cada taxon. Trata-se assim do quociente de dois erros. O denominador tendo o grau de liberdade m , no caso de m caracteres e a estimativa do erro residual um total de $m \cdot (n_1 + n_2 - 2)$, se em cada taxon foram estudados m caracteres, em n_1 e n_2 indivíduos, respectivamente. Podemos considerar o valor como um quociente com distribuição de F de BRIEGER, ou o seu quadrado com distribuição de F de SNEDECOR. De outro lado, levando em conta a divisão das diferenças pelos resp. erros residuais, isto é, processo de estandardização, o quociente representa uma estimativa média de diferenças estandardizadas, com erro, unidade e grau de liberdade igual ao número de diferenças reunidas, ou seja, m . Neste caso pode-se aplicar o teste χ^2 .

(a) **Diferença do valor zero.** A significância do valor do coeficiente de MAHALANOBIS, considerado como um quociente de dois erros, resp. o seu quadrado como quociente de duas variâncias, pode ser facilmente verificado pela comparação direta com os valores competentes de tábuas de BRIEGER, SNEDECOR, FISCHER etc.

Se o valor for insignificante, podemos concluir que não há diferenças de qualquer importância entre os dois taxons comparados.

(b) Avaliação da contribuição de cada caráter. Se calculamos o valor do coeficiente de MAHALANOBIS, tanto para todos os m caracteres, como para $(m - 1)$ caracteres, obtemos assim um valor D_m , resp. vários valores de D_{m-1} . É lícito comparar estes valores depois de reintroduzidos os valores de K , antes deixados de lado para simplificar o cálculo. Consideramos então cada valor de D como uma estimativa de um erro de m , resp. $(m-1)$ diferenças estandarizadas, e comparamos o valor do quociente de dois deles com valores limites de teta, com m e $(m-1)$ graus de liberdade, resp. de F .

Caso que calculamos, além do valor D_m , todos os m valores D_{m-1} , obteremos uma completa análise da importância da cada um dos caracteres. Evidentemente, todos os caracteres, cuja inclusão não altera significativamente o valor de D_m , representam caracteres sem grande valor para a comparação dos dois taxons, e assim poderemos determinar objetivamente quais os caracteres que realmente distinguem dois taxons.

(c) Comparação dos quocientes de MAHALANOBIS de dois diferentes de taxons. Se analisamos, em conjunto, um número r de qualquer de taxons, podemos considerar de novo os valores D , como valores médias ponderadas de diferenças estandarizadas, todos com m graus de liberdade, se este for o número de caracteres usados para todos os taxons, torna-se então lícito a próxima comparação dos seus quocientes com os limites da tábuas de teta (resp. dos seus quadrados com as de F). Neste teste, diferente dos precedentes, desprezamos as desigualdades de correlação dos erros residuais e correlações que podiam ocorrer. Agimos, porém, analogamente ao processo adotado em todos os testes de decomposição de variâncias, nos quais reunimos numerosos componentes residuais em uma única componente, sem teste justificar este confundimento. Desprezamos a possibilidade de que, na estandarização das diferenças, o erro residual de umas diferenças entre um par de taxons possa ser diferente daquela do outro par de taxons, embora que se trate sempre do mesmo caráter. Também não levamos em conta que as correlações entre caracteres poderiam ser diferentes nos diferentes pares de taxons a serem contrastados. Como na decomposição de variâncias, postulamos que qualquer diferença significativa, que pode ocorrer, sempre estaria encontrada nos componentes "entre" a nunca nos componentes residuais "dentro".

A aplicação dos testes sugeridos sob (a) (diferença em relação a zero) e sob (c) (diferença entre coeficientes de diferentes pares de taxons), são de uma aplicação muito fácil, exigindo tão pouco cálculo, que não se pode entender porque eles são desprezados. Eles são indispensáveis, dentro da linha da metodologia da Taxonomia Numérica, pois somente eles permitem uma avaliação objetiva dos valores D dentro dos conceitos da Estatística Analítica. O teste, sugerido sob (b), envolve considerável volume de cálculo, de modo que pode-se admitir que o resultado, eventualmente a ser obtido, nem sempre justifica a despesa e o trabalho requerido.

A precisão dos testes cresce, também neste caso, com o aumento do número m de caracteres em estudo, mas deve-se lembrar que podemos aplicar o processamento de análise idealizado por R. A. FISCHER e que tinha por fim, permitir o uso de dados não muito numerosos, isto é, do que se chama de amostras pequenas. Não surge a necessidade, mencionada como necessário no caso da análise do coeficiente r , de inclusão de um número muito grande de caracteres, da ordem de uma centena. Os testes de teta, F ou X^2 já são perfeitamente eficientes quando usamos uma dezena de caracteres, embora que evidentemente a eficiência dos testes sempre cresce com o aumento do número de ados, neste caso de caracteres a serem analisados.

5. Modelos e demonstrações ~~gráficas~~. O que denominei neste trabalho o "coeficiente de MAHALANOBIS", é conhecido na literatura em geral, como a "distância de MAHALANOBIS", pois foi neste sentido que o mesmo tem sido introduzido inicialmente na análise de aglomerados ("cluster analysis"). Foram aplicados duas modalidades de representações, uma de um modelo multidimensional e outro ^{na forma de} uma representação gráfica num plano.

5-1. Representação multidimensional. A técnica será explicada em primeiro lugar. Começa-se com o valor D_{12} entre dois taxons, que é o menor ou pelo menos um dos menores entre todos os valores, e se fixa no espaço, a sua posição, por exemplo, por meio de uma arame de comprimento, em centímetros igual ao valor D_{12} . Assim está localizado no espaço a posição que cabe aos dois primeiros taxons. Agora, com o uso de dois outros pedaços de arame de comprimento, em cm, iguais aos valores D_{13} e D_{23} , referentes a um terceiro taxon, determina-se a posição relativa do taxon 3 e assim em diante, sempre dando preferência ao emprêgo dos valores menores de D . Certamente começam a aparecer no decurso deste processo dificuldades, em escala crescente, pois torna-se impossível reunir num único ponto todos os arames, referentes a um taxon e que devem ligar o novo ponto com os outros já fixados, tornando necessário fazer "ajustes". As incongruências podem ser explicadas do modo seguinte.

É perfeitamente permissível dar a um valor D a interpretação de que ele corresponde a uma distância, mas não há uma razão convincente para postular "a priori", que o valor corresponda sempre a menor possível entre dois pontos, de modo que ~~se~~ pode ser representada por uma linha reta. No caso da fixação de três pontos, representando três taxons, ficou ainda menos compreensível, por que os três valores D devem agora formar os lados de um triângulo, dentro de um só plano. Com outras palavras, embora que sempre será possível colocar um plano que inclua três pontos no espaço, não encontrei razão para exigir que os lados deste triângulo sempre devem ser iguais aos valores D , o que exige que a soma dos dois menores valores de D deverá ser sempre maior do que o terceiro valor. Todavia, embora não ter encontrado uma explicação, a situação é esta que resulta. Talvez as complicações, que surgem no feito do modelo deve-se ao fato de que algumas distâncias D não devem ser medidas em linha reta, mas em linhas curvas.

O segundo ponto a mencionar é o seguinte: admitimos que exista uma justificativa dos valores de D entre três taxons resultaram sempre na formação de um triângulo plano. ^{As} estimativas de D , obtidas por cálculo, não são exatas, mas variam com referência a um valor D ideal, que seria obtido se fosse possível incluir, na análise, todos os indivíduos existentes de cada taxon. As estimativas dos valores D , sendo apenas aproximadas, não seguiriam rigorosamente qualquer regra, talvez existente, como de formação de um triângulo plano entre 3 pontos.

Não pode haver a mínima dúvida, que a construção dos modelos permite uma visão, e consequentemente uma interpretação mais fácil das relações entre os taxons em estudo. Eles indicam de melhor forma, quais os taxons mais aproximados e aparentados e que formam aglomerados dentro da constelação geral, quais destes aglomerados são mais próximos ou mais afastados de outros, etc.

Quero mencionar um ponto que me parece interessante. Aplicando o processo dos modelos multidimensional num caso concreto, da com-
que

paração de seis grupos de taxons de orquídeas e tendo feito o cálculo dos valores de $\bar{D}$, usando primeiramente todos os caracteres, em número de 8, e depois eliminando cada vez um caráter. Observei um resultado muito inesperado. Cinco dos taxons mantiveram, de um modo geral, a sua posição relativa no espaço, mas o sexto ocupou posições extremamente variáveis, conforme os caracteres usados. O uso dos valores $\bar{D}$ como medidas da diversificação evolutiva, na filogenia de taxons, baseia-se na hipótese de que os taxons tinham uma parte da sua evolução em comum, durante a qual o valor $\bar{D}$ deve ter sido praticamente igual a zero. Iniciando uma evolução divergente, cresce o valor $\bar{D}$ de uma tal forma que serve para medir o progresso da diversificação evolutiva. Neste caso a exclusão de um ou outro caráter não deveria causar alterações substanciais. De outro lado, um taxon, com evolução totalmente independente de um agrupamento, poderá, por evolução paralela ou mesmo convergente, aproximar-se a um ou outro taxon, em determinados caracteres, ficando, porém, totalmente afastados, quando desprezamos este caráter, que indica apenas uma evolução paralela, independente. O cálculo de $\bar{D}$, com a inclusão ou exclusão de certos caracteres, dará valores erráticos. Desta forma, o cálculo de $\bar{D}$ com a inclusão e exclusão sistemática de caracteres, poderá servir para indicar, se um taxon faz parte de um conjunto de pontos ou se a sua evolução ocorreu em separado.

4-2. Representação bidimensional na forma de árvores filogenéticas. Segundo uma técnica, já bastante antiga, representam-se as relações filogenéticas de taxons na forma de uma árvore filogenética. Neste esquema, se representa as fases de uma evolução comum de taxons, colocando-os num ramo comum da árvore, que depois se divide e em geral bifurca, para indicar que, deste ponto em diante, os dois taxons seguiram uma evolução divergente. O comprimento da parte comum e a dos ramos, depois da bifurcação, são escolhidos para indicar o grau de diversificação. Assim quando a semelhança entre os taxons é grande, supõe-se que a parte comum deverá ser mais longa e a parte depois da bifurcação relativamente mais curta. Simultaneamente, BRIEGER (1933) e CAVALO SFORZA e FINNEY, em 1937 sugerem de usar os valores $\bar{D}$ para indicar distâncias dentro da árvore filogenética. Usarei, como exemplo, na comparação de três taxons, indicados pelos números 1, 2, e 3. Foram achados os seguintes valores: $D_{12} = 2$; $D_{13} = 8$ e $D_{23} = 7$.

Então marcamos numa linha horizontal, com distâncias iguais, pontos que simbolizam os taxons na sequência 1-2-3. Em seguida fixamos o ponto da primeira bifurcação, entre 1 e 2, a três centímetros da distância vertical do meio da linha 1-2. Em seguida marcamos a distância da bifurcação entre o grupo (1-2) e o taxon 3 a uma distância de 7,5 cm, aceitando a simples média aritmética entre 7 e 8 como uma estimativa razoável. Ligamos então primeiramente o ponto de bifurcação (1, 2) com os pontos 1 e 2, e em seguida ligamos o ponto de bifurcação (1, 2) tanto com o ponto 3 como com o ponto de bifurcação (1, 2), e um ramo da árvore filogenética está desenhada. Seguemos ser mais exatos, poderemos calcular um novo valor $\bar{D}_{12,3}$, reunindo primeiro, pelo cálculo de médias, os valores dos caracteres dos dois taxons 1 e 2. Em geral este cálculo, por ser muito trabalhoso, pode ser substituído pelo cálculo, embora muito aproximado, de uma média aritmética direta dos valores D_{13} e D_{23} .

Entre os dois processos *de representação* discutidos, o segundo tem a vantagem da sua fácil reprodução gráfica em publicações, quando uma reprodução por meio de uma projeção, para um plano apenas, se torna bastante difícil no caso de modelos multidimensionais.

6. Considerações Finais. Os defensores da Taxonomia Numérica, como SOKAL e outros, definem como objetivo principal dos novos métodos a substituição do julgamento intuitivo, ou se quiser subjetivo, do cientista, que usa apenas o que denominarei: o seu computador cerebral, por um processo, considerado mais objetivo e assim considerado como mais exato e mais satisfatório. Da discussão desta nova metodologia quero, ao fim, destacar dois pontos:

Em primeiro lugar, na obtenção de dados é totalmente impossível aplicar processos quase que mecânicos e totalmente objetivos. A começar, a exigência da análise de "todos os caracteres sem seleção alguma" representa não somente um trabalho muito excessivo, mas pode até ser prejudicial, por exemplo, quando incluímos todos os caracteres que não mostram diferenças, em comparações de táxons, a respeito das quais queremos saber se são diferentes. *é uma verdadeira utopia esperar poder eliminar qualquer subjetividade na escolha da técnica na qual tomamos os dados numéricos referentes aos caracteres, ou por (também) medição direta, ou pela transformação de dados qualitativos em quantitativos.*

Em segundo lugar, obtidos os parâmetros estatísticos finais, seja o coeficiente de associação r ou o coeficiente de MAHALANOBIS D , não se deve esquecer que o que temos calculado, a partir dos dados originais, são apenas estimativas, de maior ou menor precisão e sujeitos as imprecisões causadas pela variação de acaso já na tomada dos dados originais. Uma vez que a execução dos testes estatísticos de significância, exigem um volume relativamente pequeno de cálculo adicional, não se entende porque os testes têm sido totalmente omitidos pelos representantes da Taxonomia Numérica, a julgar menos pela bibliografia que consulto. Procedendo assim e omitindo os testes, surge uma situação muito curiosa. Depois de amplo e exaustivo trabalho, feito para introduzir uma objetividade máxima científica, se abandona toda a objetividade no fim do trabalho, aceitando uma avaliação totalmente subjetiva de dados que representam apenas estimativas de parâmetros, atribuindo a eles uma precisão absoluta, que eles não podem possuir.

Comparando os dois métodos, do ponto de vista da eficiência estatística dos respectivos testes, surge a conclusão de que o coeficiente de associação, embora de computação muito mais fácil, é pouco satisfatório do ponto de vista da análise estatística, de modo que parece-me muito preferível a aplicação do coeficiente de MAHALANOBIS, apesar do volume muito maior de cálculo que a sua computação exige.

O método de MAHALANOBIS tem ainda a grande vantagem sobre o coeficiente de associação, pois pode ser usada como distância (a) em montagens multidimensionais de modelos, indicando a posição dos táxons em conjunto, ou (b) na forma tradicional de árvores filogenéticas.

Finalmente, deve-se apreciar em que extensão esta nova forma de trabalho taxonômico poderá, ou não deverá substituir os métodos tradicionais. Evidentemente, devido a complexidade dos cálculos, o método numérico somente se recomenda para trabalhos monográficos e assim bastante especializados, nos quais o método clássico não oferece resultados satisfatórios. Considero a metodologia da Taxonomia Numérica não como de aplicação generalizada, mas sim com um processo ~~que~~ recomendável em casos especiais. Em numerosos casos, porém, o método clássico continuará o único aplicável, e que dá, em geral, resultados suficientemente compensadores.

Como por exemplo a opção de um grupo. Muito grande, quando queremos mostrar quais espécies pertencem a um e não a outro, então ficamos a vontade dos diferentes, no nível das espécies, podendo assim

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

April 16, 1968

Dr. David J. Rogers
Department of Botany
University of Colorado
Boulder, Colorado

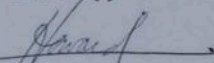
Dear Dave:

Thanks for your letter to Brasilia, but unfortunately the material you wanted was not available. I don't remember the details, and I don't have your letter here to refresh my memory. Despite the usual round of problems, everything came off pretty well, including our collecting a goodly number of Manihot species. I expect everything to arrive here in June.

Meanwhile Dr. F. G. Brieger, former head of the Instituto Central de Biologia of the Universidade de Brasília will be here the end of May to give a seminar on Numerical Taxonomy. He has written a paper on the subject (in press somewhere), and a xerox of his typescript is on its way to you under separate cover. If you or any of your people are intrigued, we'd be happy to see you.

Please let your herbarium curator have the enclosed form and convey our thanks for his attention to the return of this loan.

Cordially,



Howard S. Irwin
Head Curator

HSI:mi
Enc.

HERNER
COMPANY
Progress through effective use of information

May 23, 1968

D. J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101A
University of Colorado
Boulder, Colorado 80502

Dear Mr. Rogers:

Thank you for your response to our questionnaire regarding Current Research & Development in Scientific Documentation No. 15.

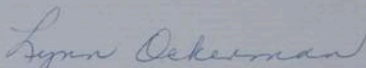
We note that you had a project reported as continuing in the previous issue of this series, CRDSD No. 14 when you were at the New York Botanical Garden. We must account for the status of that project in No. 15. If the project has been completed or discontinued please fill out Part A of the questionnaire and give the completion date. If any reports or publications describing the project have been issued please fill out Part J as well. If the project is continuing and the description in No. 14 still applies, please so indicate.

Since you are no longer at the New York Botanical Garden, you may not be able to fill this out. In that case, could you please forward it to someone there who can furnish us with the current status of that project.

We would appreciate your prompt attention in this matter, as we are nearing our final deadline.

Thank you.

Sincerely yours,


(Miss) Lynn Ockerman

and then groups of highly similar words are found. Nearly 13,000 documents (titles and abstracts in the fields of electronics and radio communications) are stored on magnetic tape. The 13,000 documents are indexed in terms of the groups (of similar words) used as descriptors, and a print-out of documents is produced in answer to each inquiry. A set of realistic enquiries will be processed, and the results will be evaluated subjectively.

Progress to date: 1,536 abstracts have been indexed in terms of one set of descriptor groups. Trial retrievals have begun.

Current and future plans and activities: Careful evaluation will now proceed, together with the generation of new sets of descriptors by different methods. At the same time, modifications to the system will be made in the light of the initial retrieval results, e.g., it seems likely that retrieval quality might be improved by permitting some weighting of the individual words composing a request.

Materials produced during the project: 13,000 radio abstracts on magnetic tape, 8 hole paper tape, and ICT cards.

Equipment used: English Electric KDF9 Computer and peripheral equipment (most work to date programmed on the ACE Computer).

Project dates: 1961-1966.

References:

- (1) Meetham, A. R. "Probabilistic Pairs and Groups of Words in a Text," *Language and Speech*, vol. 7, part 2, 1964, pp. 98-106.
- (2) Vaswani, P. K. T. "Mechanized Storage and Retrieval of Information," *Revue Internationale de Documentation*, vol. 32, No. 1, 1965, pp. 19-22.

5.65

NEW YORK BOTANICAL GARDEN

Bronx Park, Bronx, N.Y. 10458

Principal investigator: DAVID J. ROGERS.

Project title: THE USE OF COMPUTERS IN CLASSIFICATION.

Objective: (a) To provide computerized mathematical models for the purposes of general classification (cluster analysis). The major emphasis is for classification of biological entities in a hierarchical manner, with maximum information concerning each of the classified entities and the amount of connectedness to other entities. The classification should be useful to those interested in any type of classification problem, be it taxonomic, ecologic, or other. (b) To develop computerized models for the determination of the value of defining data to be used in the process of classification. Involved in this project are several levels: how to recognize patterns (or characters); how to optimally divide patterns into discrete states to define the objects under study; how to determine the value of each pattern as it affects the classification; how to deal with the situation in which not all desirable information is available for an object of importance in the study. (c) To develop retrieval systems to make the classifications useful for various purposes: correlations for various users; identification of unknowns; entry to the classification from any point.

Scope, methodology, and approach: Two cluster analysis methods have been

completely programmed and are available. The first, using an inhomogeneity factor, is non-hierarchical and is particularly useful in determining, within clusters, the amount of variability, "noise," tightness of the cluster, etc. This model assumes that all clusters are essentially spherical. The second cluster analysis, based on graph theory, gives a hierarchical classification and is more appropriate to biological taxonomy in that clusters may assume any shape. Both clustering methods start with a developed pair-wise measure of similarity between all the objects in the study. The model for the determination of the value of input data in the clustering analysis (character correlation) is based on information theory. This model is intended to be a complement to the clustering analysis method. In any classification scheme, the classification is only as good as the descriptive data used to make the classification. We must know whether the input data will indeed be useful in constructing clusters. The method will indicate the value of the states within a character, give the value of the character over the number of objects it will define, and give a measure of predictive ability for each character in relation to other characters.

Progress to date: Two cluster analyses are immediately available. The inhomogeneity clustering method is described in (2). In the second clustering method, a descriptive paper has been submitted for publication. The graph theory clustering method has been tested on a complex group of tropical legumes of the genus *Cassia*, and on the genus *Manihot* (Euphorbiaceae). In both cases, the results have been extremely valuable in that a classification was produced that could not have been achieved otherwise. The classification, based on a large number of descriptive states (over 400 in the case of *Cassia*, 250 for *Manihot*), is by far the most objective classification of a group of plants yet accomplished.

Current and future plans and activities: Studies in pattern recognition. Botanical and zoological characters are "patterns." Symptoms of diseases are "patterns." It seems reasonable to use the methods used by taxonomists to establish characters as a take-off point to study the basic problem of how patterns are recognized. As many different applications of models for character correlation and clustering as can be adequately handled will be run. This will be done on a collaborative basis, in which experts in various fields will be asked to provide data, and the Department of Quantitative Taxonomy will provide the expertise on best methods of handling. The areas of immediate interest are botanical, but others will also be attacked, such as various types of disease, anthropological classifications, etc.

Materials produced during the project: Two program decks (or program listings) are available to workers with similar problems who are willing to work with the group to develop models of more general and useful nature. The programs are written largely in FORTRAN II. (Some machine language included.)

Equipment used: CDC 1604 32 K Memory Computer.

Project dates: 1962-continuing.

External sources of financial support: National Institutes of Health; U.S. Navy, Office of Naval Research.

References:

- (1) Rogers, David J. "Taximetrics: New Name, Old Concept," *Brittonia*, vol. 14, No. 4, 1963, pp. 285-290.
- (2) Rogers, David J. and Fleming, H. S. "A Computer Program for Classifying Plants. II. A Numerical Handling of Non-numerical Data," *BioScience*, vol. 14, No. 9, 1964, pp. 15-28.
- (3) Rogers, David J. and Fleming, H. S. "Data Recording: an Inexpensive and Efficient Method," *AIBS Bulletin*, vol. 13, No. 2, 1963, p. 42.

5.66

NORTHAMPTON COLLEGE OF ADVANCED TECHNOLOGY

St. John Street, London, E.C.1., England

Principal investigator: J. E. L. FARRADANE.

Project title: INFORMATION RETRIEVAL BY RELATIONAL INDEXING AND NEW CLASSIFICATORY METHODS OF CONCEPT CONTROL.

Objective: To investigate the efficiency of retrieval of information (references to documents) indexed in structured form whereby terms (controlled by new classificatory methods) are linked by explicit relational signs (operators—nine in number, plus possible negation forms of them) derived from a study of the psychology of thinking, and to investigate methods of test and derive an adequate mathematical basis (statistical or otherwise) for such testing.

Scope, methodology, and approach: Material for indexing comprises abstracts of papers on sugar technology and related subjects taken from *Sugar Industry Abstracts* over several years. A new general scheme of classification is being developed in outline, and those parts needed for the present research are being worked out in detail. The classification method is to group words (terms) by type of concept under the broad headings of entities, activities, abstract terms, and properties. These groups are divided, for convenience, into a small number of subgroups, e.g., Entities into Physical, Chemical, Living and Artefacts; Activities into Physical, Living and Abstract; and Abstracts into Physical, Behavioral and Symbolic. Each group can be subdivided, in a second direction (i.e., rows, if the groups are in columns) by levels of complexity. The terms thus entered are individual things, actions, etc. Classes can be provided (as true generic terms only) in a third "dimension," considered as if at right angles to the plane of the paper carrying the two-dimensional table first constructed, and the class groupings can be illustrated, separately, as sectional elevations. A further table of heterogeneous terms, called "systems" needed to allocate terms that are combinations of the fundamental types, is also being worked out; such heterogeneous terms are, e.g., "waste," which is an entity plus a property, "planet," which is an entity plus an activity, "chemistry," which is a combination of entities, activities (reactions), abstracts (theories) and properties. The classification system thus builds up from the particular to the general. A special index, showing groupings of "parts" into "wholes" (which is different from the generic relation) will also be made. The relational symbols and their use have been fully described in several articles. Linear representations of structured information

HERNER
 COMPANY
Progress through effective use of information

March 1, 1968

Gentlemen:

The National Science Foundation has contracted with the firm of Herner and Company for the preparation of the 15th edition of Current Research and Development in Scientific Documentation (CRDSD). The publication is issued for the purpose of disseminating, on the widest possible basis, information regarding ongoing research and development projects in information science and technology and related fields. It is hoped that this edition will cover all ongoing research and development in the field of information processing, manipulation, dissemination, communication and use patterns.

If you are engaged in research or development in documentation, we would appreciate your completing the enclosed questionnaire. We are sending this Project Description Form to all those researchers whom we have reason to believe are active in this field.

The survey material consists of duplicate copies of a questionnaire (BOB Approval No. 99-R0217), one for retention in your files, one to be returned in the pre-addressed envelope to Herner and Company. If additional questionnaires are needed to report other ongoing projects, they will be dispatched immediately upon notification of Herner and Company, at the address below.

We ask your cooperation in observing the following deadlines: Questionnaires addressed to individuals or organizations in the United States and Canada should be completed and returned within 30 days of this date. Questionnaires addressed to other countries should be returned by air mail no later than 30 days after receipt of the survey material. Completed questionnaires, and any other communications regarding CRDSD should be addressed to:

Project CRDSD No. 15
Herner and Company
2431 K Street, N. W.
Washington, D. C. 20037

If you do not have any research or development to report at this time, please respond by filling in your name and organization title in Part A with a clear notation "no research."

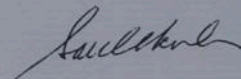
If you have previously reported an ongoing project in CRDSD and the project has now been completed, please fill out Part A of the questionnaire and give the completion date. If any reports or publications describing the project have been issued, please fill out Part J as well.

In the event that one of your projects was reported in No. 14, and the description is still applicable and up to date, please indicate "continuing" in Part A, and write into the body (B, C, & D) "the description in No. 14 still applies."

A copy of CRDSD No. 15 will be sent free to each cooperating organization.

Thank you for your prompt cooperation.

Sincerely,



Saul Herner

NATIONAL SCIENCE FOUNDATION

Current Research and Development
in Scientific Documentation No. 15

If you have more than one project to report for Current Research and Development in Scientific Documentation, No. 15, please describe each one on a separate form.

Attach extra sheets of paper if more space is required.

Mail completed form to:

Project: CRDSD No. 15
Herner and Company
2431 K Street, N.W.
Washington, D.C. 20037

PROJECT DESCRIPTION FORM

A Institution at which work is being done:

TAXIMETRICS LABORATORY
UNIVERSITY OF COLORADO

Address of institution: Taximetrics Laboratory, Dept. of Biology, U. of Colorado, Boulder, Colo.

80302

Title of project: TAXIR, an information retrieval system for biology

Person directly in charge of project: Dr. David J. Rogers

Person to whom inquiries may be addressed, if different from above:

Source of financial support:

☐ in house

☒ other sponsor, please specify: National Science Foundation

Project began: July 1967

Project scheduled to end: July 1969

B Objective or purpose of project:

To provide a computerized information retrieval system for the use of biologists, particularly systematists, and museum curators.

Over Please

C Methodology and/or approach: An information bank has two functions. 1) It serves as a repository of information, 2) it serves as a device for gaining access (by means of locating criteria) to selected portions of the information. Our efforts are concentrated in the second area. We have devised a computer algorithm based on set theory and information theory which compresses a large body of locating criteria into a small area of computer storage. Retrieval is realized by boolean arithmetic performed directly on this compressed information rather than by the usual method of searching. This technique permits 1) fast retrieval, 2) the use of locating criteria with many possible alternative conditions, 3) the use of a large number of locating criteria. A paper on the technique is in preparation.

D Results or progress to date: The design of the system is virtually complete. Computer implementation is about half finished. Plans underway for collaboration with other institutions for mounting large data banks on the system.

E Current and future plans and activities: Finish implementation and collaboration. Produce manual of instruction and program documentation.

- F Specify name, brand, and model number of all equipment used or developed in project, and, in each case, give the specific purpose for which the equipment was used.

The first version of TAXIR will run on a CDC 6400.

-
- G List any specialized dictionaries or thesauri, machine-readable texts, recordings, programming documentation, or other research materials generated in connection with project. Specify physical form (book, punched cards, magnetic tapes, etc.) and conditions of availability.

Program documentation will be produced. It is not yet available.

A manual of instruction in the use of the TAXIR system will also be produced.

-
- H Definitions of unique, specialized, or local terms used in this project description (if not defined in Glossary, pp. 567-571 of the 14th edition of *Current Research and Development in Scientific Documentation*):

Over Please

- I Definitions of abbreviations or acronyms used in this project description (if not defined in list, p. 572, of the 14th edition of *Current Research and Development in Scientific Documentation*):

- J References to publications or reports resulting from project (if available, please enclose copies of papers or reports; if copies are not available, please give complete bibliographic citation below):

not yet available

- K Copies of the publications cited above may be distributed by the National Science Foundation through the Clearinghouse for Federal Scientific and Technical Information. (Please check block below)

☐ Yes ☐ No

Authorized Signature

11 March 1968

Dr. Daniel F. Merriam, Editor
Computer Contribution Series
State Geological Survey
The University of Kansas
Lawrence, Kansas 66044

Dear Dr. Merriam:

Thank you for sending your brochure on computer programs and publications. I was interested to know that the contributions mentioned by you as being of special interest are ones for which we have our own programs.

Your idea for advertising programs^a seems a worthwhile one and in time such activity may prevent the "reinvention of the wheel" in each computer center.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm



The University of Kansas

STATE GEOLOGICAL SURVEY

LAWRENCE, KANSAS
66044

FRANK C. FOLEY
State Geologist and Director
WILLIAM W. HAMBLETON
Assoc. State Geologist and Assoc. Director

5 March 1968

Dr. D. J. Rogers
Taxometrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dr. Rogers:

From your participation in the recent Conference on Numerical Taxonomy at The University of Kansas, I assume that you are interested in computer applications. Inasmuch as it was impossible for all participants to visit the Geological Survey and to learn of our work in this area, I am taking the liberty of enclosing a list of our publications for your information.

We have several programs presently available which should be of interest to you in your research work. In particular, I would point out Computer Contributions 4, 6, 17 and 20. Of special interest to numerical taxonomists will be Computer Contribution 7, the proceedings of a "Colloquium on Classification Procedures" that was held at The University of Kansas in 1966.

If there is anything else I can do to be of help, please let me know.

Sincerely yours,

Daniel F. Merriam
Chief, Geologic Research and
Editor, Computer Contribution Series

DFM:jad

Enc.-1

Computer Applications In The Earth Sciences

Publications listed here include papers published by the Geological Survey that deal with computer applications in the earth sciences. These have been published as Special Distribution Publications, Computer Contributions, and also include reprints of articles from other journals, which are available upon request for a limited time.

As an additional service, the Geological Survey makes available the computer program for a limited time at a nominal cost. Their prices are listed herein. It would be helpful when ordering a deck to specify the type of machine because we may have several variations of one program. In this way, we can better judge which deck would be most useful to you. Unless otherwise specified, decks will be sent on small magnetic tapes with accompanying retrieval instructions. If card decks are required an extra charge of \$5.00 is made. Also, when decks are ordered the publication describing its use will be sent automatically.

For a complete list of publications, write the Geological Survey at The University of Kansas, Lawrence.

Editorial Staff

DANIEL F. MERRIAM, *Editor*

Assistant Editors

JOHN C. DAVIS OWEN T. SPITZ

Associate Editors

JOHN R. DEMPSEY
Northern Natural Gas Company

RICHARD W. FETZNER
Sun Oil Company

JAMES M. FORGOTSON, JR.
Pan American Petroleum Corporation

JOHN C. GRIFFITHS
Pennsylvania State University

JOHN W. HARBAUGH
Stanford University

RICHARD G. HETHERINGTON
University of Kansas

SIDNEY N. HOCKEY
Phillips Petroleum Company

J. EDWARD KLOVAN
University of Calgary

WILLIAM C. KRUMBEIN
Northwestern University

R. H. LIPPERT
Shell Oil Company

WILLIAM C. PEARN
Mobil Field Research Laboratory

MAX G. PITCHER
Continental Oil Company

FLOYD W. PRESTON
University of Kansas

WALTHER SCHWARZACHER
Queen's University of Belfast (UK)

PETER H. A. SNEATH
University of Leicester (UK)

COMPUTER CONTRIBUTIONS

1. Mathematical simulation of marine sedimentation with IBM 7090/7094 computers, by J. W. Harbaugh, 1966. \$1.00
2. A generalized two-dimensional regression procedure, by J. R. Dempsey, 1966. \$0.50
3. FORTRAN IV and MAP program for computation and plotting of trend surfaces for degrees 1 through 6, by Mont O'Leary, R. H. Lippert, and O. T. Spitz, 1966. \$0.75
4. FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer, by J. C. Davis and R. J. Sampson, 1966. \$0.50
5. FORTRAN IV program using double Fourier series for surface fitting of irregularly spaced data, by W. R. James, 1966. \$0.75
6. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040, by R. L. Barcher, 1966. \$1.00
7. Computer applications in the earth sciences: Colloquium on classification procedures, D. F. Merriam, editor, 1966. \$1.00
8. Prediction of the performance of a solution gas drive reservoir by Munkat's Equation, by Apolonio Baca, 1967. \$1.00
9. FORTRAN IV program for mathematical simulation of marine sedimentation with IBM 7040 or 7094 computers, by J. W. Harbaugh and W. J. Wahlstedt, 1967. \$1.00
10. Three-dimensional response surface program in FORTRAN II for the IBM 1620 computer, by R. J. Sampson and J. C. Davis, 1967. \$0.75
11. FORTRAN IV program for vector trend analyses of directional data, by W. T. Fox, 1967. \$1.00
12. Computer applications in the earth sciences: Colloquium on trend analysis, D. F. Merriam and N. C. Cooke, editors, 1967. \$1.00
13. FORTRAN IV computer programs for Markov chain experiments in geology, by W. C. Krumbein, 1967. \$1.00
14. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer, by R. D. Hobson, 1967. \$1.00
15. FORTRAN II program for progressive linear fit of surfaces on a quadratic base using an IBM 1620 computer, by A. J. Cole, C. Jordan, and D. F. Merriam, 1967. \$1.00
16. FORTRAN IV program for the GE 625 to compute the power spectrum of geological surfaces, by J. E. Esler and F. W. Preston, 1967. \$0.75
17. FORTRAN IV program for Q-mode cluster analysis of non-quantitative data using IBM 7090/7094 computers, by G. F. Bonham-Carter, 1967. \$1.00
18. Computer applications in the earth sciences: Colloquium on time-series analysis, D. F. Merriam, editor, 1967. \$1.00
19. FORTRAN II time-trend package for the IBM 1620 computer, by J. C. Davis and R. J. Sampson, 1967. \$1.00
20. Computer programs for multivariate analysis in geology, D. F. Merriam, editor, in press.

7 March 1968

Dr. Kenneth H. Zabriskie, Jr.
BioSciences Information Service
2100 Arch Street
Philadelphia, Penna. 19103

Dear Ken:

Enclosed is the paper I mentioned which makes efficient the storage of bit strings. This also gives you a little line on the rest of the program. The several aspects of programming such as scanning, queries, identification of query type, etc. are well along. We hope we can put on a demonstration this coming summer of at least some data.

Thanks for your call; we'll keep in touch.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

23 February 1968

Dr. Diego Jordano
Departamento de Zootecnia
Consejo Superior de Investigaciones Cientificas
Facultad de Veterinaria
Cordoba, Spain

Dear Dr. Jordano:

I am pleased at the interest you show in our methods.
I am sending under separate cover (surface) the documents
you request. These include

1. Fortran source list of Program Graph
2. Systems document for Program Graph
3. Reprint of "A General Method of Taxonomic Description for
a Computed Similarity Measure," BioScience, Vol. 16, #11,
1966.
4. Reprint of "A Mathematical Model in Graph Theory for Biolo-
gical Classification," J. Theo. Biol., Vol. 12, 1966.
5. ~~The~~ user's documents.
 - I. How to design the input (data) deck.
 - II. How to read the printout.

I hope the documents will prove sufficient. Please feel
encouraged to communicate with me about any further questions or
comments you may have.

Very truly yours,

George F. Estabrook

GFE:gm

26 February 1968

Dr. D. D. W. Goodall
Department of Population and Environmental Biology
University of California
Irvine, California 92664

Dear Dr. Goodall:

I enjoyed meeting and talking with you last November in Lawrence, Kansas. You may remember our discussions on information retrieval when I promised to send you a confidential preprint of a recent paper of ours. I received the list of participants for that conference this morning and was reminded of my outstanding obligation. Please find enclosed this paper. I would be pleased to hear any questions or comments you might wish to make.

Very truly yours,

George F. Estabrook

GFE:gm

29 January 1968

W. M. Johnson
Rocky Mountain Experiment Station
240 West Prospect
Fort Collins, Colorado 80521

Dear Dr. Johnson:

It was a pleasure meeting you recently. Please find enclosed a few reprints which I hope you will find instructive.

Encourage Frank and Meredith to send a programmer to Boulder soon to learn how to implement the "Skyline" feature on the Fort Collins machine.

I look forward to seeing you again soon and to discussing in more depth your very interesting problem with CAREX and the relation which our techniques might have to it.

Very truly yours,

George F. Estabrook

GFE:gm

16 January 1968

C. J. Jardine, N. Jardine, R. Gibson
Wipple Science Museum
Free School Lane
University of Cambridge
Cambridge, England

Gentlemen:

Please excuse my delay in responding to your very interesting communication. I have read your publication, "The Structure and Construction . . .", *Mathematical Biosciences*, 1 (1967) with great interest. I am not sure that the power of your arguments will be universally appreciated within the biological community for some time yet, although I find in this kind of mathematical eloquence much promise for the future of mathematics in biology.

I had the pleasure recently of speaking with Dr. Gordon Sheals of the British Museum. Dr. Sheals is familiar with your work (up to a point) and spoke highly of it. If you have not already done so, you should send a reprint of the above paper to Dr. Robert Sokal, Dept. of Entomology, University of Kansas, Lawrence, Kansas 66044, for inclusion in "Principles of Numerical Taxonomy" which he and Dr. Sneath are presently revising. I'm sure they would be most interested in your contribution.

With respect to your continuing work, I can make no constructive suggestion at present. I am, however, most interested in your efforts and I hope the quality of your recent results is a precedent for your future achievements.

My own area of interest has shifted somewhat in the past few months during which time I have been working on problems in numerical cladistics. Quite recently I have derived some interesting theoretical results which allow the direct computation of all branching sequences meeting the conditions of Camin-Sokal, *Evolution*, Vol. XIX, #17 (1965). As the result is but a few days old I do not have a manuscript prepared as yet. I expect to submit to *J. of Theoretical Biology*.

I have also become interested in problems in information retrieval for Biology. There is a remote possibility that I will attend the tentative meeting for Information Retrieval in Edinburgh this coming August. If so, perhaps I will have the pleasure of meeting some of you in person then.

If any of you be interested in continuing this exchange, I shall try to be a better correspondent.

Very truly yours,

George F. Estabrook
Research Mathematician III

GFE:gm

Taximetrics Laboratory
Department of Biology
Armory 104
University of Colorado
Boulder, Colorado 80302
29 January 1968

C. J. Jardine
Whipple Science Museum
Free School Lane
Cambridge University
Cambridge, England

Dear Sir:

My colleague G. F. Estabrook has recommended to me your paper
"The Structure and Construction of Taxonomic Hierarchies" which
appeared in Mathematical Biosciences for June 1967.

I would be most grateful if you could send me two reprints
of that paper if still available. Your work is highly regarded
here.

Yours sincerely,

Howard Christensen

HC:gm

TO: Dr. Dan Bailey
FROM: David J. Rogers
SUBJECT: Grant Classifications
23 January
DATE: 23 January 1968

Dear Dan:

Do you recall the name of the man in M.I.M.H. who wants to run a classification of grants? I'd like to get in touch with him, to try our approach on his data.

You can call me at 6712.

Thanks.

The Catholic University of America
Washington 17, D. C.

DEPARTMENT OF BIOLOGY

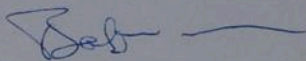
Dear Dave,

How are chances of getting ^{copies} ~~copies~~ and operating instructions ~~for~~
of your programs.

We have a bunch of data of chromosomes that we would like to commit
to your analyses. Naturally, you will be given due credit if anything
comes of it.

How are things going in Boulder?

Sincerely,



Bob Davidson

Dear Bob -

Here are the necessary
documents to get you
going for our 2 biggest
efforts -
Have fun - Dave

13 February 1968

Dr. Robert Davidson
Department of Biology
The Catholic University of America
Washington, D. C. 20017

Dear Bob,

Here are the necessary documents to get you going on
Similarity Clustering (GRAPH) and a start on Character
Analysis (CHARANAL). The flow chart will be sent along
on completion.

Have fun!

David J. Rogers
Professor of Biology

DJR:gm

P. S. Richard A. Barnes, Berss, Reitzel, and Associates, Inc.
of 134 Mt. Auburn Street, Cambridge, Mass. 02138, (phone 617-

864-4320) has successfully converted our programs and might be able
to give you some helpful hints if you have trouble.
to

Copy
to
R.H.F.

THE UNIVERSITY OF TEXAS
DEPARTMENT OF ELECTRICAL ENGINEERING
AREA CODE 512, GR 1-1881
AUSTIN, TEXAS 78712

January 31, 1968

Professor David J. Rogers
Taximetrics Lab - Armory 101A
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Professor Rogers:

A copy of the preprint, "A Computer Classification of Psychiatric Patients", which you requested is enclosed. The principal purpose of the paper is to present the results obtained from the application of our numerical classification procedure to a set of psychiatric patient data. A brief discussion of the numerical approach is also included.

Dr. Turner and I have recently written a more detailed account of our procedure for publication, and the manuscript is currently being revised. When these revisions are completed, I will forward a copy to you.

Sincerely yours,

R. H. Flake

Robert H. Flake
Assistant Professor

RHF:rpf

Enclosures

2 January 1968

Dr. David W. Goodall
Division of Biological Sciences
University of California
Irvine, California 92664

Dear David:

Thank you for sending the program for computer identification. I hope to be able to get to it, and understand it soon.

Best wishes for the New Year.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

8 January 1968

Dr. Arturo Gomez-Pompa
Instituto de Biología
Universidad Nacional Autónoma de México
México 20, D. F.

Dear Arturo:

First let me congratulate you on a really fine meeting. We were all very pleased by the arrangements for hospitality and the general scientific level of the proceedings. I hope this is the first of many such fine gatherings.

I do hope that you will be able to find time for Sergio Ahumada to come here to see us. I was serious about paying for his travel funds; I think we can afford it. I would recommend, if travel arrangements can be made, that he spend a week with us and I think it would be better if he could come sometime in the months of April or May. I will be glad to have him stay with me and thereby reduce his living costs. It is my firm conviction that we can contribute considerably to your efforts for your Information retrieval system and I hope that you will take advantage of our offer.

Once again let me thank you for your great efforts on our behalf at the symposium.

Very sincerely yours,

David J. Rogers

DJR:gm

10 January 1968

Dr. Mario Gutierrez G.
CIMMYT
Londres 40, 3er Piso
Mexico 6, D. F.

Dear Mario,

It was indeed a pleasure to see you again in Mexico. I have recently written a letter to Arturo Gomez-Pompa recommending strongly that he send us Sergio Ahumada from the computing center of the University. The purpose of sending him, of course, is to familiarize him with our techniques so that he can be of service to you and Gomez-Pompa. I have agreed to pay the bill for his travel.

In the meantime I have sent reprints to you and Dr. Gil. I would suggest very strongly that you pay particular attention to one by Rogers and Estabrook, entitled "A General Method of Taxonomic Description for a Computed Similarity Measure." While that paper pays attention to the structure of data for a classification program, very similar attitudes can be adopted for structuring data about your germ plasm bank to be used in an information retrieval program. It would be very helpful if you could let me have a sample of the data which you wish to put into an I/R system. With this we could advise you on the proper construction of these data. Then with Ahumado's assistance you could go ahead and develop the system necessary for your own operation. We will be very pleased to work with you and give you the benefit of our thinking along these lines. I look forward to cooperation with you and hope we can be of service to you in this type of endeavor.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm