



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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

December 14, 1966

Mr. David L. Largent
Botany Department
University of Washington
Seattle, Washington 98105

Dear David:

Please find enclosed a reprint from each of Science 132(3434) and Systematic Zoology 15(1): 59-69. We have not received from the publishers reprints of the other publications yet. Let me suggest that the Science article is of little more than historical interest as the techniques described therein have been replaced by those described in the more recent Systematic Zoology publication, to which the current BioScience article may be considered a sequel.

You have expressed some desire to use our techniques on a classification problem in the genus *Rhodophyllus*. We would strongly encourage you to work with us in applying our techniques to your problem. There are many reasons for this. Among them:

1. The field of computer applications in taxonomy is young and experimental. The methods are not "cut and dried" and in fairness to your efforts and our methods we would like to work with you to ensure that you develop an understanding of what the method is, what you as a biologist must do to prepare data, and how the results of a computer analysis can be used by you to help solve your classification problems.
2. We are set up here with all the programs and computing machinery necessary for the analysis, as well as the personnel trained to operate the machinery properly. In general, it has proven unexpected trouble to run rather sophisticated programs such as the ones you mention on different machines and by operators who don't have a thorough understanding of what technically is to be done. If you are to prepare your data properly, it will be a fair amount of effort and expense on your part and this effort should be protected by having the analysis run properly. We suggest that it will be faster, cheaper, and more reliable to run your data with us here at Colorado State University.
3. We ourselves would hope to gain farther insight to the structure of biological data and the nature of classification problems through a cooperation with outside workers.

Mr. D. L. Largent

- 2 -

December 14, 1966

If you do choose to cooperate with us in this, we would expect a contribution in difference of operating expenses and machine time. If you are not really interested in cooperation, or did not realize the time and effort involved, we would be willing to send you program listings and wish you luck.

Very truly yours,

George F. Estabrook

GFE:ch

Enc.

P.S. Please also find enclosed a reprint of BioScience 16(11).

NORTH CAROLINA STATE COLLEGE

THE COMPUTING CENTER
POST OFFICE BOX 5445
RALEIGH, NORTH CAROLINA

November 28, 1966

Dr. David J. Rogers
Taximetrics Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

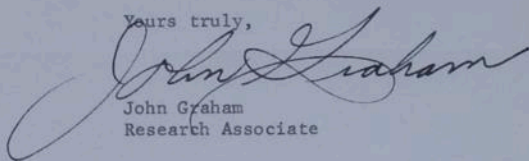
Dear Dr. Rogers:

Thank you indeed for the general write-up for the similarity-clustering program and for the separate from Systematic Zoology. With regard to the latter, I am a member of the associated Society, and therefore have a copy of the article.

Incidentally, I shall be aiding Dr. Lucas and Dr. Wellman in Plant Pathology with the international phytopathology project if it gets off the ground financially. Naturally, the material you have sent could have relevance for such a project as well as for the taxonomic projects I originally had in mind.

Here's hoping (a) I will have some time in the near future to examine the write-up in detail and (b) I will be able to communicate with you again on these matters.

Yours truly,

A handwritten signature in dark ink, appearing to read "John Graham", with a long, sweeping underline that extends to the left.

John Graham
Research Associate

JG:gl

THE UNIVERSITY OF KANSAS
COLLEGE OF LIBERAL ARTS AND SCIENCES
LAWRENCE, KANSAS

DEPARTMENT OF ENTOMOLOGY

SNOW ENTOMOLOGICAL MUSEUM

1 September 1966

Dr. David J. Rogers
Taxonomy Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado

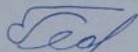
Dear Dave:

Thank you very much for your recent letter and for the thesis which you returned with comments. I spent the month of August back East attending AIBS meetings and collecting plant and insect material, and so I have not been able to answer your letter until now.

I agree with most of your comments; there are one or two things which we should discuss but which would be a bit lengthy and awkward in a letter, especially those relating to character and character state. I am still not sure how vigorous a hybrid can be produced between a biologist and a mathematician. I am, of course, trying to obtain as much math and statistics as I can while holding on to the biology I have. I may well come to the same conclusion you have, that it is just too much for one person to be both a mathematician and a good biologist.

Again, thanks for your comments.

Sincerely yours,



Theodore J. Crovello
Research Associate

TJC:cm

July 28, 1966

Dr. Theodore J. Crovello
Department of Entomology
The University of Kansas
Lawrence, Kansas 66045

Dear Ted:

Under separate cover I am returning your thesis. I have made some comments both in the margin and on pieces of paper along with it. You can take these for what they are worth. They are my opinion based on our experiences in this field.

Most of these comments are intended to try to sharpen up our language and what must be done in order to bring the whole bit of mathematical models and the science of taxonomy up to a more scientific level. When one embarks upon the process of employing mathematics and computers in any field, one is obligated to express one's self in such a way that mathematicians as well as biologists can understand the meaning. Under these circumstances, unless you follow procedures given in mathematics, that is, the statement of premise, the establishment of rules, the discussion and certainty and the certainty that the rules used do not contain any illogical process, you will not succeed in the discussion. I know, I have had this difficulty myself. Our training as taxonomists does not give us any of the precision required for this field, nor, unfortunately, do courses labelled as biostatistics do it either. It apparently takes more training and more disciplining in the mathematical field itself than we can expect to have. It is for reasons such as this that I personally have employed mathematicians trained broadly in that discipline rather than to try to wrap up the needed mathematics myself. I find that being a taxonomic botanist is a full-time occupation and requires the best efforts of my part to understand what is going on mathematically but to insist that the mathematician do the mathematics only after he and I thoroughly understand what is wanted.

It is clear, I think, from my little comments that I am trying to get across an attitude that I think we must assume if we are to get any kind of models to work for us and any kind of computer programs. Really when it comes down to it, many of the problems deal with the identifying pieces of information which we frequently refer to as characters and character states. These problems of characters and character states are not new. They have been with us for years and years and years. I think

July 28, 1966

it is time that we recognized that a character is strictly a construct in the mind of the biologist and has very little significance from the standpoint of the biology of the plant itself. This refers to any classification, not only those in which we try to get a computer to make the necessary comparisons for us, or whether we are doing a strictly 19th-century, straight forward classification as perhaps Asa Gray may have done it.

Although I still feel that we are doing pioneering work in helping the science to grow, I have lost, as you can see, some of the missionary fervor which usually accompanies the first shifts out an old and into a new way of doing things. Although we may eventually end up with some good useful computer programs—I think there are some running around now—we still have a very long way to go before we can claim to have a satisfactory procedure, and I still think that all we are doing is developing procedures, we are really not changing our attitudes, goals, or desires in the field of taxonomy.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

THE UNIVERSITY OF KANSAS
LAWRENCE, KANSAS 66045

DEPARTMENT OF ENTOMOLOGY

SNOW ENTOMOLOGICAL MUSEUM

18 July 1966

JUL 21 1966

Dr. David Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers:

Under separate cover I am sending you a copy of my thesis. I hope it will be of value to you.

I am just getting into my work with Dr. Sokal, which right now is not numerical taxonomy, but a study in geographic variation. It is a continuation of his work on Pemphigus, a parasite that inhabits the petiole of Populus, but this is the first time that we are trying to analyze variation within the host species as well.

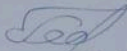
I enjoyed your recent article on ^{the} Systematic Zoology. If you have an extra copy, I would appreciate it. Have you been able to add another mathematician to your staff to replace Estabrook?

I have decided to take that job at Notre Dame and will be starting there in September 1967. Besides curating an herbarium, I should also be teaching biostatistics and doing research in numerical taxonomy and data retrieval.

I hope to see you at the AIBS meetings in August. If not there, perhaps we can get together sometime before the year is out, either here at Kansas or back at Fort Collins.

Dave, I would appreciate your candid comments on any parts of my thesis.

Sincerely yours,



Theodore J. Crovello
Research Associate

TJC:cm

August 12, 1965

Mr. Theodore J. Crovello
Department of Botany
University of California
Berkeley 4, California

Dear Mr. Crovello:

I will not be at the AIBS meetings nor will I be in New York at a time convenient to you. We are transferring our department complete from The New York Botanical Garden to the Department of Botany, Fort Collins, Colorado. We will begin our operations there about the first of September.

One of my assistants, Mr. Henry Fleming, will be at the AIBS meetings, and I will ask him to talk to you about the various problems involved. While he has not been in on the discussions about this particular problem, I am sure that he can help you. I will show your last two letters to Mr. Fleming so that he will at least have those available for your discussion.

Another of my assistants, Mr. Robert C. Jancey, will be giving a paper with me along the lines that we had intended to give. His discussion may be useful to you.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

Department of Botany
University of California
Berkeley 4, California
July 29, 1965

Doctor David Rogers
The New York Botanical Garden
Bronx, New York 10458

Dear Doctor Rogers:

Although I have not yet received a reply to my letter of July 15, I have two things to write to you about.

Larry Heckard mentioned that you had withdrawn from the Urbana meetings. I was sorry to hear that, because I would have welcomed the chance to discuss the problems my research is posing. Also, I would have liked your opinion on some methods I have been developing.

But perhaps I can still see you. I had planned to go from Urbana to New York to spend the weekend with my parents in Queens before returning to Berkeley. My airline reservations, as they are now, will have me land at JFK on Friday night. However, if it would not be inconvenient for you, perhaps I could fly to New York in time to come up to the Garden sometime Thursday or Friday. If you would be available to talk with me Thursday or Friday, please let me know so that I can try to get my flight changed. A less convenient time would be to meet Saturday morning. However, I do not want to impose upon your free time if you follow a regular weekly schedule.

The second reason for ~~this~~ letter is to let you know that I am at least aware and thinking of the problem of expressing characters in terms of the usual statistical parameters of the mean, standard deviation, etc. However, I would appreciate your comments on the scale which is used to measure similarity as a distance. If I understand Mr. Eatsbrook correctly, he would rather the distance (similarity) be measured in standard deviates. While most of the applications of distance have used standardized characters, I am not quite sure yet that one can give an objective reason why this is the correct way and the estimate of distance by raw measurements (after transformation to give equal weight) is incorrect. How does one measure similarity between two OTUs? Is it in terms of the number of OTUs that stand between them in the expression of a character or is it in terms of the fraction of the entire range of expression of a character in the group under study that separates the two OTUs? I hope I can discuss this with you.

Please let me know if you will be available in New York that last half week so that I can change my flight if necessary.

Sincerely yours,

Ted Crowell
Theodor J. Crowell

Department of Botany
University of California
Berkeley 4, California
July 13, 1965

Doctor David Rogers
The New York Botanical Garden
Bronx, New York 10458

Dear Doctor Rogers:

Thank you for your letter of July 8 and for returning my manuscript.

The comments of your Mr. Estabrook cover quite a bit of territory. Actually, some of it seems to be directed towards the entire field of taximetrics and its shortcomings instead of to the drawbacks of my small research note. From the comments in your "loose preamble," I am eagerly awaiting the next publications from your department, since now I have only your published work through the Bioscience article to assess your methods.

Returning to the strong statements you used in your review, I find that neither you nor Mr. Estabrook really expended on most of them. I await a list of the articles you think I should have read. Furthermore, you have not told me, "exactly what is supposed to be done." You reject my methods but leave me hanging. How shall I finish my thesis? I am sure you have a substitute, "better" method.

Also, will you or Mr. Estabrook please tell me how to handle the example I gave in my letter of June 20?

Since Sheals' work involved the method I have called condensation, do you think his work is also shallow?

Are you or Mr. Estabrook going to be at the AIBS meeting in Urbana? I shall be giving a paper on some of the problems of the application of numerical taxonomy to the Angiosperms. If you are I would welcome the opportunity to meet with you and discuss the problems more freely.

Thanking you in advance for your constructive comments, I await your reply.

Sincerely yours,

Theodore J. Crovello
Theodore J. Crovello

October 22, 1964

Mr. Theodore J. Crovello
Department of Botany
University of California
Berkeley 4, California

Dear Mr. Crovello:

Thanks for your letter of the 11th. I am glad to hear that you are working into taximetric methods with genera that are "difficult." I know that Salix is in great need of such work.

You asked about our input cards. Our design is intended to be used either way--with specimens or with formed taxa. I do not think it profitable to employ a method which starts with a subjective judgment. We used a completed monograph (Halimeda) only to test our program. We imply in the BioScience article, although we do not definitely state it, that one should or could start with individual specimens.

As soon as reprints are available, one will be mailed to you.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF
Enclosure

Department of Botany
University of California
Berkeley 4, California
October 11, 1964

Doctor David J. Rogers
The New York Botanical Garden
Bronx 58, New York

Dear Doctor Rogers:

Congratulations on your recent article in Bioscience. It contains a number of strong points, among which may be listed the fact that you invited the author of the monograph to review your interpretations and your method of centroid determination.

This latter point recalls a method that Doctor Herbert Meson has been experimenting with, although he has expressed his ideas in the formal language of set theory. Unfortunately he has been away on a cruise for the last five months and I can only hope that he has kept up his interest in this subject.

The way you coded such characters as the number of utricle layers also interested me. Currently I am a PhD candidate under Professor Lincoln Constance. My research involves the genera Fremontia and Salix, the former genus serving as a pilot project in numerical taxonomic method, which will then be used in Salix. As such, I am confronted with many characters with overlapping attributes, such as your utricle layers. Sokal and Sneath in their early studies just avoided such characters, yet, if I were to do that, I feel I would be left with too few characters to work with. Leaf length is a good example.

Will you clear up a question for me? In the Bioscience article you talk of specimens and illustrate a specimen card, yet your project was the analysis of the information in a monograph. Did you actually process individuals, giving each a card, or did you only use one card per taxon used in the monograph, incorporating its variation on it?

May I please have a reprint of your Bioscience article as well as a list of the characters and attributes used? A sample IBM card would also be appreciated for demonstration purposes. Have you written up the work reported in abstract form in AJB this past summer?

If you are interested in more details of my study, please let me know and I will expand. Thank you for your cooperation and I await your reply.

Sincerely yours,

Theodore J. Crowello
Theodore J. Crowello

THE NEW YORK BOTANICAL GARDEN
BRONX • NEW YORK 10458 LU 4-8500

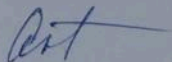
June 6, 1966

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

Dear Dave:

Thanks for your letter of June 2. As you know, I am sympathetic to your efforts to develop the theory of numerical taxonomy and the use of computers in taxonomy. I believe you also know that I do not plan to do any work in that field myself. I am much more interested in coming up with what I consider to be the right answers than in proving their correctness to somebody else. I can already see more work ahead of me that I would earnestly like to do than I am likely to get done in my life. People are always coming to me with things that would be wonderful for me to do. A lot of those things are worth doing, and some of them I could effectively do, but only at the expense of something else I am more eager to do. So thanks, but no thanks. On the other hand, I would be glad to discuss the progress of your work with you at any time that is mutually convenient. One of the things I really do want to do is keep up with what is going on.

Yours,



Arthur Cronquist
Senior Curator

AC:gg

THE ACADEMY OF NATURAL SCIENCES OF PHILADELPHIA

FOUNDED 1812

NINETEENTH AND THE PARKWAY, PHILADELPHIA, PENNA. 19103 LOCUST 4-3921

JOHN W. BODINE
PRESIDENT

December 13, 1966

H. RADCLIFFE ROBERTS
DIRECTOR

Dr. David J. Rogers
Botany Department
Colorado State University
Fort Collins, Colorado 80521

Dear Dave:

Thank you so much for your letter of November 30th. I have not had a chance to really study the reprint in detail which you sent, but I do understand the instructions for filling out the characters for the various specimens. We will get to work on this in the near future.

With best wishes.

Yours sincerely,



Ruth Patrick
Chairman
Department of Limnology

RP:lt

November 30, 1966

Dr. Ruth Patrick
Chairman, Dept. of Limnology
The Academy of Natural Sciences
of Philadelphia
Nineteenth and the Parkway
Philadelphia, Penna. 19103

Dear Ruth:

I'm glad to know you're willing to have a go with our clustering methods. Below I say a few words about what, why and how. Also, I've included a reprint of an earlier paper, and have marked some pertinent words which may help a bit in preparing your data. But the most significant paper to understand is the one we've already sent as a preprint. (This last paper appeared in BioScience, the November issue.)

As a trial run, and before you commit yourself to a very large task, I suggest that you select a sample of critical specimens from G. parvulum and G. angustatum. In selecting the specimens, pick those with the most complete information, because you can't make a good classification with missing information. How many specimens you select is a matter for your judgment. Without knowing the complexity of the problem, I can only suggest that perhaps the trial could include 50-75 specimens. This is probably sufficient to determine the value of the method. But the suggested figure is no magic number—you could include less if, in your judgment, you could see the effectiveness of the program with a smaller number.

The number of characters you choose to describe each object (specimen, or population sample) is again a matter to be decided by you. We have used as few as 14 characters and 110 states, and as many as 85 with 400 total attributes (states). The upper limit in the program is something like 115 characters, but I have yet to see anyone who used many more than this number. You know what information you want, and no one else can be as good a judge of it as you. Do not assume any statistical technique can dictate how many you have to have. The most significant question to ask about each character is: does it contribute to the task of comparison of the objects?

You could select, say, 15 of the most interesting characters, if you have more, and see what happens to the resulting clusters. You could include 20, putting those you consider the least reliable at the end of the list, and then, after seeing one run, decide to strike out some of the later listed characters, and run it again for comparison. Time for running as many as 100 objects is very short--10-15 minutes at most, so you can play around without the cost becoming prohibitive. It should be noted that the number of characters used for each object has very little influence on the running time--the number of objects included in the study is the time-determining factor.

You may have some difficulty in setting up characters to get the most benefit from our methods. I am referring specifically to those conditions referred to in part III, Considerations of Intrastate Relations, of the preprint we sent earlier. After you have selected your characters, and find that some of them may be structured in one of the ways suggested in that paper, you may want to check with us before coming to a final decision. While we attempted to duplicate the thought processes of taxonomists in character use, when it comes out in black and white, the method seems to be far from anything the working taxonomist does. So, having worked up the method, we can sometimes suggest the most profitable way to structure the character to fit your needs.

A sample coding sheet with instructions is included.

I hope this information gets you started--but please feel free to query us on any doubtful point. We want to help out.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

THE ACADEMY OF NATURAL SCIENCES OF PHILADELPHIA

FOUNDED 1812

NINETEENTH AND THE PARKWAY, PHILADELPHIA, PENNA. 19103 LOCUST 4-3921

JOHN W. BODINE
PRESIDENT

November 22, 1966

H. RADCLYFFE ROBERTS
DIRECTOR

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

Dear Dave:

Thank you so much for your letter of November 7th. I would like very much to work with you and to use your clustering method to see whether or not it elucidates the problem of Gomphonema parvulum and Gomphonema angustatum. Kindly let me know what kind of data I should send you and I will try to get to together in the next month or two--that is, by the end of January. This will be fun, and I will enjoy working with you.

With all best wishes.

Yours sincerely,



Ruth Patrick
Chairman
Department of Limnology

RP:lt

- Taximetrics Laboratory

November 7, 1966

Dr. Ruth Patrick
Philadelphia Academy of Natural Science
Philadelphia, Pa.

Dear Dr. Patrick:

It was indeed a pleasure to see you at the Yale symposium. I have only one regret that we did not have longer to chat. I think I promised to send you a copy of a paper of ours that has something to say about the structuring characters to be used in our clustering method. Perhaps you will find this of interest. I am also enclosing a discussion of our graph clustering technique.

We have found this clustering method extremely useful. It is now programmed and has been used in a number of interesting applications on groups of organisms as diverse as strains of streptomycetes to a subtribe of orchids. I have used it myself for Manihot esculenta and am now satisfied that I have a classification of the cultivars. This latter, of course, has been uppermost in my mind in the development of any computer methods.

If you are interested in testing this program, we could accomplish this in one of two ways. We here at Fort Collins could run your data through our hardware and send you the results. Or two, you could reprogram for whatever computer facilities are available to you there. Of the two courses, I recommend the former, purely because of the great difficulty in getting a new program transferred from one machine to another. It is no simple task, since this program is in the words of my mathematician "a nontrivial task." I do not recommend the job of the transfer to another computer unless the computer be of the same configuration and number as the one for which the program was originally written.

I would be very happy to collaborate with you on some problems amongst the diatoms.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

December 15, 1961

Dr. Ruth Patrick
Philadelphia Academy of Sciences
19th Street and Benjamin Franklin Parkway
Philadelphia, Pennsylvania

Dear Dr. Patrick:

It was indeed a pleasure to visit you yesterday. We certainly do appreciate the good advice on our project and the nice luncheon which we had with you and Dr. Roberts.

I have just inquired about the diatom collections and find that they are available now. Dr. Rogerson and his assistant, Mrs. Fleming, are looking forward to your visit.

We shall keep you informed as to our progress. Thanks for a most cordial visit.

Very sincerely yours

David J. Rogers
Curator of Economic Botany

DJR:MDF

December 8, 1961

Dr. Ruth Patrick
Philadelphia Academy of Sciences
Philadelphia, Penna.

Dear Dr. Patrick:

I am sorry to have delayed so long in confirming my telephone conversation with you last month for a conference on December 14. I had hoped that the contract with the Office of Naval Research under which we are doing the computer work would have been signed long before now. Only today has our business office signed the contract, and I feel confident that we can go ahead with the project now.

If it is still convenient for you, we would like to come to see you some time around 11:30 (I think there is a bus which arrives in Philadelphia around 11). The purpose that we have in visiting you is to get your opinion of some of the most suitable algal taxonomies--those which have stood the test of time, those which are more modern, and ones in which we may have some confidence in working with. I will give you more details about our procedures and hoped for results when I see you. I will be bringing along with me Mr. Henry Fleming who will be working hard on this project with me.

Looking forward to seeing you, I am,

Sincerely yours

David J. Rogers
Curator of Economic Botany

djr;mdf

INTERNATIONAL ASSOCIATION FOR PLANT TAXONOMY

Office:
International Bureau for Plant Taxonomy
and Nomenclature
106 Lange Nieuwstraat
Utrecht
(Netherlands)

UTRECHT, 22 May 1966.

MAY 25 1966

Bankers: Messrs Vlaer & Kol, Utrecht
Telephone: (030) 15891

Prof. David J. Rogers
Department of Botany
and Plant Pathology
Taxonomy Laboratory
Colorado State University
Fort Collins, Colorado 80521 - U.S.A.

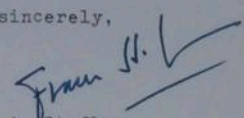
Dear Dr Rogers,

Thank you very much for your kind letter of 16 May 1966. I am very grateful to you for your interesting remarks. The difference between ADP and EDP is known to me and I realize quite well that Mr Gould used a rather simple form of automatic data processing machines. This is exactly one of his weaker points because he now had to let the machinery limit his input of information. I certainly hope that we shall move rapidly toward more sophisticated data processing in biology, the times are over-ripe. I do not agree that we should ignore Gould and his work. I think he has shown dedication and enthusiasm and these may well be acknowledged. Furthermore it is necessary not to take his shortcomings too tragically; it is necessary to outline them without too much emotional involvement.

As soon as you feel that you want to write on these subjects, and on steps doing away with the naiveté mentioned by you, I hope that you will think of Taxon as a medium to reach the taxonomic public!

With all good wishes,

yours sincerely,


F.A. Stafleu.

- Taxonomy Laboratory

May 16, 1966

Dr. F. A. Stafleu
106 Lange Nieuwstraat
Utrecht, Netherlands

Dear Dr. Stafleu:

I wish to congratulate you on the sensitive, accurate and appreciative review of the publication "Authors of plant genera" by Gould. Yours is a much more objective attitude than the more frequent emotion-filled reactions we have come to fear when we say anything about computers for taxonomy.

There are a few observations I should like to place before you, however, for there are still many misunderstandings involved that we as taxonomists must overcome before we can really put the complex devices generally known as electronic data processing machinery (EDP) to work. I think that I am somewhat qualified to make the following comments, having been deeply immersed in this type work for some eight years. In this time, we (I speak of myself, my assistants Henry Fleming, George Estabrook and Robert Brill) have made much progress from complete naivete to something of a sophistication relative to computers not to be found in the large body of taxonomic scientists.

First of all, there must be a clear understanding of the various devices which generally, mistakenly, get lumped together under the term "computer." There are many levels of complexity of the machinery. There are less complex devices which will punch cards, sort the cards, rearrange the data contained on the cards, and print the information contained on the cards in any specified manner. These types of devices are so-called "peripheral" machinery to the main computer: They are the hand-maidens to the central machine. They are frequently designated as automatic data processing (ADP) machines, and as the designation implies, process the input data, but seldom do any computation on the data. The methods of directing these devices are quite simple, much as an automatic loom is directed with a Holerith system to weave certain patterns of cloth. These ADP devices have been used exclusively by Gould in his work. In a real sense, Mr. Gould has never used a computer. As a result, he has had to limit his format and capability to the capacity of these essentially limited devices. These devices cannot cope with the variable word length of botanical epithets, largely because the set of instructions which the peripheral equipment can follow is extremely limited. It is unfortunate that Mr. Gould

May 16, 1966

has not recognized these shortcomings, and furthermore, has done nothing to remedy the situation. He is not only limited in his knowledge of the biological problems, but is, as well, limited in his knowledge of computers.

Second, if we are to achieve any sort of acceptable application of the computers, we must resort to the "main frame" of the computer, the device where most of the excitement of the computer is. The major advantages of computers are their extreme speed and accuracy, but these are available to us only because of the concept of a stored program. The stored program is a set of predetermined instructions which directs the action to be taken by the computer on our biological data. Even with the computer, we are not requesting computation to be done with names, authors and bibliographies, but we want to have sufficient flexibility to permit us to see the results of computer manipulation in a usable fashion, without resort to standardization on an artificial basis. With the stored program, such flexibility is available because the stored program is first established by biologically knowledgeable people in collaboration with computer programmers.

Our own naivete has unfortunately led to such as Mr. Gould's stepping into the picture. I hope that we can forget this unfortunate episode. I trust that taxonomists will not continue fighting against something that is not worth their efforts to fight against. The best treatment (though I do not advocate this because of a desire to hurt the individual) is to totally ignore the man and his work.

We must get on with efforts to apply these marvelous devices, but we must recognize that we are the masters, the directors, and not the machine. Our fear of machine direction is fear based on ignorance, and this is not tolerable if we are to call ourselves scientists.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

THE UNIVERSITY OF MICHIGAN

SCHOOL OF PUBLIC HEALTH

ANN ARBOR, MICHIGAN 48104

August 1, 1966

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

Dear Dr. Rogers:

I have received your letter of August 10 and the mimeographed paper on taxonomic description. At the moment I have no questions on the paper but perhaps when I begin to work with your computer program there will be some questions.

I have been looking forward to hearing from you, particularly to receiving the listing of your computer program, mentioned in your last letter. If this listing has not yet been sent, can you suggest when I might be receiving it? I am in the process of analyzing (1) my "real" data on influenza virus and (2) a large number of computer simulated experiments. Whether your cluster technique using information theory is applicable to my problem and how much computer time it will require are two questions I hope to answer soon.

Thank you again for your assistance.

Sincerely yours,

Anna M. Lee

Anna M. Lee
Department of Biostatistics
University of Michigan
School of Public Health

THE UNIVERSITY OF MICHIGAN

SCHOOL OF PUBLIC HEALTH

ANN ARBOR, MICHIGAN 48104

November 9, 1966

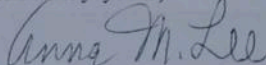
Dr. David J. Rogers
Taxonomy Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

Thank you for the materials which you sent to me last month--the flow chart and program listing for your graph theory model for clustering. Unfortunately the materials were tied up in the post office for most of the month of October (due to a labor shortage); and they arrived too late for me to make use of the program in my dissertation, which should be finished this month.

I am happy to receive your materials, however, and hope to investigate the possibilities of using your graph theory model for my classification problem at another time. Thank you again.

Sincerely yours,



Anna M. Lee
Department of Biostatistics
School of Public Health

AML:ss

- Taximetrics Laboratory

November 3, 1966

Dr. Warren H. Wagner, Jr.
Director, The Botanical Garden
University of Michigan
Ann Arbor, Michigan

Dear Herb:

Thanks for the reprints.

What has happened to one of your advisees, Anna M. Lee? We sent her a big, long, detailed flowchart of our computer program, and have not had an acknowledgement of it. To refresh your memory (if that need be) she is in the Department of Biostatistics, working on influenza virus.

Come see us sometime.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

- Taxonomy Laboratory

October 3, 1966

Miss Anna M. Lee
Department of Biostatistics
School of Public Health
The University of Michigan
Ann Arbor, Michigan 48104

Dear Miss Lee:

Following your telephone call and several letters, enclosed is a corrected flow chart and a program listing. Please pay careful attention to the handwritten instructions on the listing.

I have a copy of my letter to you on July 7, indicating that we had given up on the Information Theory model devised by Tanimoto and me. I suspect that that letter didn't get to you, or else didn't ring a bell with you. The basis of this suspicion was your comment over the phone to me a few days ago. However, be advised that the program we're sending is not the Information Theory model, but is one based on graph theory. It is far superior to the Information Theory model, and accomplishes much more from the biologist's point of view.

Also, with the July 7 letter, I sent a preprint of a paper on the establishment of characters which will be critical if our program is to do the job for you. The gist of this preprint is that there is much needed to be done by the biologist before getting to the clustering program, and no computer can substitute for the biologist's knowledge of his own material, and the proper manner of setting up his data. Such statements can be made no matter what computer method is employed--ours or Sokal's or Sneath's or any of several others. But for our method, the techniques described in the preprint work. I think this is one of the unique features of our method--there is no hand-waving about the input. Follow the instructions given, and you can do a job.

Please also refer to our paper in Systematic Zoology, Vol. 15. This gives you some of the needed background for work with the graph theory model. There are several other items that we could help you with, and for this, I recommend strongly that you find some way to visit us.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch
Enc.

- Taxonomy Laboratory

August 10, 1966

Miss Anna M. Lee
Department of Biostatistics
School of Public Health
The University of Michigan
Ann Arbor, Michigan 48104

Dear Miss Lee:

I have your letter of July 28. We would have sent the listing requested sooner but have delayed pending a few minor changes in it which will improve the running efficiency. This would have been done but my programmer has been tied up because of the airline strike. I am hoping he will be back next week. At that time we will have a listing made for you.

In the meantime, I am sending a mimeographed copy of a paper which I think is necessary and essential to anyone who wishes to utilize the clustering methods employed. We have discovered that the most critical element in the clustering analysis is in the understanding and preparation of the input data. Please note that the enclosed mimeographed paper is not for distribution. I send it to you because I expect that you would like to follow this format (or the ideas herein in setting up your information). If there are any questions after reading this paper, please do not hesitate to inquire. We have discovered that the input methods require more than average knowledge of the classification procedure throughout than is generally available to most workers. The listing of the program should be forthcoming soon.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

THE UNIVERSITY OF MICHIGAN

SCHOOL OF PUBLIC HEALTH

ANN ARBOR, MICHIGAN 48104

July 28, 1966

Dr. D.J. Rogers
Department of Botany
Taxonomy Laboratory
Colorado State University
Fort Collins, Colorado 80521

Dear Dr. Rogers:

I inquired last month about obtaining a listing of your taxonomy program for the computer. Professor Fleming replied to letter, indicating that you "have written a new program for the analysis of characters using the information theory base employed in the former program but with an entirely different base of definition." He further indicated that the program is presently being debugged for the IBM 7044 and 7094.

I should like very much to obtain a listing of this program when it is available, should it be available in the near future. Can you suggest when it might be ready? Our computer is an IBM 7090. It should not be too difficult to adapt your 7094 program to our system since we have a programmer here to help us.

Presently my data are ready for analysis and I am interested in applying several "cluster" techniques to the data and then comparing the results. I shall look forward to hearing from you.

A copy of my original letter to you is enclosed.

Sincerely yours,



Anna M. Lee
Department of Biostatistics

June 27, 1966

Dr. David J. Rogers
Curator of Economic Botany
New York Botanical Garden
New York 58, N.Y.

Dear Dr. Rogers:

I am presently working on my doctoral dissertation, one aspect of which is a comparison of various clustering or clumping techniques as evidenced by their use in analysing serological data on influenza virus. My distance measure is the immunological distance (or its negative, mentioned by Sokal and Sneath, Numerical Taxonomy, p. 43).

It appears from the article written by you and Dr. Tanimoto, "Computer Program for Classifying Plants," Science, 1960, that your "cluster technique" would be applicable to my data. I should like to have a listing of the computer program to which you refer in that article, along with any instructions necessary for its use. Is this possible? Needless to say, I would very much appreciate receiving the listing and enclose a self-addressed envelope for your reply. My data are nearly ready for analysis, so I can assure you that your program will soon be put to use.

You and Dr. Tanimoto will, of course, be acknowledged in my dissertation. Thank you.

Sincerely yours,

(Miss) Anna M. Lee
Department of Biostatistics

CURATOR OF ECONOMIC BOTANY
NEW YORK BOTANICAL GARDEN

JUN 28 1966

THE UNIVERSITY OF MICHIGAN

SCHOOL OF PUBLIC HEALTH

ANN ARBOR, MICHIGAN 48104

June 27, 1966

JUL 5 1966

Dr. David J. Rogers
Curator of Economic Botany
New York Botanical Garden
New York 58, N.Y.

Dear Dr. Rogers:

I am presently working on my doctoral dissertation, one aspect of which is a comparison of various clustering or clumping techniques as evidenced by their use in analysing serological data on influenza virus. My distance measure is the immunological distance (or its negative) mentioned by Sokal and Sneath, Numerical Taxonomy, p. 43.

It appears from the article written by you and Dr. Tanimoto, "A Computer Program for Classifying Plants," Science, 1960, that your "cluster technique" would be applicable to my data. I should like to have a listing of the computer program to which you refer in that article, along with any instructions necessary for its use. Is this possible? Needless to say, I would very much appreciate receiving the listing and enclose a self-addressed envelope for your reply. My data are nearly ready for analysis, so I can assure you that your program will soon be put to use.

You and Dr. Tanimoto will, of course, be acknowledged in my dissertation. Thank you.

Sincerely yours,

Anna M. Lee

(Miss) Anna M. Lee
Department of Biostatistics

Enc.

- Taxonomy Laboratory

July 6, 1966

Miss Anna M. Lee
Department of Biostatistics
School of Public Health
The University of Michigan
Ann Arbor, Michigan 48104

Re: Letter to Dr. D. J. Rogers
New York Botanical Garden
June 27, 1966

Dear Miss Lee:

The program you are referring to is available through IBM. The listing is available through the title "A taxonomy program for the IBM 704" produced by T. T. Tanimoto and Robert A. Loomis. The program deck is also available through the IBM Applications Library at 590 Madison Ave., New York 22, N.Y.

Our own laboratory has since abandoned this approach. Though the information provided by the concept was an improvement on what had been previously available, the biological premises were too restrictive and the mathematical definitions inadequate biologically. We have written a new program for the analysis of characters using the information theory base employed in the former program but with an entirely different base of definition. This program at present is being debugged for the IBM 7044 and 7094.

As a clustering program, we are using a computer program based on graph theory--a relatively new, 10 to 15 years or so, math discipline--and find it far more satisfactory for developing classifications than the original Rogers-Tanimoto program you referred to. Unfortunately, we do not have listings and program decks available yet, though it has been debugged and is running on the IBM 7044 and 7094. If you are interested in this program, write Dr. Rogers toward the end of July as he will be in Mexico until then (but see Systematic Zoology 15: 59-69).

You will have noticed from the above letterhead that we have moved from the New York Botanical Garden. Any future correspondence should be sent here to the Taxonomy Laboratory.

Sincerely,

Henry S. Fleming
Professor of Botany
(Assistant to Dr. D. J. Rogers)

HSF:ch

- Taximetrics Laboratory

November 17, 1966

Dr. Wm. Bridge Cooke
1135 Wilshire Ct.
Cincinnati, Ohio 45230

Dear Bridge:

Following our telephone conversation, I checked with Henry Fleming who had the opportunity to talk with you last summer about your particular interests. After talking to him, it turns out that my estimation given over the phone fell somewhat shy of the needs. Inasmuch as your requirements will take some additional research on our part, I am sending herewith a preliminary budget which indicates, (1) our contributions to your project and, (2) what will be needed from you. These are not final figures but they are estimations made on the basis of past experience and should not be too far out.

What we will do for you:

1. Develop a program specifically designed for your yeast identification.
2. Make tests and trial runs on your data as demonstration of the method.

What we cannot do under this contract is to continue as a "service bureau" once the programming is complete. You may wish to consider that problem with your fiscal authorities when we complete the research work.

Note that the budget has: no travel funds for us and no funds needed by you for your own work or travel.

Let me know what the possibilities are for this work.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

OUR CONTRIBUTION

1. Biological analysis of the problem
 - a. establishment of the needs in a key program
 - b. setting the limits or rules for such programming _____ \$10,000
 2. Interpretation of the program
 - a. how the computer key program works
 - b. what are the "side benefits" of the key program _____ 10,000
- Total contributions on our part _____ \$20,000

NEEDED FROM YOU (your contributions)

1. Research requirements
 - a. development of algorithms _____ \$ 5,000
 - b. programming _____ 5,000
 - c. checking out the program (debugging) _____ 2,000
 - d. testing the classifications of yeasts separately and merged _____ 1,250
 2. Consulting _____ 1,000
 3. Punch operators and secretarial time _____ 1,500
- Total direct costs _____ \$15,750
4. Overhead 42% _____ 6,615
- Total costs _____ \$22,365

- Taxonomy Laboratory

June 20, 1966

Dr. Wm. Bridge Cooke
1135 Wilshire Ct.
Cincinnati, Ohio 45230

Dear Dr. Cooke:

My home telephone number is 667-7286. The area code is 303 which covers the whole of Colorado. Since your time of arrival is indefinite, it is somewhat superfluous to give you my lab number but here it is anyhow--491-5326. Let the home number ring as we are likely to be out in the garden working. We live just north of Loveland on the only road going east and west north of Horseshoe Lake (a westerly adjunct of Boyd Lake). Horseshoe Lake is north-east by north of Loveland and the road is the first east west road north of Loveland city limits on the road 287 from Loveland to Fort Collins.

I'll alert Lawrence Durrell that you are likely to come through on the 4th p.m.

Motels should not be a problem between Loveland (your entrance town to Rocky Mountain National Park) and Fort Collins.

In the time available, we won't be able to do more than sketch out the program in broad outlines and give you some idea of what it does. The time will probably be best spent in dealing with the nature and manner of entering the taxonomic information into the computer for analysis. A list of your characters and character states as you have coded them will be far more pertinent to the discussions than your data cards.

Will be waiting upon your arrival.

Sincerely yours,

Henry S. Fleming
Professor of Botany

HSF:ch

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

15 June 1966

Dear Dr. Fleming,

As you may know, Dr. Rogers suggested that I contact you concerning a possible stop at Fort Collins en route west on our regular vacation trip.

The biggest problem of our routing ourselves through Fort Collins is that we will not arrive there until the afternoon (late) of July 4 and will have to leave on the morning of July 5. On the fifth I plan to collect some fungi in Rocky Mountain National Park and then go on as far as Craig or Vernal. But there will be some time on the evening of the 4th for discussion and exchange of information. We have not made reservations for that night, usually leaving ~~up~~ that up to luck on arrival. We have stayed at a small, old motel on the north side of town on the road toward Laramie. It may have changed hands since last July since the women who ran it were trying to sell it. It may not exist or it may not be what we would want this year. Usually we try to find something at lower rates than AAA motels.

I will have with me a "master" set of punch cards carrying the published characteristics of each yeast species described up to last Sept. 1. On the basis of this set of cards I hope to be able to do two things: First, identify by comparison unknown strains of yeasts for which punch cards have been ~~exp~~ prepared on the basis of laboratory note sheets on which results of experiments have been entered. In some cases there are more data for the unknowns than for the knowns. Second: I would like to determine whether on the basis of punch card systems the present standard method of classification is more workable than that presented by Novak and Zsolt in Hungary recently. The yeasts seem to me to present the best characters, among fungal groups, for punch card and machine taxonomy techniques. You may have had access to the correspondence I have had with Dr. Rogers on this matter.

I hope I will be able to contact you when we reach Fort Collins. If so will you please send me a phone number to call when we reach town and get settled into a motel? Of course, I will also want to see Dr. Durrell at least briefly.

Very truly yours,

Wm. Bridge Cooke

Wm. Bridge Cooke

- Taxonomy Laboratory

February 11, 1966

Dr. William Bridge Cooke
1135 Wilshire Court
Cincinnati, Ohio 45230

Dear Bridge:

So far I have had no luck whatsoever in getting a student interested in the problems that we discussed at Berkeley. Unfortunately, the majority of the students here in mychology, at least at the moment, have interests other than in classificatory processes.

I am sorry to report this because I am vitally interested in the type of data that you have for yeasts and really do want to do something to help out if at all possible. I will keep stirring around to see if some student does not show some glimmering of interest in this area. In the meantime, it might be interesting to consider continuation of your vital studies on yeast with a separate grant request to the National Institute of Health. This strikes me as a very fundamental need and if anybody can handle the problem, certainly you can. If you wish to explore a joint application let me know and I will prepare some basic materials to support your request.

Sincerely,

David J. Rogers
Professor of Botany

DJR/pam

FEB 3 1966

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

28 Jan. 1966

Dear Dr. Rogers,

It was good that we had the opportunity to discuss something of my yeast taxonomy problem last month in Berkeley. Thank you for taking the time to do this.

Several hundred cultures are still available for transfer to fresh agar in case that phase of the project is still of interest. These would be used to complete tests which are still incomplete for one reason or another. If it looks as if I must abandon my present space more rapidly than I now think, then this might not be practical. However, these cultures might survive shipment.

Data from some 4000 data sheets have been transferred to punch cards, as indicated earlier. These represent some 75 species out of a potential of about 450-500 species which have been described for the world. Most of the species, more than 80%, are imperfect, and most of the punch cards, maybe as many as 90-95%, are imperfects. Two species probably make up 25% of the white yeasts and two species make up about 75% or more of the red yeasts. Many characters in these larger ~~gm~~ represented species are somewhat variable. My technique for entering data on the cards may have eliminated some of this variability.

Presently, yeast identification is strictly subjective. Dr. Herman Phaff at Davis, under whom I learned most of what I know about yeasts, prefers this technique. One of his students, Dr. Moshe Shifrine, suggested the punch card outline which I have modified for my own use. If yeast isolation should become an important area of concern it is possible that large numbers of cultures would have to be identified. It is further possible that in a large laboratory with a small number of professional workers, a subprofessional would be handed the job of gathering data on cultures and subjecting these data to some process for substitute professional opinion. This applies to other groups, also. So a machine which would align the data assembled subprofessionally with a set of master cards would be a useful adjunct to the laboratory.

In addition, the machine could tell whether or not it would be more advantageous to use one system of classification rather than another. The data cards as far as I have gone to date should supply enough information to answer both questions.

Monday, Tuesday or Wednesday, when Dr. Weinberger visits our laboratory, we will learn more of the fate of my operation. In the meantime I will continue to develop groups of cultures to be sent out for one reason or another - identification, storage, etc.

Very truly yours,

Bridge

May 19, 1965

Dr. William Bridge Cooke
1135 Wilshire Court
Cincinnati, Ohio 45230

Dear Dr. Cooke:

In response to your letter of the 13th, there are several comments. First of all, I do plan to be at the AIBS meetings at Urbana, and we may be able to get together there. Whether or not I can make it to Cincinnati afterwards is a moot question, what with getting a daughter off to college at that time, etc.

We definitely should have an opportunity, however, to talk about the data that you have. It is quite likely that the man in charge of your statistical work is not the most important person at this stage of the game. Certainly he will want to know whether or not he can program the model that we have to be run on the computer hardware that you have there. It seems that no two data processing facilities have the same hardware and that for each one there are some special sets of instructions unique to that operation. In that respect the statistician may be important.

However, it has been our experience working with quite a number of different classification problems at varying levels from the Actinomycetes onward that the most important thing is the "input" data itself, and only the specialist can actually sense the questions needed to be answered. These questions deal with exactly how you establish the characteristics and their defining attributes most efficiently. I do not intend by this statement that there is anything wrong with anyone's characters and attributes. It is just that when one has to fit them to a precise procedure there are problems which only the specialist can answer.

I would like to talk to you about this problem. Let's see if we can arrange something while at the AIBS meetings at Urbana. As far as Botanical Review, Volume 28, is concerned, this must be ordered from Stechert Hafner, 31 E. 10th St., New York, N. Y. However, it will probably not be available for a month or two as it is in the process of being reprinted.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

Actinomyces
Actinomyces

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

13 May 1965

Dear Dr. Rogers,

I am sorry that we could not get together last Christmas week but business is business and the meetings in Montreal were probably quite interesting. I had an interesting week at the National Fungus Collections adding to my information about families of fungi and Ohio fungi in the Lloyd Herbarium.

This summer I will again have a college student working for me. Since this will be her fourth summer she is quite familiar with my routines and with the type of information I will need to complete data sheets and punch cards for my yeast cultures. Together with other jobs around the Lab.

We have about as complete a set of data as we can get on all published species of yeasts to serve as background material on which to base matching identification of unknown cultures.

My next problem will be a little more complicated. We now have a "statistician" on our water program staff. He has access to our IBM data processing equipment which he uses in cooperation with the staff of that operation. None of these people are familiar with your work or work like it except through noting that there is a literature on the subject. However, even though he is a bit snowed under with work from all the 4 or more units in the Branch he will be expected to be the local resource person to help me and since there is a local resource person why should I have to go to New York to learn the technique myself. I have answers but need some sort of additional excuse or reason. We plan to go to San Francisco or Berkeley next Christmas week so we cannot go to New York. I have been trying to work out a schedule so that my wife can go too during her Christmas vacation but that did not work out. Problems, problems, decisions, decisions.....

This summer we plan to stop a day in Davis where I hope to get Dr. Herman Phaff to confirm some of my more recent yeast identifications. This helps to build up my confidence that I can match data sheets and published descriptions. He is not particularly happy with the idea of punch card identification although the person who proposed it for yeasts, Moshe Shiffrine was one of his doctoral students. Moshe is now living in Africa and last I heard was not sure he would return soon to the States. You may have seen his short paper in Antonie van Leeuwenhoek.

You may also be in Urbana in August? Since I will leave here early to attend the Foray I cannot ask you to stop here on your way except before the 12th. Or you could stop on your way back on the 20th when I have to be back at work.

Comments appreciated,

Very truly yours,

W. W. Coker

Can I get a copy of Botanical Review, Vol. 28, No. 1, 1962? I need Whittaker's paper. Thanks for any information available. WWC

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

14 Nov. 1964

Dear Dr. Rogers,

Thank you for your informative letter and for the reprint of your latest article. It is obvious, of course, that I need to know a bit more about the details of numerical or computer taxonomy as applied to the problems I would like to have solved for me.

I have two problems: first, on the basis of data available from published descriptions, which of currently proposed systems of classification are better, or which one is the best for general purposes. Second, on the basis of the characters used at present, obtained from cultures available for study, can an identification be obtained for an unknown culture more quickly by machine or by scanning the available data and which would be the more accurate?

Presently there are maybe 500 species of yeasts in the literature. Some of the taxa are nomina nuda, but this detail can be cleared up as required because published names may be needed and there is no need to discard work already done when it can be updated. When there were 100 species, or when a person is familiar with a genus containing 1-20 species, it is or was not difficult to keep the details of each species in mind and come up with an identification, less and less tentative as results of morphological and physiological tests became available. I have seen this done. As more and more characters are taken into consideration, or as more and more species are described - as more species are described in a genus more characters are needed to separate each successfully or easily - it becomes less and less easy to keep the details in mind or the literature at one's finger tip.

Headed by Dr. Lodder, a committee or persons is developing a second edition of ~~XXXXXX~~ "Yeasts". Groups of genera have been parceled out to various collaborators some of whom I am fortunate to correspond with and to get identifications from. There is evidence that the generic lines of the second edition may parallel those of the first. On the other hand, workers in Hungary have broken up the larger genera on bases which until recently were of specific level. Since the classification which is used at the species level is at least partially artificial - to a morphologist, since it is based on single or combinations of physiological test characters - it might be convenient to upgrade some of these characters to the genus level as the Hungarians have done. Maybe not enough characters are being used.

In routine isolation work we obtain 100-200 cultures from a series or a project, of which sometimes 20-50 species are involved. In reaching this number of cultures some isolates have been discarded because of visual comparisons. The residue must be run through the whole series of tests and finally a partial or complete set of data are available for species identification. If a partial set of data is all that is necessary, which tests are being run unnecessarily? On the basis of some of the tests partial identifications can be made as the tests proceed. In some cases all tests must be completed before an identification can be completed or even suggested, beyond genus.

I would like to discuss these problems with you to try to determine the best course to follow with the data available to date and to be gathered in the future. The easiest way for me to do this is to go to New York at my earliest convenience and talk it over. My earliest convenience is the Christmas holidays. This is not necessarily the best time to expect to find someone in his office - he may also have an itching foot at the same time. If the Christmas holidays, Dec. 28-31, is a good time to do this then I could bring some of the material I have on hand as a basis for whatever discussions are necessary. In addition to this there are some things I would like to look up in the herbaria at the Garden. And my wife and I would use the period as a Christmas holiday.

Several years the head of our statistical section, Dr. Eugene Harris, was interested in testing some equations he saw in an article in Science. He found that an acquaintance was to be a summer student trainee in my laboratory. Since I have been building up a file on different numerical approaches to taxonomy, he thought that it would be an idea if I had the student work on a start to develop a file for his use. The difficulty was that my data sheets were not readily translatable into the information he needed. Assuming that I could get this data translation job done he would check out his equations for \$500.00. I never did find out how this idea would work out since Gene transferred to the Washington office and has not expressed an interest since. His successor, who apparently only operates the IBM machines for those who need the work, has expressed an interest but I have never gone across town to another of our buildings to try to discuss the matter with him.

I will be interested in hearing whether you will be in New York Christmas week, or not.

Very truly yours,

Wm Bridge Cooke

Wm. Bridge Cooke

November 6, 1964

Dr. William Bridge Cooke
1135 Wilshire Court
Cincinnati, Ohio 45230

Dear Dr. Cooke:

I have been mulling over your letter of October 15 in order to find the best way to answer it. I am not sure that we are on common ground, and until we can get to that point, we will need to get some of our ideas better settled.

First, I am not sure just exactly what it is you are attempting. As I read your first letter (dated October 2 and addressed to Dr. Steere), it seemed to me that the best approach would be to take all your data, code it, and punch it onto our Termatrex Card System. To accomplish that, we could supply you with the know-how of coding for the Termatrex, offer the use of our Termatrex punch machine, even provide an operator to do the actual punching. We could further provide instruction in the way to interpret what you had punched into the cards, and how to make correlations from the cards. The information to be coded, what is wanted from the information (data) would be provided by you. This is what I meant by collaborative endeavor in my letter to you. I suggested it because it seemed to me to offer the best solution at the least cost, but at this stage can you talk about money? I do not think it is as much of a problem as you have posed.

But be assured that the above paragraph does not apply to any computer application or program. Punched card reading and computer manipulation are several orders of magnitude different. Until one has fully understood this, there is little else to say. I might say that most taxonomists are confusing punch cards with computer programs, just because they haven't bothered to look into the fact that IBM makes both punch card readers and computers. That you have an IBM punch card does not mean that a computer system must follow.

"Data retrieval" from information punched into a card is another point of confusion. When you want a pure memory system, this is one way to achieve it. We have not been working with data retrieval, per se.

On reading your second letter, I find that this confusion seems to be one of our first points to clear up. I have available a computer program for classification. I will be glad to let you use this program if you have your own computer center, your own programmers, and want to do more than just sort out pieces of information. If you do not have any notion of the best taxonomic classification and wish to test several hypotheses concerning which is the best classification to use, then you want to use a computer program. If you have several hundred pieces of information for each yeast culture that must be employed to give an adequate classification, then you need a computer to help you correlate that much information--you cannot by yourself, in a practical amount of time, make all the correlations among 100 bits of information for more than a very few specimens. You need a computer, in much the same way that you must use a microscope to see the organisms. As noted in your second letter, you wish to determine among the various proposed classifications which one is most applicable, and which pieces of information are most useful in forming the classification. From the number of both specimens and data, it would seem to me that recourse to the computer is essential. The computer is, of course, only as good as the instructions given it to use. We have attempted to provide the best set of instructions possible to reflect what a taxonomist wants from his data.

I think that your fears on the cost of computer use may be a bit over emphasized. Unless you yourself get involved in the studies needed to make the computer work for you, the costs are not as prohibitive as first imagined. As with most scientific endeavor, the cost of initial investigation runs higher because many false starts are made in attempting to solve the problems. We have taken on that cost and responsibility here in our department. Others have been at work in various universities where better facilities than those available to us exist. I think that we have achieved a measure of success, however, and welcome any taxonomist who would like to avail himself of our skills and techniques. We can, if desired, provide instructions for data preparation necessary for computer runs, advice on the operation of the program (what's being done to the data), interpret the computer print-outs, etc. I doubt that you would wish to spend much time becoming a computer expert. But I also expect that you would want to know what it is that a computer does to your data, and the value of the output. For that purpose, you would be best advised to talk to a group of botanists who have gone to the trouble to find out about the computer, the mathematics necessary, and the potential implications. We can provide this in this department, and this is what we mean by collaborative endeavor. If you feel that what we intended by collaborative endeavor was to ask for large sums of money, I trust that this letter will immediately clear up that impression.

Dr. William Bridge Cooke

-3-

November 6, 1964

As far as work on your part in collaborative endeavor, you are the expert on yeast. It is therefore up to you to see that the correct information is collected from the yeast culture, that it is appropriately subdivided into categories (characteristics and attributes), that the sampling of yeasts is adequate to the business of classification, etc., etc. If you cannot provide such information, then who can? We cannot. We can suggest techniques in data gathering, notation, etc., but we would only serve in whatever capacity seems best suited.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

15 Oct. 1964

Dear Dr. Rogers,

Thank you for your letter of the 9th. I am especially interested in the last paragraph of that letter, but before going further with it I should like to explore all the angles involved. Obviously, the work with the computer is not a cheap endeavor. Personally I would not be able to go into a program of any type which involves the expense which might be involved. Since the basic work is primarily done in my Public Health Service Laboratory, it might look as if there were a nice nest egg lying around for use on such a project. But I have a distinct feeling that the powers that be are not particularly interested in taxonomy, let alone the setting up of a complete computer taxonomy of the yeasts. In monthly newsletter items, quarterly and annual reports I try to do what I hope is a sort of softening up paragraph, but little interest is evoked. In a paper which has been prepared for publication on enumeration of yeasts in polluted waters I have tried to make it clear that accurate identification will lead to more efficient reporting, but the idea of developing yeasts as an indicator of pollution has yet to catch on and until it does no one has any idea what the demand will be for ready, quick, or what-have-you, identifications. So what does the expression "collaborative endeavor" mean in terms of work and money on my part?

During the last 3 summers I have had a "student trainee". I have had her prepare a file of yeast descriptions from the literature. All but 16 entities are now included in this file. After the preliminary screening for duplication and synonymy, I find that there are about 421 species and 32 varieties or forms published (including *nomina nuda*), in both perfect and imperfect yeasts. Lodder and Kreger-van Rij's book "Yeasts", the present "bible" of the yeast identifiers, lists 176 species and varieties in 28 genera. Using the broadest interpretation of genera available at present there are only 61. Taking the characters used by Novak and Zsolt (in Hungary) for their tables, excluding morphology, there are 26. Add 14 more for morphology and there are 40. Add 40 more for the biochemical tests now popular. And I have chosen about 90 characters for use on the Unisort cards I have had set up so far. There are only a few more characters which could be used with the exception of certain inflationary types of splitting of morphological characters. To develop these properly would require a careful study of whatever cultures might be available.

The present "Yeast" book is in process of revision. This will take another year or two to complete so far as each of the separate compilers is concerned. Certain groups of genera have been farmed out from Delft to Davis, Pretoria, Lisbon, and other laboratories where yeasts are studied intensively. The new edition will include the Wickerham carbon compound tests as well as the tests used previously and certain of the standard tests will be revised. With one or two exceptions, I think I am using all the standard tests.

The biggest difficulty at present is that Everyone describing a yeast follows the older Delft scheme. A few describers add the Wickerham carbons. A few are careful about morphological detail. So at present, that is, before the revised book appears, there are only 25-40 characters which can be positively assigned to any one species. With the supplemental data obtained through the newer tests this can be brought up to my total, but this will leave some species with unanswerable questions about certain characteristics. With my total I find I have a few blanks. I have ignored certain characters when they do not apply in general to the groups of yeasts with which I am working. These can be inserted when or if necessary.

On the other hand, in the area of water pollution study it may be that only a few laboratories will want to bother with the lengthy process of getting all the necessary data from a culture. For these few laboratories the maximum available information may be useful. For others the minimum information may be all with which they would be able to work. But with the increasing interest in yeasts and the consequent discovery of additional species in habitats previously untapped, or as a result of increasingly fine slicing of bits of data, the largest amount of data available may be more practical.

Any comments will be appreciated.

Very truly yours,

Wm. Bridge Cooke

Wm. Bridge Cooke

October 9, 1964

Dr. William Bridge Cooke
1135 Wilshire Court
Cincinnati, Ohio 45230

Dear Dr. Cooke:

In reply to your letter of October 2 addressed to Dr. Steere requesting information and potentialities of developing a technique for the classification and identification of yeasts, I have the following suggestions.

We have here in our office a new type of punch card system (the Termatrex) that permits the entry of information up to 10,000 bits of information per card and which can be very quickly and rapidly searched using a light box rather than needles as is true in many card systems. We think that the Termatrex offers a very great opportunity for keeping up with the kinds of taxonomic information that seem to have been collected by you in your laboratories. We think that it further offers an additional opportunity if you would eventually care to use the information on these cards to make a classification scheme. The information from the Termatrex cards is readily transferable to IBM cards which may then be run into a computer program which is now available through our department. This computer program is intended to make a taximetric analysis of the data and end up with a series of "clusters" which will represent the species, genera, etc.

We would be pleased to consult with you concerning the possibilities of collaborative endeavor. If you see any merit in such a collaborative endeavor, we would be happy to talk to you in considerably more detail about it.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

CC: Dr. Steere

October 6, 1964

Dr. William Bridge Cooke
1135 Wilshire Court
Cincinnati, Ohio 45230

Dear Dr. Cooke:

In response to your letter of October 2, I can say only that I have a great interest in computer taxonomy and information retrieval, but very little technical knowledge about how it works. For my money, this is a technique that will replace many of what we now think of as standard procedures in taxonomy. However, since I do not know the technical or mathematical side of the thing, I am forwarding your letter on to Dr. David Rogers, in the hope that he will have something constructive to offer you in the way of advice on how to proceed for the maximum benefit and the least work.

Sincerely yours,



William C. Steere
Director

WCS/gd

cc: Dr. David Rogers

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

2 Oct. 1964

Dear Dr. Steere,

For some time you have been interested in computer taxonomy or information retrieval. I came to this conclusion because of your sponsorship of the IPIx program and because of your participation in one of the programs at Edinburgh. I heard David Rogers' paper at that symposium and looked forward to seeing his paper in "Bioscience". Now that I have seen it I want to learn more about the techniques and how they apply to a problem I have.

When I started to isolate fungi from polluted waters and sewage I found that yeasts were also present. I isolated them, too, more or less regularly, until I had a number of cultures accumulated. Subcultures of these were sent to Dr. Herman Phaff at Davis for identification. You may be aware of the fact that he is the only mycologist in the country who identifies yeasts of all groups (when he has the time). As time permitted he or his students worked on my cultures but by 1957 no results had been reported. On a trip to survey our Milk and Food Laboratory, Dr. Emil Mraz suggested that one way to solve this dilemma was to go out to Davis myself and learn the techniques. This I did in September, 1958. Since then I have studied some 5000 cultures. One of Phaff's students, Moshe Shifrine, several years ago suggested a punch card technique for entering information about a yeast culture. I amplified the punch card Moshe suggested a bit and now have a file of about 4000 of these with about 90 types of information recorded for each of as many isolates. Included among these cards are "master" cards for more than 95% of the published species of yeasts. The remaining species of which we have knowledge, are in still inaccessible literature. Phaff's "Yeast" News Letter is useful for this.

Yeast taxonomy is a hybrid between bacterial and mycological taxonomy. It is based in part on morphology, in part on physiology. Morphology and cytology play a part in setting up perfect genera, morphology in imperfect genera, and physiology in setting up species in all groups. In fact, some species are separated only on the basis of one reaction or another. This is the situation in the Delft scheme, the technique used by all "free" world mycologists. The Kudriazev system is not much different. In Hungary, Novak and Zsolt have separated some genera into 4 on the basis of certain physiological reactions, but the nomenclature is a mess because of ignoring certain aspects of the International Code.

To try to determine whether my cards can be used effectively I need further indoctrination into the punch card techniques. In addition, I would like to know whether certain types of data are valid for use at the species level, and the extent to which certain groups of species can be maintained at the generic level. This work does not need to conflict with the proposed second edition of the Lodder work now in preparation. However, that will become the "bible" of yeast identifiers, just as the 1st edition is now, and this kind of thing sort of results in an ingrown discipline within Mycology, assuming that a good mycologist will admit he is a mycologist.

I hope that there is some way that the data I have on hand can be put to use.

Very truly yours,

- Taxonomy Laboratory

June 13, 1966

Dr. Wm. Bridge Cooke
1135 Wilshire Ct.
Cincinnati, Ohio 45230

Dear Bridge:

Although I do not plan to be here during the time that you will be, I suggest very strongly that you come along anyway because Henry Fleming will be here and will be able to go over your work as well or better than I could anyway. So please make arrangements to stop by, we will be happy to see you.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

1135 Wilshire Ct.,
Cincinnati, Ohio 45230

3 June 1966

Dear Dr. Rogers,

Vacation time approaches and we may have the same period in which there is no need to go farther with this idea.

My vacation starts the evening of July 1 and we will start west then. We will go directly to Lassen Volcanic National Park and on the way it seems that it might be possible to spend one night in Fort Collins. It just happens that that night is the night of Monday, July 4. I can have with me the master set of punch cards about which we have spoken and written, for all yeast species published up to last Sept. 1. And I can have sample sets of other or unknown species.

If it is not convenient to do this please let me know so that I can forget work for the duration of the vacation. If it is convenient and practical to discuss some aspects of the problem please let me know.

We plan to go to Rocky Mountain National Park from Fort Collins and possibly collect a fungus or two on our way across the Park. This would be on the fifth.

Very truly yours,

Wm. Bridge Cooke

- Taximetrics Laboratory

November 3, 1966

Mr. Lauren B. Doyle
Language Processing and Retrieval Staff
System Development Corporation
Research and Technology Division
2500 Colorado Ave.
Santa Monica, California 90406

Dear Mr. Doyle:

Please excuse my long delay in answering your cordial letter of September 21. First of all, you will note that I and my staff have moved from New York (in September, 1965) and are now a part of the botany department of Colorado State University. The forwarding of your letter caused some loss of time in answering, but cannot be blamed for all of it. We have been extremely busy this fall seeking research funds to continue our work, and this is a nontrivial task because the education of the granting agencies to "see the problem" as we do is not easy. It is too bad that I did not attend the meetings of the ADI, an organization that I should be closer to, but you can understand, I hope, that one's allegiances shift slowly from one group of scientific societies to another.

It is certainly true that I do not have the knowledge, or skill, to follow all the various developments for information retrieval wherever these developments are made. Your area, for example, does not usually get indexed in the various journals I read. Good thing that the NSF document did put us in contact.

Frankly, we did not see the value of our work on the broad general scale that other (including you) have seen it. We were (at first) only concerned with biological classification, and we still stick to that as our primary source of inspiration. We have become aware of the really great potentials for general information retrieval problems which can be derived by careful examination of the biological classification process. Though the biologists generally use their classification tools in an intuitive way, we have found many procedures which can be put into logical sequences, rules and algorithms developed, and computer programs written. There are several more programs running that we have not yet reported on, largely for the want of sufficient personnel, time and money. All these programs are generated in the general framework of the classificatory disciplines of biology (classically called taxonomy), and all seem to have merit for the general field of literature classification, and eventually for information retrieval.

November 3, 1966

I am sending a flow-chart of taxonomy which we put together mostly for our own benefit, but reading of this chart will give you some feeling of the various bits and parts of the total biological classificatory effort. Each of the areas has some special role to play, and we hope eventually to design programs for each of the areas, each merging with the others. However, to design the programs, we have much mathematical work to do, as well as much biological analysis to test the effects of the mathematical logic. My team is integrated with this type of endeavor uppermost in our minds. The team includes two biologists, one mathematician and one programmer. The latter two are only half-time members, and we need more of their time to really do the job I know has to be done.

Our efforts will take much time to spread through the various areas of science where they will be recognized as meritorious outside the strictly biological disciplines. We have found acceptance within the biological milieu in only a few places, but more and more people are coming into the fold. I trust that through efforts such as yours we can hope for wider acceptance sooner. I am sending you some papers you may find interesting, and have your name on our mailing list to receive other works now in the mill.

Thank you for your initiative in writing. I hope we can get together soon.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch



SYSTEM DEVELOPMENT CORPORATION

September 21, 1966
L-13183

Mr. David J. Rogers
Principal Investigator
NIH Classification Project
New York Botanical Garden
Bronx Park, Bronx, New York 10458

Dear Mr. Rogers:

I have just been reading Number 14 of Current Research and Development in Scientific Documentation published by the National Science Foundation and find that your project in computerized classification is the most interesting and the most nearly related to mine of the many projects described therein. It sounds like we should be communicating, so here goes.

I am enclosing a research document, SP-2516, "Breaking the Cost Barrier in Automatic Classification," describing my work in automatic classification; this speaks mostly about document classification, but the basic method is directly applicable to any listed qualitative data, probably including data of the sort you describe. This may serve to open the dialogue. Perhaps you will attend the Annual ADI Meeting in Santa Monica next month, in which case I would certainly like to meet you. If you will not be able to make it, I would be interested in further written communication or an exchange of reprints, as you choose.

In the middle of the enclosed document is a digression away from the strictly documentation uses of this classification method. My interests in classification are not confined to the library problem, even though 95 percent of the work I do is in that domain. Examples discussed in the enclosed paper are pictured as analysis of soil samples in agriculture, quantitative study of health-relevant data, and analysis of individual cases in special problem areas, such as delinquency.

I have long emphasized the problem of describing the classification (or network of clusters) after obtaining it. It

Mr. David J. Rogers

-2-

September 21, 1966

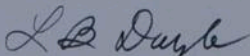
L-13183

sounds as though this would also be one of your interests. Some of the problems you explicitly mention are ones that you may find also mentioned in SP-2516. It may happen that if any of these problems are particularly central in your work, I might have some further material available on it, or it may be that you have further material of special interest to me.

I notice you have access to a CDC 1604, which sounds like a reasonable machine for this kind of work.

I shall await your return message with interest.

Sincerely yours,

A handwritten signature in cursive script, appearing to read "L B Doyle".

Lauren B. Doyle
Language Processing
and Retrieval Staff

LBD:pf
Enclosure

- Taximetrics Laboratory

November 3, 1966

Dr. Geoffrey H. Ball
Radio Systems Laboratory
Stanford Research Institute
Menlo Park, California

Dear Dr. Ball:

Enclosed are the reference materials requested in your recent letter. I was pleased to receive your review of the clustering procedures.

You will find in the enclosed reprint an entirely different clustering procedure developed by this laboratory. The graph clustering method is described more from its biological standpoint than from the basic mathematical descriptions. This aspect is covered in a separate publication we hope to have out in the near future. This paper will be of great value to you, and I shall add your name to our list to receive this descriptive material when it appears.

With regard to your form for specification of any cluster seeking technique, I fear that I cannot do it justice without some descriptive literature to accompany it. Precisely, I am not certain what to say about the first mentioned word "data", nor how to say anything about "operations on data." I think you need some sort of additional sets of directions to fill in the various categories.

Would you be so kind as to take our enclosed paper on graph-clustering methods, fill out your form as an illustration of the required information on the form. From that point, perhaps I can decide what is needed.

With regard to your request to CDC for a documented program, we now have a very detailed, documented listing, flow chart and instructions for input, written for an IBM 7044, mostly in FORTRAN IV. If you are interested, I think we can do better than CDC's charge for that service.

Hope these comments are useful to you.

Sincerely,

David J. Rogers
Professor of Botany

Enc.



STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA

14 October 1966

Dr. David J. Rogers
Taxonomy Laboratory
Colorado State University
Fort Collins, Colorado 80521

Dear Mr. Rogers:

I am in the process of preparing an analysis of and survey on cluster seeking techniques (including consideration of measures of similarity or association, and to some extent preliminary processing of the data).

I have found your work valuable and I am interested in having copies of or references to work done by you and your associates on these topics that I do not have copies of. I have attached a list of references to that work of yours of which I have copies.

In addition, I have enclosed the form for specification of any cluster seeking technique that I have used in my survey and that I plan to recommend be included by the authors in all future papers on cluster seeking. The purpose of including such a summary would be to clarify the objectives and methods used in each new technique proposed. I would like to improve the form I have used or to propose several alternate forms. Therefore I would be very appreciative of your comments as to improvements that should be made in this form.

Thank you very much for whatever help you may be able to give me in these matters. (I have enclosed a copy of a short survey I did last year for whatever use it may be to you.)

Yours sincerely,

Geoffrey H. Ball, Ph.D.
Senior Research Engineer
Radio Systems Laboratory

GHB/jb

P.S. I would greatly appreciate response within a month of the date of this letter if it would be at all possible.

*CDC didn't have the documentation
and I didn't have \$730.00 to pay
for their time to prepare it. Thanks for
your help.*

Stanley Reiter and Gordon R. Sherman, "Allocating Indivisible Resources Affording External Economies or Diseconomies," International Economic Review, Vol. 3, No. 1, pp. 108-135, (Jan. 1962).

A. Rescigno, and G. A. Maccacaro, "The Information Content of Biological Classifications," Fourth London Symposium on Information Theory, pp. 437-446, (Aug.-Sept. 1960).

Salomon Rettig, Rawson, Knoblock, Pasamanick, "A Multidimensional Computer Analysis of Some Complications of Pregnancy and Their Neuropsychiatric Sequelae," (1962)

Bernard Rimland, "Multidimensional Scatterplotting: A Graphic Approach to Profile Analysis," J. of Applied Psychology, Vol. 44, No. 6, pp. 404-406, (1960).

Bernard Rimland, "On the Objective Diagnosis of Infantile Autism," Conf. on the Role and Methodology of Classification in Psychiatry and Psychopathology, (July 1965).

David J. Rogers and Henry Fleming, "A Computer Program for Classifying Plants, II. A Numerical Handling of Non-numerical Data," Bioscience, Vol. 14, No. 9, (1964).

David J. Rogers and Taffee T. Tanimoto, "A Computer Program for Classifying Plants," Science, Vol. 132, (21 Oct. 1960).

James F. Rohlf, "A Randomization Test of the Nonspecificity Hypothesis in Numerical Taxonomy," Taxon, V. 14, (Oct. 1965).

James F. Rohlf and Robert R. Sokal, "Coefficients of Correlation and Distance in Numerical Taxonomy," The Univ. of Kansas Science Bulletin, Vol. XLV, No. 1, (June 7 1965).

FLOW DIAGRAM:

DATA // Operations // Partitioning (DISTANCE MEAS.)
 on Data // Calculations Performed

CLUSTERSEEKING CLASS:

Evaluation // Modification of Partition // Characterization
 of Partition // What is Changed, How Much // of Final Group
 (Criterion) // (degrees of freedom)

$$\{x_i\}$$

Criterion Maximized (if any);

Sequence
 DEP/INDEP

Feedback:
 Iterative:

Quantity of Calculation:

Quantity of Storage:

RIA

Lineal

INFO ASSUMED AVAILABLE

Specific Comments (Invariances, Assumptions, Obvious Limitations)

Experiments Reported

Program Exists

Convergence:

Quotes (pp)

Goal, Applications
 and Viewpoint.

- Taxonomy Laboratory

August 15, 1966

Dr. Geoffrey H. Ball
Radio Systems Laboratory
Stanford Research Institute
Menlo Park, California 94025

Dear Dr. Ball:

Your letter of 3 August was forwarded to me from New York. Our present address is given above.

Since the program we devised was supported by NIH funds, you are welcome to use them. Our only request is that any use of the programs acknowledge the source.

When we left New York, the program was still "in flux." While it did indeed do the job described to you by Mr. Mabie, we had not settled on any particular flow chart, and many of the routines were called at option, and were controlled by the programmer. We never received any documentation, and would be happy to have that program, as it stood, described from the programmer's point of view. If you can get this from Mr. Mabie, we would like to have a copy of it.

There are several other facets of the output that Mr. Mabie did not include, i.e., each cluster was listed with the minimum similarities to be found amongst the objects of the cluster; the amount of "connectedness" amongst the objects in each cluster at each partition (the number of connections existing versus the number possible), and a measure (the most) indicating the separateness of a cluster from its neighbor. Several other items are included, all of which give information concerning the objects and their clusters.

Mr. Mabie indicated in his letter to you that further work with the output is done by hand. We still continue to follow a manual routine to "graph" the output, although we hope that someday this activity will be programmed. I am sure that Mr. Mabie could have developed a program for this purpose, had he been given the opportunity. Unfortunately, our funds, which have never been plentiful, petered out before we could contract with CDC to have this job done.

We have reprogrammed the CDC program for the IBM 7044 and 7094 available to us here. This has been fully documented, and will soon be submitted to SHARE, for the two machines. In this job, a considerable

August 15, 1966

amount of former manual control of the program has been eliminated, and considerable increase in the options for varying the program have been included. If you are interested in these, please let us know.

One word of caution should be given. The program is extremely sensitive to the format of the input data, and it is no small chore to select and prepare your data. Our whole effort in designing this program was directed towards classification of biological objects, and most of the objects we wish to classify are described by qualitative information. These pieces of information (called characters in the biological milieu) are generally considered as discrete, though some types of data are continuous.

We have found that workers interested in use of this program are generally benefitted by spending a few hours with us, getting the feel for the program's operation, objectives and philosophy, as well as instructions for data preparation. You can probably save a considerable amount of time if you are aware of this problem. If you wish to do so, we can arrange for a visit here, where you can get definite instructions from our programmer, mathematician, and from the biological standpoint.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

cc: Henry C. Mabie



STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA 94025

3 August 1966

AUG 12 1966

Dr. Rogers
New York Botanical Gardens
New York, New York

Dear Dr. Rogers:

I have become very interested in the computer program described by Mr. Henry Mabie of the Computech Data Center, Control Data Corporation, which was written for some of your recent work. We would like your permission to receive a copy of the program and the program documentation.

We will be able to modify the program for operation on our CDC 3100. Our installation consists of the computer with 3 tapes, disk storage, and a display tube controlled by the SCOPE monitor system.

Enclosed are copies of correspondence between Mr. Mabie and Mr. Rivinus of CDC, and myself.

Thank you.

Sincerely yours,

Geoffrey H. Ball, Ph.D.
Senior Research Engineer
Radio Systems Laboratory

GHB:el
Enc.



June 7, 1966

Mr. Geoffrey H. Ball, Ph. D.
Senior Research Engineer
Radio Systems Laboratory
Stamford Research Institute
Menlo Park, California

Dear Mr. Ball:

Thank you very much for sending me a copy of your paper on Data Analysis in the Social Sciences. I see that you have included a description of the 1960 work of Rogers and Tanimoto, but do not refer to the more recent work of Dr. Rogers and his staff, for which I wrote the computer program.

This method of clustering starts with the computation of all pairwise similarities. These are then sorted in descending order, and the resulting table is used for building up the clusters.

All pairs having the greatest similarity are selected as the clusters at the first level. If any object occurs in two or more pairs at this level, all such pairs join to form a larger cluster.

When clustering is complete at one level, the program moves to the next level. There are now four possible cases:

- 1) The pairs found at this level do not contain members of any previously formed clusters, and so constitute new clusters in themselves.

- 2) One member of the pair is already included in a cluster, but the other is not. The unclustered object is then added to the cluster to which the other member of the pair belongs.
- 3) Both members of the pair already belong to the same cluster. No change in clustering occurs.
- 4) The members of the pair already belong to different clusters. The two clusters are then combined into a single cluster.

This process continues until all objects are combined into a single cluster.

As the clusters form, the sequence in which the objects are listed is maintained as follows:

When a single object is added to a cluster, it is placed at either the beginning or the end, whichever is nearest to the location of the other member of the pair to which it belongs.

When two clusters are combined, they are placed in such a way that the members of the pair which caused their combination are brought as near to each other as possible. To accomplish this, the sequence of one of the clusters may be reversed, but may not be changed in any other way.

After clustering is complete, a chart is drawn on the plotter, which helps the user to determine the relative significance of the clusters formed.

A further investigation, which has been done, as far as I know, only by manual methods, concerns the structure of the clusters at each level. For example, three extreme cases might be a chain, in which each object is joined only to the preceding and following objects, a wheel, in which each object is joined only to a single central object, and a tight circle, in which each object is joined to every other object. A computer program for making this investigation could readily be written.

Sincerely yours,

Henry C. Mabie

Henry C. Mabie
Senior Programmer Analyst

HCM/emc



July 22, 1966

Dr. Geoffrey H. Ball
Radio Systems Laboratory
Stanford Research Institute
Menlo Park, California 94025

Dear Dr. Ball:

Thank you for your recent letter to our
Mr. Henry Mabie regarding computer programs written for
Dr. Rogers of New York Botanical Gardens.

This work was done on contract with our clients
so we are a little reluctant to release something that is
really not ours to give. However, if we receive authorization
from New York Botanical Gardens, we would be delighted to
comply. Perhaps you should first contact Dr. Rogers.

Would you want these programs modified for operation
on your CDC 3100? What is its particular configuration?

We look forward to hearing from you further.

Sincerely

CONTROL DATA CORPORATION

A handwritten signature in dark ink, appearing to read "Willis M. Rivinus".

Willis M. Rivinus
Assistant Manager, Marketing & Sales

WMR/jet

Southern Illinois
University

CARBONDALE, ILLINOIS 62901
October 24, 1966

Department of Botany

OCT 27 1966

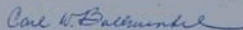
Dr. David J. Rogers
Taxonomy Laboratory
Department of Botany
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers:

Thank you so much for the reprint and mimeographed material which you sent last month. I have read through the materials and I am sure that I will find them most helpful. I have been on a trip to Mexico and now find other matters most pressing, so I shall have to postpone working on the use of the program. I have thought of using the program in conjunction with a comparison of species of Acerates within the subgenus and with other Asclepias species. I may also want to consider its use in a study of South American Asclepias.

Thanks again.

Yours truly,



Carl Bollwinkel
Graduate Assistant

bjh

- Taxonomy Laboratory

August 24, 1966

Mr. Carl W. Bollwinkel
Department of Botany
Southern Illinois University
Carbondale, Illinois 62903

Dear Mr. Bollwinkel:

I am sending you a flow chart for a program developed since the one that appeared in BioScience. This program is far superior to the one described in BioScience, and should be more satisfactory for clustering purposes. Discussion of the biological significance is given in the enclosed reprint from Systematic Zoology. I am also sending along a mimeographed paper "A General Method of Taxonomic Description, etc."

The last mentioned paper is sent because we feel that a thorough understanding of the preparation of taxonomic data is a critical element to be understood before attempting to use the clustering method. I might say this applies no matter what computer program is employed. We have written the computer program in FORTRAN IV, but you will see places in which machine language is employed. Machine language is required in order to get sufficient room in the internal memory units of any computer that we know. This will require that program be done by a relatively experienced and competent programmer.

These items, though in mimeograph form, have been submitted for publication and are, therefore, not for general distribution.

I will be interested to know of your proposed use of this program and would like to hear from you when you have had an opportunity to digest the contents of the papers sent. Please do not hesitate to write for assistance.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

DEPARTMENT OF BOTANY

Southern Illinois
University

CARBONDALE, ILLINOIS 62903

August 16, 1966

AUG 19 1966

Dr. David J. Rogers
New York Botanical Garden
Bronx Park
New York 58, New York

Dear Dr. Rogers:

The work you have been doing with computer programming for the classification of plants is of interest to me. In your BioScience article, "A numerical handling of non-numerical data", you indicated that you were working toward the development of detailed flow charts which would be needed by anyone interested in computer analysis. If these charts or a source deck listing is now available, I would appreciate the opportunity to study them and perhaps make use of them in planned taxonomic investigations. If you have publications since the BioScience article I would appreciate reprints.

Thank you.

Yours truly,

Carl W. Bollwinkel

Carl W. Bollwinkel
Graduate Assistant

bjh

*Hi - Greetings to everyone
there. Hope all is well -
Mange
Bollwinkel
m. by mistake*

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

JAN 24 1966

January 20, 1966

Dr. D. J. Rogers
Taxonomy Lab
Department of Botany
Colorado State University (Ft. Collins, Colo. 80521)

Dear Dave:

Thanks for your letter of the 17th. Things are what I thought they would be; it was mostly Dr. Wagner's letter which made me think I was wrong about the arrangements. Everything is ship-shape for our collaboration, but, as you can see, many details remain to be solved in continuing my work with Dr. Wagner. Yesterday my programmer in Chicago wrote to say he might be spending eight months on the east coast, so that a computer in Chicago would do us no good. Basically the problem is getting the programmer next to a computer. I can't see the programmer having to go through Dr. Wagner or one of his students to get to a computer, even if money is saved. There would be too many chances for mixups. It would be better to let the thing drop until the Smithsonian gets a better computer, and I could feed the machine myself.

There has been considerable high-level talk, principally from President Johnson's Committee on Scientific and Technological Information (or something like that — it's called COSATI for short), about setting up a library-based information storage and retrieval system. On its own, the automatic data processing committee in the Museum of Natural History, Smithsonian Institution, has been discussing automated records keeping plus problems that COSATI has been concerned with. The NSF's Publications Support Program, which has financed Mr. Gould, also has a continuing interest in these sorts of things. The only definite thing down here has been our hiring an ADP man to look after our System 360 (IBM 1440). In addition, the Science Information Exchange has a storage and retrieval system. The new man wants to get rid of SIE's computer, and use the money to add to the System 360, which would give us our current capability of the 1440 for fiscal, an adequate machine for the SIE, plus capability about equal to a 7090 for the Smithsonian scientists. Needless to say, this would solve all my computer problems — but no one knows when and if this will happen.

Sincerely,

David B. Lelinger
David B. Lelinger
Associate Curator
Division of Ferns

TO: D. Leffinger
Smithsonian

Taxonomy Lab. Jan. 17

Dear Dave:

Let me try to unconfuse the present confusion.

We can only run your data on our program. We cannot run your programs.

Now that I have laid it out, let me say why. I hope this explanation doesn't add another factor of confusion. There is a very good chance that we would have to reprogram your programs for the specific configurations of the 7094 at the Western Data Processing Center at UCLA, the installation we use. This institution is one of the western universities tied in to the WDPC by a teleprocessing line. We have to pay only for the cost of the ~~XXXXXXXXXX~~ transmission of the data, but pay nothing for the actual computing machine time in Los Angeles. Because of the contract terms, we can only work with our own programs.

In addition to these contractual complications, there are also the problems that we'd have in doing the reprogramming, even though you might have everything in FORTRAN IV. If you and your programmer were not around to tell us what was going on, it would be a long drawn-out process, as I'm sure you are well aware.

May I ask why you can't continue to use your program at U. Mich? Can't Herb get on there without it costing him anything? Sounds kind of funny to me that a program developed there can't be run there.

But Before I get completely off the topic, be assured that we will run your data with out any charge to you.

Now that I've said that, and hope that you are still interested in collaborating with us, let me ask you something entirely separate from it. I heard via the grapevine that someone at the Smithsonian is interested in running a retrieval program for nomenclature. Have you any information on who this might be, and is it any more than scuttelbut? I'd be interested to know. I hope that something will be done, but that it will not be the same unfortunate type of operation that has slowly died at NYBG. I can think of no better place for such work than the Smithsonian, provided the right people are involved.

Looking forward to receiving your input.

Sincerely,

JAN 17 1966

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

January 12, 1966

Ans 1/18

Dr. David J. Rogers
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dave:

Thanks for your letter of January 3. Dr. Wagner also just wrote about his conversations with you at Berkeley. I was surprised about the availability of computer time, but am a little unsure of just what you meant. Did you mean we could run my data with your programs or also my data with my own programs, once they are converted to FORTRAN IV? Dr. Wagner is anxious to know about this, as he would like to avoid having to handle input/output himself; currently my programmer, who is in Chicago, is checking up on computer time there. As for my programs, once they are translated from MAD into FORTRAN IV and debugged, they should all run in much less than an hour (on the IBM7090). I think I can talk Dr. Wagner into spending about \$250 on computer time.

Just now I am occupied evenings with preparing copy for the first number of the Fern Journal, but that should be off to the printer in a week or two, and then I'll get back to recoding data.

There is some possibility that SI will eventually have an IBM-360 system with the capability of an IBM7090. The new chief of the data processing division wants to get our Science Information Exchange to give up its computer and use a wire into the small 360 system now in the "castle." If he succeeds in promoting this, the money saved on SIE's rental would lease sufficient more hardware for the present 360 system to make it the equal of an IBM7090. So here's hoping!

Best regards,

Sincerely,

Dave

David B. Lellinger
Associate Curator
Division of Ferns

- Taxonomy Laboratory

January 3, 1966

Dr. David B. Lellinger
Associate Curator
Division of Ferns
Smithsonian Institution
United States National Museum
Washington, D. C. 20560

Dear Dave:

Thank you very kindly for sending me a copy of your dissertation, or at least the relevant parts of it. I am a little beyond where I can use it specifically in the paper that I was hoping to do. But, I am glad to have it for a full exposition of your ideas, anyway. They will certainly come in handy.

I had a chance at the Berkeley meetings to talk to Herb Wagner and he told me that you were definitely interested in going ahead with us on the collaborative project. He said that you hoped that you might put in for funds from the National Science Foundation to foot the bill. I told Herb that the situation has changed considerably since we left New York and that it was very likely that we could run the material here if it didn't get too big, without any cost to you whatsoever. If this is the only stumbling block that you have, that is, the cost, then I think that we can probably make it go here. I am not sure yet, but I think that the funds would not be at all unreasonable. Probably less than one hundred dollars. I am not even certain how much this charge will be because we have not yet had any experience with running our programs here at C.S.U.

As you may know, C.S.U. is one of the Western Universities that works with the Western Data Processing Center at Los Angeles. We have a teleprocessing line from here which shoots the data right into the 7094 in Los Angeles and we get the print out back sometimes within 24 hours by the same line. We have still to complete the transfer of our programs from the Control Data Corporation's machinery that

we used in New York to the I.B.M. hardware in Los Angeles. However, this is progressing nicely and I think that we should have ourselves in operation for our Clustering Methods and so forth within the next two months.

I hope you can go ahead and prepare your data now with the relief that you won't have to pay such a wampun big bill as you might have had to do if we were still on the east coast.

Best wishes for the new year.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/pam

NOV 12 1965

Clarke Hall
Box 41
Roseau, Dominica
BRITISH WEST INDIES
November 17, 1965

Dr. David J. Rogers
Dept. of Botany & Plant Pathology
Colorado State University
Fort Collins, Colorado 80521
U.S.A.

Dear Dave,

Thanks for your letter of November 2, which reached me today, having been forwarded from the SI on an irregular basis.

Right now, as my address will indicate, nothing is happening. Since September 9 I have been down here collecting ferns. I'll be back in Washington by mid-December, and will continue recoding my present data. Wagner and I are getting together to talk over continuing the work on December 18; we may approach the NSF for funds. My present method is in need of improvement, as we discussed at NY awhile ago. But whether anything more comes of it or not, I'll recode my data for your method, as I know the results will be of interest to you as well as to us.

I have nothing down here to send you, unless I can get the use of an office copier in the HQ of Archbold's plantations. He has put up part of the money for the Smithsonian Biol. Surv. of Dominica, and we are quartered at one of his plantations. If nothing works out with the copier here, I'll send you a xerox of the relevant parts of my dissertation by mid-December. Wish I could do more, but am helpless in this backwater. The only thing published is an explanation of Wagner's method in Benson's grey-covered Plant Taxonomy book. Background is in Danser's 1950 article in Bibl. Biotheoret. D, 4(3).

Eventually University Microfilms at Ann Arbor (no connection with the University of Michigan) will publish, as they say, my dissertation on microfilm. Xerox copies will be available. I'll let you know when I find out if it is ready in that medium. Actually, it may be out now-- I don't know.

How do you like Fort Collins. Bit of a change from NY, isn't it. Perhaps you have met Roger Peterson of the Forest

411
- Taxonomy Laboratory

November 2, 1965

Mr. David B. Lellinger
U.S. National Herbarium
Smithsonian Institution
Washington, D. C.

Dear Dave,

I wonder what is happening to our collaborative project, if anything. I should like to know if there is any advance towards the culmination of a cooperative study.

You know, of course, of our move out here. We are more or less settled now, and are in the process of rewriting our programs from CDC Fortran to Fortran IV for the 7090. But in the meantime, we are trying to pull together the loose ends from our various collaborative endeavors, of which yours is one. Let me know what the status of your project is.

Do you have any spare copies of the discussion and theoretical model behind your work for the study of evolution in ferns? I have in my file your flow chart that was made up for the symposium in St. Louis in 1962. I also have in my file Herb Wagner's little mimeographed blurb on the divergence index method for expressing evolutionary data. Other than these two unpublished items, there is no discussion that I can find on the ground-plan method for evolutionary studies. I would be very pleased if you could send me a copy of your discussions, and also if you could cite for me some published works discussing these. I need this rather quickly inasmuch as I am writing a chapter in a book entitled, "The Use of Computers in the Study of Evolution". Naturally, my deadline is upon me, and I must have this information as quickly as possible. Let me hear from you as soon as is convenient.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ec

October 22, 1964

Mr. David B. Lellinger
Division of Ferns
Smithsonian Institution
U. S. National Museum
Washington, D. C. 20560

Dear Dave:

I have had your letter of the 6th around for some time. The question, it seems to me, that bothers you is how much information you have to have, and the answer is that I don't know.

All I can say is that with the Halimeda study we got a classification using about 50 characteristics and 280 odd attributes. For my cultivars of Manihot we have about 15 characteristics and about 86 attributes. Recall that this is in a group of cultivars not formed into any kind of cluster; that is, the sub-species variants and their levels of differentiation are different from those of the Halimeda where we were working with taxa that had already been formed. However, my opinion on that is that you use what information you want or need and don't try to squeeze out information just to fit a statistical need. Therefore, we say that if you don't have so many, forget it.

We are now in the process of making a study of the methods, the logic and the program for character selection and analysis to accompany any clustering of objects. This should shed some light on the subject of your question. We don't know the answers yet, although taxonomists have for years said that they knew this aspect and some still think they know.

I hope this answers your question. Let's keep up our correspondence and discussions. As soon as our reprints are out, I will send you one.

Sincerely

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

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

October 6, 1964

Dr. David J. Rogers
The New York Botanical Garden
Bronx Park
Bronx 58, New York

Dear Dave:

Thank you for your letter of the first, and for the reprint of your Port-A-Punch paper. It makes me good and mad that I didn't use such a device when scoring characters in cheilanthoid ferns. Gould assured me that the devices were no good, caused mistakes, etc., etc., etc. Instead I spent hours keypunching and checking, and had to spend a lot of time at the desk calculator doing $r_{x,y}$'s between specimens to see which ones to use. Well, live and learn!

May I come back to the data of the Manihot and Halimeda studies again? Considering the fact that the programs worked so well, I am interested in the fewer data that you used. For instance, Michener and Sokal used 122 characteristics, and, I suppose, upwards of 500 attributes. In my last letter, I forgot to ask how many characteristics were used in each study. Judging by the 86 and 286 attributes, the number of characteristics must have been far smaller than 122.

Thereby, in my opinion, hangs the tale. If you hate to take data as much as I hate to take data, this reduction in the number of characteristics and attributes contributes greatly to one's willingness to use a taximetric system. It seems to me (correct me if I'm wrong) that your success in this reduction is because you have chosen quite carefully the characteristics and attributes because they are likely to have taxonomic meaning, rather than be merely random manifestations of the phenotypes, and without any obvious taxonomic meaning. Thus the better the way in which we could "test" for meaning, the fewer the characteristics we could use and still come up with meaningful, evolutionary systems.

With best regards,

Sincerely,



David B. Lellinger
Division of Ferns

October 1, 1964

Mr. David B. Lellinger
Smithsonian Institution
United States National Museum
Washington, D. C. 20560

Dear Dave:

Thanks for your letter of September 29. For Manihot
we had 200+ objects and 86 binary attributes. For
Halimeda we have 28 objects and 286 attributes.

In response to your question on the S_{ij} and the S_r ,
they are the same.

As soon as the reprints on the Bioscience paper are
received, I will be sure to send you one. Meanwhile,
I am enclosing the AIBS article on the Port-A-Punch
which you also requested.

Sincerely

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF
Enclosure

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

September 29, 1964

Dr. David J. Rogers
The New York Botanical Garden
Bronx Park
Bronx 58, New York

Dear Dave:

May I ask how many characteristics and how many binary attributes were used in both the *Manihot* and the *Halimeda* studies? Also were the numbers of objects 300 and 28, respectively?

The heat is on here to finish my thesis; now I'm working mornings, as well as evenings, and progress is good. Had lunch with the Director recently and was, as another colleague who is in the same boat said, "wined, dined, and roasted!"

One question about your most recent paper. What is the difference between s_{ij} and S_r . It looks like they are the same, or have I missed some essential point?

I am starting to regret that we wrote almost everything in MAD. It is such a local dialect. FORTRAN at least might be usable on some of the government computers. On the other hand, we hope to have some kind of computing center here by the end of the decade, and perhaps we will make MAD one of the languages. It seems to me to have great advantages since it can be read with so little training by non-programmers.

When I saw you last I forgot to ask if you could send a copy of your 1963 article in AIBS Bulletin. Are there any left on your shelf?

Many thanks. Best regards to all,

Sincerely,



David E. Sellinger
Associate Curator
Division of Ferns

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

DEC 27 1965
52-1

December 21, 1965

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

Dear Dave:

Here are what seem to be the relevant parts of my dissertation for your purposes. I hope they have not come too late to be useful to you. Unfortunately, I was not able to have a copy made in Dominica, where your letter reached me.

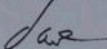
Last weekend I went up to Ann Arbor to talk with Dr. Wagner and to march at graduation. We decided to continue the work for a little while. If the improvements we plan work out, we may apply for an NSF grant together.

We plan to convert my character states to binary for use with your programs and to convert our own programs from MAD to FORTRAN and switch from rx.y to sij. Also, I will code some additional species to fill in some of the gaps.

Naturally, holiday time is busy, but I expect to get back to the recoding phase after the first of the year. As soon as I am ready to send data out to you, I'll write.

Best wishes for a prosperous New Year.

Sincerely,



David B. Lellinger
Associate Curator
Division of Ferns

Enclosure

CHAPTER II

THE THEORY AND PRACTICE OF THE VISUAL GROUND PLAN METHOD

Basic Premises

Inductive reasoning is the normal mode of taxonomic thought because deductive reasoning is too restrictive. Few premises of science, other than many of those of mathematics, can be put absolutely, which is usually necessary in framing deductive arguments. Statistical reasoning is a useful quantification of inductive reasoning. *Because statistics deals with probabilities rather than certainties,* it is an extremely useful tool for analyzing data. The scientific method of testing multiple hypotheses by matching observations against predictions of the hypotheses is a logical extension of the inductive method. The scientific method is argued against by some scientists, who claim that it causes an investigator to find what he seeks, that it biases his view (and his perception?) of data, and that it prejudices his results. But against this must be weighed the considerable advantage of stating hypotheses, by which focus and direction are given to research.

The principle of Occam's Razor, which chooses the concept that offers the greatest economy of assumptions, arose in the Middle Ages. Perhaps a simple set of premises or statements explaining a hypothesis is more likely to be true than a complex set, but the principle is fallible. Above all, the simplicity of assumptions should never be

considered a final measure of their truth; nor should simplicity necessarily be a major factor in choosing a set of assumptions.

Traditionally, taxonomists have made syntheses of data involving many hypotheses and assumptions. Some numerical taxonomists, however, have departed from this approach. Their greatest successes, in my opinion, have been in the taxonomy of the bacteria, a group in which it is difficult to state many taxonomic assumptions.

According to some people, evolution is itself the "grand assumption," and questionable. Evolution is occurring, this opinion notwithstanding, and what it has produced seems to be an accomplished fact. It should be unnecessary to defend the existence of evolution; the proofs of it are manifold. But, of course, models describing how evolution occurred, how it worked, and what it produced may be true or false in varying degrees.

There are a few points important to taxonomy which can be deduced from what we know about evolution. First, evolutionary distance can usually be inferred from morphological difference. Convergent or parallel evolution is the biggest exception to the general rule that things which look alike are probably closely related.

Second, it becomes increasingly difficult to find single, constantly valuable, definitive characteristics as one advances upward in the taxonomic hierarchy. Blockages of gene systems may be responsible for the "loss" of some characteristics. I believe that norms of characteristics among taxa may be more useful to evolutionary classification than the morphology of single exemplar taxa; classifications at the higher taxonomic levels in particular may have

to be based to some extent on potentialities, norms of characteristics, or ground plans, and not on actual characteristics of specific taxa. Ground plan characters and trends of specialization radiating from them are often useful because they illustrate what potentialities certain groups have, even when misleading absences of characteristics, which could be caused by blockages of gene systems, occur.

Lastly, and with special reference to the present study, five concepts regarding the diversity of nature have proved useful to me in making a model of evolution as a basis for classifying the genera and subgenera of the Adiantoidae.

- 1) Evolution is a more or less orderly changing of characteristics within sequences of organisms, in which both the characteristics and the organisms fall into patterns. Evolution is patterned rather than random because of the barriers to random assortment and expression of genetic material, which include spatial, genetic, and ecological isolating and differentiating mechanisms.
- 2) Most of these character gradients are divergent, which is to say, most evolution is divergent. Once isolating mechanisms come into play, differential selection pressures, however slight, cause once-similar stocks to diverge, as with Darwin's Galapagos finches. Of course, parallelisms, convergences, and even simple atavisms must sometimes occur. Indeed, it is these phenomena which constitute the principal barrier to successful intuitive interpretation of morphologically-based relationships.
- 3) Only natural groups, which are by definition monophyletic, have at their origin a common evolutionary ancestor. Natural groups

may be homogeneous in terms of their attributes, but homogeneity and naturalness are not necessarily coincidental. The former term does not imply the latter. The common ancestor may not exist today, but sometimes it can be inferred from the data of extant taxa.

4) Because of the divergent gradients of characteristics from original to derivative taxa, it is possible to infer the most probable characteristics of the common ancestor by interpolation and centripetal extrapolation, using the characteristics which, judging by their patterns in extant taxa, form the most highly patterned gradients.

5) Geological time is not an important dimension in describing evolutionary patterns, since rates of character change differ among characters and organisms. Because of this, organisms show different degrees of advancement. Some are highly specialized, having many advanced characteristics, while others are "living fossils" of ancient and primitive organisms.

The possibility of a natural system of classification of organisms is a notion which intrigued the ancients and remained a (mostly undiscoverable) certainty to Linnaeus. From the time of Aristotle, people had been sure that plants had an "essential" reality (see Cain, 1958, pp. 154-156). But it was Darwin's (1859) Origin of Species which allowed elaboration of the concept of "essence" within a sound framework of reality. Plants and animals were shown to be what they were and to be related the way they were because of their evolution (see Sokal and Sneath, 1963, pp. 12, 20). Hence it is only over the past hundred years that the search for a natural classification has been able to proceed on a basis of scientific concepts.

Although many classifications have been natural in part, the best natural system is the classification which would best mirror the evolution of organisms; it would be the one used if everything were known about all organisms and their evolution since they first existed on the earth. Classifications that attempt to be "natural" in this sense have this special purpose (cf. page 5), and the methods used to construct them are somewhat different from assortive, purely phenetic methods.

Nature is not a conscientious herbarium keeper. Although there is a diverse living flora, relatively little is known about the kinds of plants which preceded this flora. And the paleontological data are sometimes interpreted only with difficulty (see pages 26 and 27).

Any set of data which can be quantified and summarized in a model can be analyzed by operations which are based on the model or on its constituent hypotheses. Hypotheses about the nature of the data concern us first; until we have a concept of the data, we cannot make a set of methodological operations for dealing with the data. Among the hypotheses concerning variation in the data, one might assume that variation is random, hence there would be no evolutionary pattern. Here the resultant operational principle might be: "I quit!" On the other hand, the variation may be non-random. This is generally true of most biological data, which do show patterns, and any hypothesis or model ought to take this fact into account.

With horse phylogeny, Cuvier would perhaps have suggested that the model be linear advancement. In the evolution of the families and orders of dicotyledonous plants, the model might be always-

divergent evolution. In the Appalachian *Aspleniums*, Wagner (1954, 1963, pp. 145-146) has found a large series of interspecific hybrids all stemming from three basic diploids. The model would be one of always-reticulate evolution. To a taxonomist concerned with evolution within a genus of ferns or flowering plants in which some species were suspected hybrids, the model used to describe the data should allow for taxa produced either by divergent or by convergent evolution. The point is that a suitable model must be used to account for the structure of the data, and an adequate methodology must be used to analyze the data.

The system probably will evaluate characters in some way; it will make assumptions about the mechanisms and products of evolution, and will relate the taxa in a graphical way. Compared with strictly assortive methods, the evolution-based technique is bound to be complex and to suffer from shortcomings. Sometimes it will fail to work at all; at other times it will be wrong, and its users may be hard put to learn this, or to discover what in the data, model, or operations has caused the error.

Definition of Terms

Wagner (1961) has long advocated the use of a visual ground plan analysis in determining the evolutionary relationships among related groups (usually species or genera) of plants. His method, also briefly discussed by Benson (1962, pp. 273-277 and 415-417) attempts to uncover the threads of consistent (patterned) character evolution from among the multitude of available characters. The next

few paragraphs will define the technical vocabulary of the ground plan method.

A ground plan is the sum of primitive characteristics held by a group of taxa which have diverged along evolutionary pathways from a common (i.e., ground plan) ancestor. The ground plan method seeks to construct this totally primitive ancestor by inference, from a correlative examination of the characteristics of the organisms descended from it. The ground plan ancestor may exist as a living plant or a fossil, or it may at present be hypothetical and presumably extinct.

Of the two fundamental processes involved in the ground plan method, the first is determining characters useful to the analysis. Characters are defined here as features--any features--which can be described by two or more mutually exclusive character states and subsequently designated by numbers (character state codes). Character states are the set of attributes under any character. For example, under the character "leaf size" might be the states "<1 cm., 1--3 cm., and >3 cm." Most characters exist in two or more states, one of which is primitive in the group at hand, while the other(s) are divergent character states. For example, in the subgenus Coptophyllum of the fern genus Anemia the character state "horizontal" is primitive and the character state "oblique or erect" is the advanced state of the character "rhizome habit" (see Mickel, 1962, p. 456). A few characters are found in only a single character state in a group. "Plant body constructed of many cells" is such a null character in the group of higher plants. Fluctuating characters, which in their divergent

(advanced) character state(s) are scattered and not patterned throughout the group under study, cannot be included in a visual ground plan analysis because they do not indicate evolutionary patterns. Taxonomists often call fluctuating characters "trivial." "Lamina glandular" is a fluctuating character in the Adiantoidae, although it may be a patterned or a null character in certain genera of the subfamily. Non-fluctuating characters are called patterned characters. They are useful to the analysis, since they indicate evolutionary pathways, but only after their primitive versus divergent character state(s) have been determined. The distribution of character states in the operational taxonomic units (OTU's) and in adjacent groups determines whether a character can be considered patterned and whether one character state can be called primitive. Operational taxonomic units are the coded items, usually taxa, although they may be characters and attributes.

For the most part, primitive (unspecialized or ancestral) character states are present both in the majority of taxa of the group under study and also in adjoining groups of the next highest hierarchical level, such as adjacent genera in the study of the species of a single genus. Divergent (specialized or advanced) character states usually exist in few of the taxa under study and are not likely to exist frequently, if at all, in nearby groups (see Figure 5A). A rather different case occurs when a character state is found in most of the group under study, but is usually absent from the adjacent groups (Figure 5B). Within group a it is highly likely that this is a primitive character state of a patterned character, although the

distribution of this state in groups b-e is vastly different from the previous case. Both of these examples of patterned characters blend gradually into the more nearly random distribution of states that marks the non-usable fluctuating characters (Figure 5C).

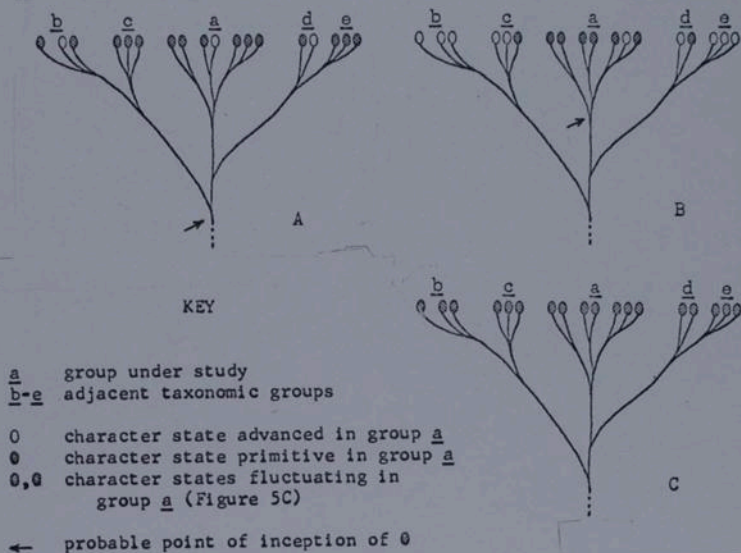


Fig. 5.--Models of character state distribution

Evolution has distributed character states into taxa, but whether we call a certain state primitive or advanced is determined by the limits of the group. In Figure 5B, the character state which is primitive in group a belongs to a patterned character. But in the group a-e neither character state has a clear majority, and the character is fluctuating.

A taxon x character table ($i \times m$) is the way data from the OTU's are stored. Each taxon is examined for every character, and the body of the table is filled with the character states pertaining to every taxon and character. The characters of the raw data, which may have two or more character states under them, are for some purposes converted to binary characters, and the data arrayed in a binary taxon x character table. A taxon x taxon matrix ($i \times i$) serves to associate or relate the similarity of every taxon with every other taxon. This is done by means of a measure of association, which mathematically assesses similarity (but not necessarily relationship) by means of evaluating the quantified characteristics of the two taxa at hand. Rogers' index of similarity (similarity ratio), s_{ij} , and Pearson's product-moment correlation coefficient, $r_{x,y}$, are examples. A character x character matrix ($m \times m$), which is the inverse of a taxon x taxon matrix, is constructed by assessing the similarity of pairs of taxa until every character has been compared with every other character.

Outline of the Method

The first step of the ground plan method is to collect the data into a taxon x character ($i \times m$) table. In the case of a small study not intended for computer analysis, this can be a mimeographed data recording sheet listing the characters as rows and the taxa as columns. In a larger study the data can be transferred to IBM cards, punching in the required format with a printing punch. This allows a visual check of the punches against the $i \times m$ table. Or IBM cards which are

to be punched out directly with a stylus as data are accumulated can be used (see page 58). The collection of the data into an $i \times m$ table, either on paper or cards, is analogous to mentally enumerating the various plants' characteristics.

It is best to be rather unselective of characters at this beginning stage. In fact, if the group is large, it may be useful to write short descriptions of the various OTU's, collate these, and select and codify the characters that are present. This decreases the likelihood of mistakenly overlooking some characters and, at the same time, including others which are correlated due to improper codification. In the body of the $i \times m$ table the character states for each character and taxon are shown as character state codes.

All of the data exist as characteristics of the various OTU's which have been entered into the $i \times m$ table. The task of the investigator or the computer is to evaluate the significance of each character as an indicator of evolutionary pattern. Only some, and perhaps just a few, of the characters are reliable or patterned; these characters belong to character complexes describing evolutionary patterns. For that reason they are usable in constructing a system of classification based upon those evolutionary patterns.

The non-null characters must be evaluated, and those which have a single primitive character state (versus one or more divergent states) selected. The model used to describe and explain the evolution of the primitive versus advanced states was discussed in the definitions on pages 39-40. Between the extremes of distribution patterns there is a range of patterns in which advancedness or primitiveness

cannot be decided upon surely. Ideally, the development of measures of probability for these intermediate patterns might permit the use of these characters. Sporne's use of the Chi-square distribution in this regard is not satisfactory, but it does point to the possibility of assigning probability values to characters as indicators of evolutionary trends. However, until satisfactory methods are developed, judging which characters to use depends (for reliability's sake) upon selecting only those characters which approach the extreme distribution patterns shown in Figures 5A and 5B. There is little doubt that much information is thereby lost, particularly in highly polythetic groups, where the defining set or "key characters" are not usually found all at once in an OTU, but a slightly varying subset is found in each of the various OTU's (see Sneath, 1962, p. 291). When doing a small ground plan analysis, within a genus, for example, the primitive states of the characters can often be discovered by inspection at the specific and generic level. In most small studies only one or two advanced states accompany the primitive state in each character. But if the project is large and machine computation needs to be used, then it is likely that an investigator will not have sufficient insight into the distribution of character states to permit a significant analysis of primitiveness or usefulness of characters as indicators of evolutionary patterns.

In this study three series of tests are used to distinguish patterned characters having a single primitive state. It is assumed that the most frequent states are the primitive ones, and it is the frequency distribution of these states which is tested.

A frequency test is performed to find what proportion of the total frequency lies in the most frequent state of each character. Pass lists are printed out for $f_1 \geq 0.5N$, $0.6N$, $0.7N$, $0.8N$, and $0.9N$ (see Figure 6 and Appendix E). Examination of the frequency values in Appendix D yielded data for $f_1 \geq 0.3N$ and $0.4N$. This test increases in selectivity rapidly as N 's coefficient rises, as shown in Table 1. It excludes character state distributions where the most frequent state passes the proportion test, but the states are otherwise quite frequent and perhaps evenly divided in a long series, as shown in Figure 6A. Null characters, which automatically pass the distribution

Table 1.--Numbers and percentages of characters passing the frequency distribution test at five values of N

<u>N</u>	<u>Tribe 1</u>	<u>Tribe 2</u>	<u>Tribe 3</u>	<u>Tribe 4</u>	<u>Tribe 5</u>	<u>Tribe 6</u>
0.50	82 76%	67 62%	71 66%	58 54%	67 62%	82 76%
0.60	67 62%	58 54%	54 50%	50 46%	48 44%	82 76%
0.70	56 52%	40 37%	38 35%	39 36%	44 41%	58 54%
0.80	45 42%	26 24%	29 27%	33 31%	38 35%	58 54%
0.90	29 27%	15 14%	14 13%	28 26%	27 25%	27 25%

tests, but are of no value to the construction of the ground plans, account for, respectively, 29, 11, 7, 14, 23, and 27 characters in the six tribes (see the Introduction to Chapter IV) within the sub-family. Percentages are calculated by dividing the number of characters passing the tests by 108, the total number of characters used in the study.

$$N = \sum_{i=1}^n f_i$$

$$f_2 \leq X f_1$$

$X = \text{a constant}$
(see text)

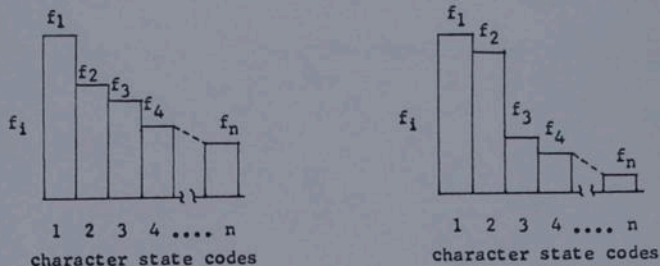


Fig. 6.--Character state frequency distribution histograms

The character state proportion test measures the relative frequency of the most frequent and second most frequent character states. Pass lists are printed out for $f_2 \leq 0.60f_1$, $0.65f_1$, $0.70f_1$, $0.75f_1$, and $0.80f_1$ (see Figure 6B and Appendix E). This test excludes characters which have (at least) two character states of high and fairly equal frequency, which is illustrated in Figure 6B. By definition there can be only a single primitive state, and if there seem to be more than one on the basis of character state distribution, then the character must be rejected. The number of characters passed in the six tribes is shown in Table 2. Increasing the proportion causes a less stringent selection, since the frequency of the second most frequent state can more closely approach that of the most frequent state. The number of null characters in the tribes is as was stated previously.

Table 2.--Numbers and percentages of characters passing the character state proportion test at five values of X

<u>Prop.</u>	<u>Tribe 1</u>	<u>Tribe 2</u>	<u>Tribe 3</u>	<u>Tribe 4</u>	<u>Tribe 5</u>	<u>Tribe 6</u>
0.60	81 75%	77 71%	69 64%	61 56%	75 69%	75 69%
0.65	84 78%	80 74%	76 70%	66 61%	79 73%	75 69%
0.70	93 86%	86 80%	81 75%	72 67%	83 77%	91 84%
0.75	99 92%	91 84%	88 81%	81 75%	89 82%	91 84%
0.80	99 92%	91 84%	88 81%	81 75%	89 82%	91 84%

The distribution of some character states is inconsistent across the diversity of taxa within the group under study; imagine a situation midway between the cases illustrated in Figures 5A and 5C. Here the adjacent group test helps to determine whether the most frequent character state is indeed primitive. If, as has been hypothesized, primitive character states are descended from a common ancestor, then the adjacent, coordinate groups of equal taxonomic rank are likely to show the same primitive character states. OTU's selected from the diversity of the adjacent taxa are used for this comparison. If highest frequency of a character state also occurs in the 75% or 80% majority of adjacent group OTU's, then that state is deemed primitive within the group under study (see Appendix E).

To allow for passing characters with the character state distribution hypothesized in Figure 5B, a frequency test result of $f_1 \geq 0.9N$ is taken as sufficient proof of common descent within the group, and the adjacent group test is not used. Greater risk in this bypassing of the adjacent group test would attend lower values of the

passing frequency and smaller groups, where statistical uncertainty is greater.

Once reliable information in the form of usable characters has been winnowed from the unsorted mass of data, these data are used to determine the most probable evolutionary patterns. These patterns are made by the association of complexes of characters which tend to be present as groups of divergent (advanced) attributes along the pathways of descent. The ground plan method seeks to determine the most probable pathways from among those that could be hypothesized within the limits of the data.

It seems reasonable to recognize economy of change as a guiding principle. It is less likely that the differences of each character between two taxa are the result of a long series of changes flip-flopping back and forth between advanced and primitive character states (Figure 7A1) than that the observed differences are the result of single changes (Figure 7A2). By rejecting the unpatterned characters, this conclusion is usually insured.

Similarly, the arrangement of taxa which requires the fewest number of changes--gains and losses--of characteristics to accomplish the patterns is sought (Figure 7B). The selection of patterned characters and careful ground plan analysis yields the result depicted in Figure 7B2, rather than that in Figure 7B1.

Of particular note are the so-called "losses" of characteristics. Sometimes after the data are fitted into a generally economical pattern, certain peripheral taxa lack advanced characteristics which were present in their progenitors. If a presumably derivative taxon has many of

these "losses," one suspects that the pattern that was worked out is not the most probable one (Figure 7C1) and that rearrangements are needed (Figure 7C2). If rearrangements of even the entire pattern make little improvement, or there is contravening biological evidence, such a lack of an advanced characteristic may be an extreme specialization, like the loss of chlorophyll in certain Santalaceae. Figure 7C1 illustrates this in character a.

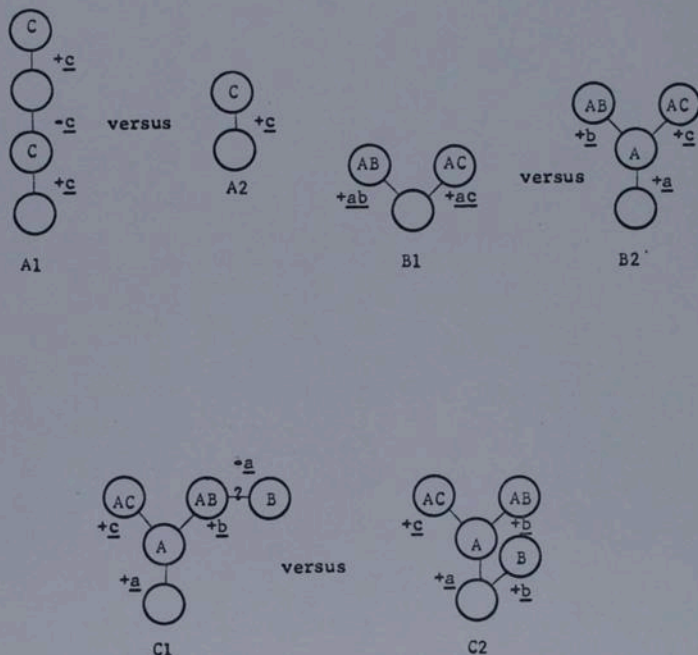


Fig. 7.--Construction of probable evolutionary pathways

Once a set of constant characters, each having a single character state that is primitive with high probability, has been found, these characters are used to generate a taxon $\times$ taxon ($i \times j$) matrix from the relevant rows of the $i \times m$ table. Such a process is in part the analogue of an intuitive assessment of relationship.

Pearson's product-moment correlation coefficient serves this purpose in studies of many multi-state characters in many taxa. Rogers' similarity ratio is perhaps even better suited, and it or Sneath's index of similarity can also be used with binary characters. The full data set of chosen characters should be used to calculate the $i \times j$ matrix, whatever measure of association is employed.

The next step is to prepare a visual ground plan diagram, which is based on a series of concentric rings. The radial distances indicate divergence from the ground plan ancestor, which lies at the center of the rings. The problem is to place the taxa on the diagram so that their connections, which are functions of similarity of character changes, trace the most probable evolutionary pathways. The finished diagram should illustrate the probable course of evolution of the taxa, and may resemble a family tree such as a classical taxonomist might develop from his un verbalized knowledge of taxon relationships.

In making a small scale visual ground plan diagram, the taxon with the lowest divergence index is first placed on the diagram. Then the remaining unplaced taxa are searched to see which of these is most like the taxon already placed. This process is repeated, each time trying to add and to rearrange the taxa so that there are the fewest

gains and losses of characters. Occasionally a hypothetical taxon will be added in an evolutionary line between two real taxa where a third taxon or evolutionary line branches off. After the ground plan diagram is complete, the fluctuating characters are assessed. They are either eliminated from a re-analysis and revised diagram or the taxa are rearranged to reduce or eliminate the fluctuations on the original diagram.

When the computer is used to process data for a ground plan diagram, the primitive character state, coded (0), is used against all the divergent states--no matter how many--which are lumped together and coded (1). This has the advantage that programming the computer is far simpler than with many states. But more important, the evolutionary relationships of the various advanced states is usually uncertain in large studies. Lumping has the disadvantage of losing some information and, especially, equating different advanced states on the diagrams. While the loss of information by no means precludes satisfactory results, the lumping becomes increasingly questionable at the higher taxonomic levels, and may reduce the effectiveness of evolutionary ground plan methods at those levels.

In most of the computer studies, the character selection tests have been stringent enough so that fluctuating characters which did pass the tests have not overpowered the patterned characters. The computer has produced a largely satisfactory result. But in almost all of these studies, one or two of the taxa seem to be out of place. Here, fluctuating characters isolate a highly advanced OTU from its nearest attachment to the remainder of the ground plan diagram by a

long series of character losses. The taxonomic importance of the fluctuations has been difficult to evaluate, since the studies have been at the generic and tribal level. In the computer technique, which will be described in greater detail in Chapter III, the computer makes the $i \times m$ and $i \times i$ tables. Divergence index relates taxa on the ground plan diagram at the beginning of an evolutionary line from a ground plan ancestor. Pearson's $r_{x,y}$ is used to associate taxa in other positions on the diagram. Hypothetical intermediate taxa are calculated for every character change (at what would be each ring of the visual ground plan diagram, so to speak) on the basis of the characters' frequencies among the real taxa. The results are printed out by the computer as sequences leading from the point of attachment of the evolutionary line through the intermediates to the taxon to be added. The lists of taxa and their characters appear in such a way that a standard visual ground plan diagram can be readily produced from the computer output.

An $m \times m$ matrix was used to estimate the correlation between all the pairs of the 108 characters of the same seventy-four OTU's of the Adiantoidae. Four characters showed perfect correlation. A matrix made with the characters passed by the character distribution selection mentioned in the paragraph above was compared against a matrix made with those characters less the four fully redundant ones. The similarity of the two matrices was 0.999. Almost no information was lost by dropping the four redundant characters.

Systematization Through the Evolutionary Ground Plan Method

At the heart of any taximetric system is the method which is used to convert the similarity matrix data into a dendrogram, evolutionary ground plan diagram, or other visual representation of the data. It is this procedure which in the evolutionary ground plan method, but not in other current taximetric methods, is closely associated with a model of evolution.

Because the evolutionary ground plan diagram method utilizes a model of divergent evolution, the diagrams it produces consist of dichotomizing (or rarely trichotomizing) pathways extending along divergent gradients of advanced characteristics from the ground plan ancestor to OTU's which stand for taxa included in the study. Some real OTU's lie along the pathways or at branching points, as well as at the termini of all branches of the diagram.

The data, consisting (for real taxa) of character states which are most likely to be primitive and to show evolutionary patterns, are converted into an $i \times i$ matrix and are stored in an $i \times m$ table (M),

one taxon to a column. From this the divergence sums (CSUM) and the divergence indices (PSUM) are totalled as the sum of primitive character states across each row and P (the number of characters) minus the sum of the primitive character states down each column, respectively.

As with other taximetric methods, OTU's are associated with one another in terms of the similarity of characteristics which they share. In constructing ground plan diagrams of the tribes, the similarity of multi-state characteristics is assessed with Pearson's product-moment correlation coefficient, $r_{x,y}$. Rogers' similarity ratio, s_{ij} , is used with the binary characters of the subfamily ground plan diagram. Either measure of association can range from +1. (100% positive correlation) through 0 (no correlation) to -1. (100% negative correlation), but for the most part, correlations between taxa range upward from -0.2.

To save computer time, the standard deviations (STD) used in the character correlation selection program are included as data. An internal function calculates the indices of similarity from STD, N, P, and PSUM. The number of correlation coefficients which must be calculated for an $i \times i$ matrix of both real and hypothetical intermediate OTU's seems to be controlled by: (1) the number of real OTU's, (2) the number of characters, (3) the distances, in terms of the number of character differences, between the OTU's, and (4) the degree of pattern shown by the characteristics. The details of the evolutionary ground plan diagram program follow in the paragraphs below, and are summarized in Figure 16.

The ground plan ancestor, which is defined as that real or hypothetical OTU with all primitive characteristics, is generated by filling P spaces with character state (1). Code number 501 is assigned, and the data for this OTU are stored in TAPE. Code numbers 502 through 999 remain available for the hypothetical OTU's yet to be generated by the computer; in the figures these are labelled upwards from 1, rather than 501.

The real OTU most similar to the ground plan ancestor is the one with the fewest advanced character states, which in this program are coded (0). This OTU has the lowest divergence index (highest PSUM), and is thus chosen by the computer from the list PSUM.

A comparison is made between the two OTU's, the one "previously placed" and the other to be "added." Characters which are advanced in the added OTU but are primitive in the previously placed OTU are put in temporary list ADV. Those which are advanced in the previously placed OTU but are primitive in the added one go in temporary list LOSS.

Whether there are one or several differences in advanced characteristics between the previously placed and the added OTU's, the code numbers of the previously placed, the hypothetical intermediate(s) (if any), and the added OTU's are printed out in order, along with their sets of characters, whether in the primitive or advanced state (coded (1) and (0), respectively, in the output). In addition, for all but the previously placed OTU, the one character which has changed in each OTU is also printed out. Thus the evolutionary ground plan diagram is built up OTU-by-OTU and character

change-by-character change, like beads on strings (see Appendix G and Figures 17 through 25).

When there is more than one character difference between the previously placed OTU and the OTU to be added, the CSUM values are used to sort these differences in decreasing order of their frequency in the real OTU's of the group being diagrammed. They are attached outward in this decreasing frequency of advanced state order, both to the hypothetical intermediate OTU's which will be intercalated between the previously placed OTU and the OTU to be added. The census of character states of a hypothetical OTU is the sum of advanced and primitive character states of other OTU's leading up to it from the ground plan ancestor plus the one change that it has in a character state that its proximal, neighboring OTU does not have. For example, in going from a previously placed OTU with advanced character b to an OTU to be added with advanced characters bcde, if d had the highest divergence sum (highest CSUM) and e the lowest of characters c, d, and e, then d would be the first to be added and e the last. The sequence of advanced character state gains would be b-bd-bcd-bcde. The first and last character state groups belong to real OTU's. The middle two have been constituted for hypothetical OTU's, and code numbers are assigned from the top of a list called HTNUM.

Just storing the full $i \times i$ matrix for the N real OTU's takes $(N^2-N)/2$ spaces. To generate and store in addition the indices of similarity of all the hypothetical OTU's times the real OTU's is beyond the capabilities of most computers. Therefore, only the one highest correlation coefficient for each unplaced OTU and its real or

hypothetical mate is stored in HCOL. The code numbers of the OTU pairs are stored in HRN.

Assuming that there are more real OTU's to place in the ground plan diagram list (list UNPL is scanned to determine this), the OTU to be added is chosen on the basis that it is the highest one in HCOL. Parallel list HRN indicates the code numbers of the previously placed OTU and the OTU to be added.

If the correlation between the previously placed OTU and the one to be added is not statistically significant (the value, called RHO is assigned as a constant on a control card) for a test of $\rho = 0$ at the level of $p \leq 0.05$, then the OTU in list UNPL with the highest PSUM is chosen to be put on the output list next, starting a new line from the ground plan ancestor. The added OTU's code number is removed from list UNPL.

But if RHO is exceeded, and the correlation is statistically significant, then the code number of the OTU to be added is removed from UNPL, the data of the added and previously placed OTU's are compared, with temporary lists ADV and LOSS used as before, and the added OTU is placed on an evolutionary line stemming from some real or hypothetical OTU more advanced than the ground plan ancestor.

When all the data of real OTU's have been processed, the list UNPL is empty, and the program terminates.

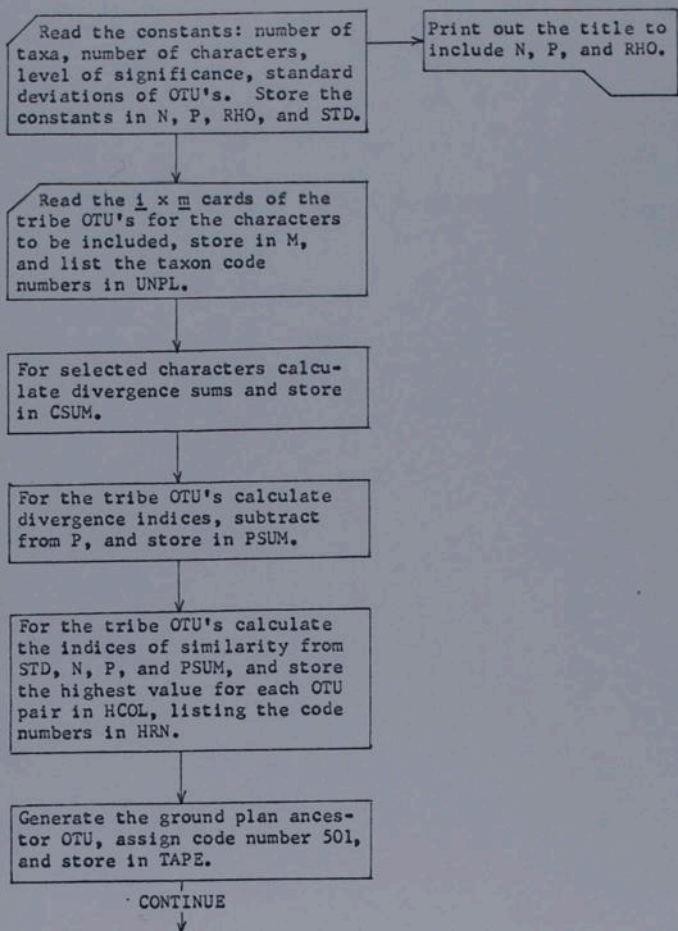


Fig. 16.--Flowchart of the state code ordering program,
begun

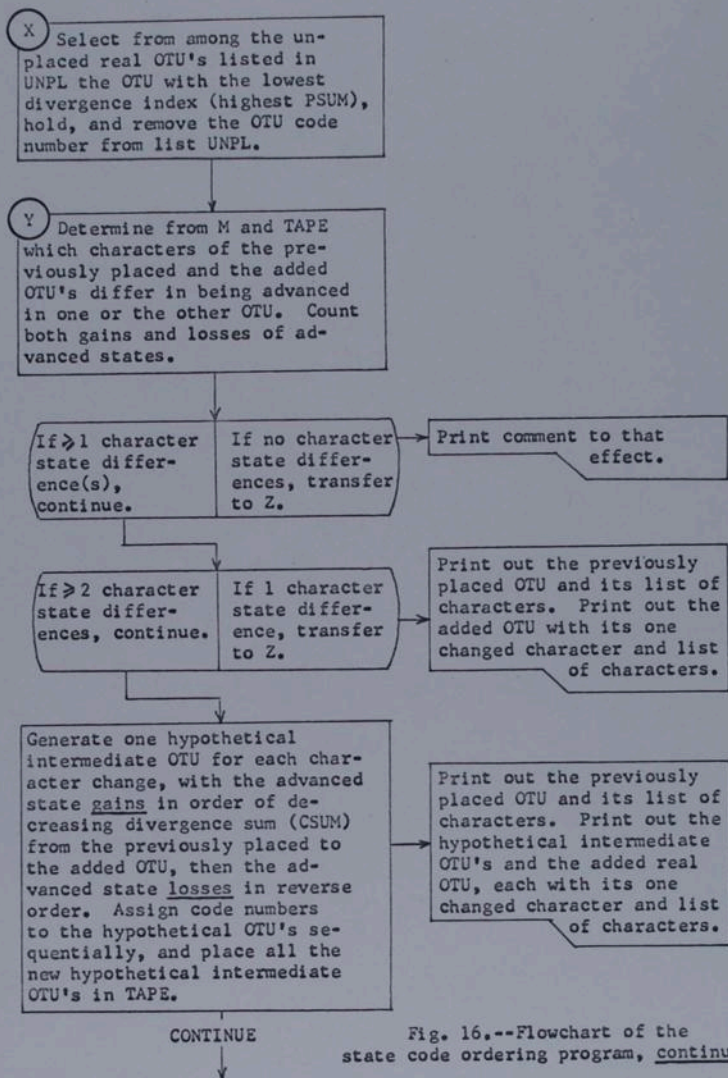


Fig. 16.--Flowchart of the state code ordering program, continued

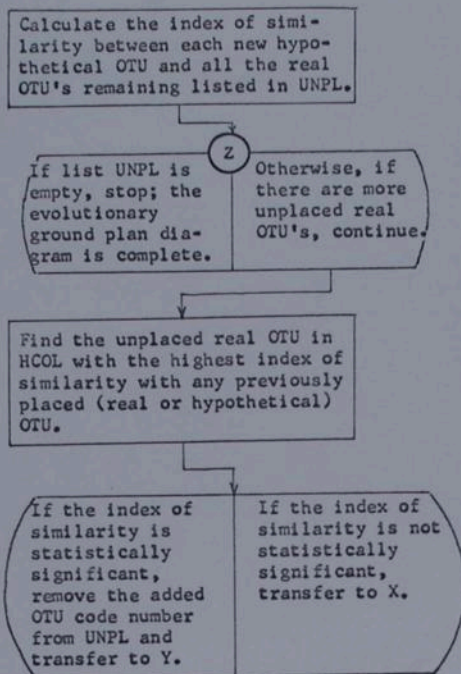


Fig. 16.--Flowchart of the state code ordering program,
concluded

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF ARTS AND SCIENCE
COLLEGE OF SCIENCE



DEPARTMENT OF BOTANY

LONDON, CANADA

September 28, 1966.

Dr. David J. Rogers,
Department of Botany and Plant Pathology,
Colorado State University,
FORT COLLINS,
Colorado,
U.S.A.

Dear Dr. Rogers:

Many thanks for sending me the reprint from Systematic Zoology vol. 15 (Graph-Theory model), also the general write-up for the similarity clustering program. They will be of much use to me.

My own work concerns a systematic study of the tribe Arundinelleae (Gramineae) though I have published nothing numerical as yet. Dr. Orloci, of this department, whose papers you may have seen in J. Ecol., is another worker in this field.

Yours sincerely,

A handwritten signature in cursive script that reads "J. B. Phipps".

J. B. Phipps,
Assistant Professor of Botany.

JBP:kdb

- Taxonomy Laboratory

August 30, 1966

Mr. J. B. Phipps
Department of Botany
College of Science
Faculty of Arts and Science
The University of Western Ontario
London, Canada

Dear Mr. Phipps:

I am sending you a flow chart for a program for our graph theory clustering program. The flow chart will be of more general value to you than would be a listing of the FORTRAN statements. This would be the case under any circumstances since our experience has been that no two computer installations have the same instruction set. I am also sending a mimeographed paper "A General Method of Taxonomic Description, etc."

The last mentioned paper is sent because we feel that a thorough understanding of the preparation of taxonomic data is a critical element to be understood before attempting to use the clustering method. I might say this applies no matter what computer program is employed. We have written the computer program in FORTRAN IV, but you will see places in which machine language is employed. Machine language is required in order to get sufficient room in the internal memory units of any computer that we know. This will require that program be done by a relatively experienced and competent programmer.

These items, though in mimeograph form, have been submitted for publication and are, therefore, not for general distribution.

I will be interested to know of your proposed use of this program and would like to hear from you when you have had an opportunity to digest the contents of the papers sent. Please do not hesitate to write for assistance.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

THE UNIVERSITY OF WESTERN ONTARIO
FACULTY OF ARTS AND SCIENCE
COLLEGE OF SCIENCE



DEPARTMENT OF BOTANY

LONDON, CANADA

August 26, 1966

Dr. David J. Rogers,
Taxonomy Laboratory,
Dept. of Botany and Plant Pathology,
Colorado State University,
FORT COLLINS,
Colorado 80521,
U. S. A.

Dear Dr. Rogers:

I would be very grateful for a copy
of your program for clustering by graph theory, as announced
in this March's issue of Systematic Zoology, if available.

Yours sincerely,

Jane B. Phipps

J. B. Phipps
Assistant Professor of Botany.

JBP/L

THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210

~~XXXXXXXXXXXXXXXXXXXXXXXXXXXX~~
~~XXXXXXXXXXXXXXXXXXXXXXXXXXXX~~
College of Biological Sciences

BOTANY AND PLANT PATHOLOGY
BOTANY AND ZOOLOGY BUILDING
1735 NEIL AVENUE

July 25, 1966

JUL 29 1966

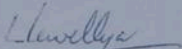
Dr. David J. Rogers
Taxonomy Laboratory
Department of Botany and Plant Pathology
Colorado State University
Fort Collins, Colorado 80521

Dear Dave:

Thank you for your letter which I'll consider carefully when the reprint mentioned arrives. I can't find it, however, so would you please send me another copy.

I hope your Mexican trip was successful. I leave August 11 for Denmark, Sweden . . ., with Catherine Martha. Paul is coring Galapagos craters at present.

Sincerely,



Llewellyn H. Colinvaux

LHC:mv

THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210

COLLEGE OF AGRICULTURE AND HOME ECONOMICS
Resident Instruction—Research—Extension

BOTANY AND PLANT PATHOLOGY
BOTANY AND ZOOLOGY BUILDING
1735 NEIL AVENUE

December 29, 1964

Dr. D. J. Rogers
New York Botanical Garden
Bronx Park
New York 58, New York

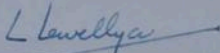
Dear Dave:

I had hoped to get a letter to you before you left for Montreal, saying I would not be there. I would indeed have liked to hear the Rogers' paper - but had such a delight at home. Our little girl arrived December 18 - the day of the eclipse! - and she has charmed both of us already. It has been truly a wonderful Christmas!

I will write you sometime in January when my schedule is worked out and I know how my time is used. The Ohio State University is on the quarter system, and I don't yet know what my assignment is for this term.

Our very best wishes for the New Year.

Yours sincerely,



Llewellyn H. Colinviaux
Assistant Professor

LHC:ja

November 24, 1964

Dr. Llewellyn H. Colinvaux
Department of Botany
1735 Neil Avenue
The Ohio State University
Columbus, Ohio 43210

Dear Llewellyn:

Thanks for your letter of the 17th. I will jump right in to answer the question on the joint paper, the subject of your second paragraph. The idea of the paper was that there should be an analysis made of the attributes which define Halimeda and an analysis of the species thereof. I had thought that you were at the point of changing some of the characters and attributes which we used in the analysis and that some modifications would be required in the input data concerning several species which you have additional specimens of. I also thought that you had a couple of new species that you would like to incorporate with the others. In other words, what would be needed now would be a recoding of all this data if these do not conform to that used in the monograph and a rerun of these in the computer to see what results would come from the new data. This would then give us a sufficient amount of information for a paper.

The "baw" that I had intended was the recoding of the data and the addition of the data for any new specimens and the old species. The computer program to which I refer is indeed a new program different from the one given in the BioScience paper.

If this fits your schedule, which looks to me like a very tight one, then this would give us something to talk about if you have the opportunity to come to Montreal at Christmas time. I trust that all of your activities, including the forthcoming blessed event, will be successful.

Sincerely

DJR:MDP

THE OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210

COLLEGE OF AGRICULTURE AND HOME ECONOMICS
Resident Instruction—Research—Extension

BOTANY AND PLANT PATHOLOGY
BOTANY AND ZOOLOGY BUILDING
1735 NEIL AVENUE

November 17, 1964

Dr. David J. Rogers
The New York Botanical Garden
Bronx Park
New York 58, New York

Dear Dave:

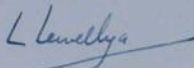
I was delighted as usual to receive your letter, and to hear that more work is planned on *Halimeda*. And I would indeed like to continue co-operating in the project. I have, however, an exceedingly full schedule. Our first offspring is expected sometime before Christmas - and I begin teaching here January 4, for two quarter terms. If either N.S.F. or O.N.R. come through with support of my planned program on the culturing of Siphonales and the working up of Indian Ocean *Halimeda*s I am committed to spending half time on these projects-for they will be paying half my salary. This being so, without an assistant to help with some of my research, I don't feel I can fairly undertake data punching. Perhaps something along the line of an assistant could eventually be arranged, but my grants will not take care of that immediately.

What have you in mind for the joint paper? In theory I should like to do one with you, and it is the natural follow-up to your work. Would this paper be based on the already existing data, or will it be the outcome, possibly in one or more years time, of our future work? I am somewhat confused by your adding the word "new" to "computer program" in the third line of your letter.

A discussion with you on possible content of a joint paper and the scope of our future work would, I feel, be useful. I note you are giving a paper at the AAAS meetings the end of December. This is the most likely time I could travel about - if our little one comes conveniently early in December. How is your schedule at Montreal and do you think something might be arranged for then? I would then be able to hear your paper which I have regretted having to miss at Edinburgh and Quebec.

With best wishes,

Sincerely yours,



Llewellyn H. Colinvaux

LHC:ja

October 23, 1964

Dr. Llewellyn H. Colinvaux
Dept. of Botany and Plant Pathology
Ohio State University
Columbus, Ohio 43210

Dear Llewellyn:

I hope by now that you have seen our BioScience article; if not, you will be receiving a reprint of it as soon as they are available. I am prompted to write to you because we have more interesting things with *Halimeda* and I wonder if you wish to continue in this direction.

First of all, it would seem profitable to me to punch in whatever new information or new species you have and to re-work whatever old information needs it. We will run the data without cost, but we certainly hope that you would punch out for us all the new punching necessary. We believe that a new development here will even be more useful in classification procedures.

Although we have not yet gotten the program written for the computer, it would further seem to me to be useful for us to do a joint paper with you as a senior author in which the findings from the computer program will be discussed in terms of their taxonomic merit.

Let me know if you are interested in pursuing this further, and we will get together when and if you so desire. Looking forward to hearing from you,

Sincerely

DJR:MDF

- Taxonomy Laboratory

September 7, 1966

Dr. Perry R. Stout
Director
Kearney Foundation of Soil Science
University of California
265 Hoagland Hall
Davis, California 95616

Dear Dr. Stout:

I am remiss in not acknowledging your letter of August 22.

It seems that in the search for applications for computers in medicine in biology that the more "classical" areas of biology get overlooked. I note, for instance, in the citations of the book "Digital Computer Needs in Universities and Colleges" that the biological applications of those more specifically oriented to medicine.

I think that I will soon be in a position to make something of a bibliography of the more strictly biological applications inasmuch as I have just finished writing a paper reviewing these applications. Perhaps I will submit these notes and bibliography to the National Academy to have added to their information store.

Once again, thanks for your several letters and references.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

KEARNEY FOUNDATION OF SOIL SCIENCE

265 HOAGLAND HALL
DAVIS, CALIFORNIA 95616

August 22, 1966

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

Dear Dr. Rogers:

Relative to the question raised in your August 11 letter, it occurs to me that S.L. Empey might have suggestions on how to discover the various computer applications in botany.

The cue comes from Note 33, page 86 of "Digital Computer Needs in Universities and Colleges", where reference is made to Empey's "Computer Applications in Medicine and the Biological Sciences; Bibliography", Communications of the ACM, 1963.

I doubt that anyone in the Academy would have direct leads although Mr. Robert E. Green, Executive Secretary of the Committee on Science and Public Policy of the National Academy may be able to suggest a procedure.

In making the present series of studies on Science in American Universities, the Academy appointed study committees who in turn dug out the necessary information in the best way they could.

Sincerely,

A handwritten signature in dark ink, reading "Perry R. Stout".

Perry R. Stout
Director

PRS:vla

- Taxonomy Laboratory

August 11, 1966

Dr. Perry R. Stout, Director
Kearney Foundation of Soil Science
University of California
265 Hoagland Hall
Davis, California 95616

Dear Dr. Stout:

I am pleased to have your two letters of July 29 and August 1 and your copy of "Plant Sciences." I am returning your copy under separate cover and have ordered one for myself. We already have a copy of the report "Digital Computer Needs in Universities and Colleges." These are all instructive pieces of information. As you suggest, I think it will be useful to discover over the next few years the various applications of computers in botany. I am not certain how this should proceed, but we will give it some thought. I should think that the National Academy might be interested. Can you suggest the most appropriate person in the Academy to whom I might address inquiries for aid or suggestions?

The report on the plant sciences is certainly a comprehensive document. I am particularly impressed with the section on applied and economic botany, and I trust the recommendations included therein can be put into action.

Again, thank you for your considerations and letters.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:ch

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

KEARNEY FOUNDATION OF SOIL SCIENCE

265 HOAGLAND HALL
DAVIS, CALIFORNIA 95616

August 1, 1966

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

Dear Professor Rogers:

This morning's mail brought me a copy of the new report on "Digital Computer Needs in Universities and Colleges" which is now available from the

Printing and Publishing Office
National Academy of Sciences
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

*Call Nat. Acad.
Library
LIB 1027.5
N3*

No price is named. The Library of Congress Card number is 64-60097. I am sending verifax copies of the Table of Contents.

In building cases for the computer needs of life sciences, interested parties should have the aforementioned document for ready reference. I am sending you verifax copies of pages 90 and 91 which show that treatment of computer needs for life sciences (medicine excepted) is relatively cursory, though increasing needs are freely predicted.

The amount of attention paid to computer uses among science and engineering sectors appears to reflect the proportionality of relevant published papers.

As I have indicated in an earlier letter, the struggle for public support for computer facilities is spread through all segments of science with the result that policy makers now look upon the supplying of computers as an item unto itself.

I trust that this information will aid in bringing the needs of taxonomy for computers into a routing most likely to produce the desired support.

Sincerely,

Perry R. Stout
Director

PRS:vla

CONTENTS

I. INTRODUCTION AND RECOMMENDATIONS	1
Introduction	1
Recommendations	1
II. DIGEST OF FINDINGS	3
Introduction	3
Need for Computers	3
General Usefulness	3
Adjunct to Other Scientific Equipment	3
Time-Saver for Human Workers	3
Key to New Developments	3
Estimates of Future Needs	3
Need for Training	4
Need for Increased Federal Support	6
Cooperation Between Government Agencies	6
Needs for the Future	8
III. COMPUTERS AND THEIR UTILIZATION	9
Introduction	9
What is a Computer?	9
Importance of Software	10
Rapidity of Development	11
General Use of Computers	12
Problems in Computer Science	16
Computing Needs in Scientific Research	16
Physics	16
Engineering	18
Other Physical-Science Areas	20
Biological Areas	20
Other Areas	22
Computing Needs for Teaching	23
Using Computers to Teach	23
Teaching about Computers	23
Future Expectations	23
Educating Experts	25
Computing Needs in Computer Science	25
An Estimate of the Support Needed for Computers in	
Universities and Colleges	26
Past Support	26
Future Support	26
Availability of Computers	27
Availability of Staff	27
Availability of Funds	28
Financial Arrangements	30
The Need for Coordination	30
The Support of Software	30

The Educational Discount	30
The Special Role of the National Science Foundation	31
IV. CURRENT COMMERCIAL & INDUSTRIAL APPLICATIONS.	32
Introduction	32
Applications	32
Clerical Applications	33
Process-Control Applications	34
Management Applications	34
Design Applications	36
The Aircraft and Airline Industries	37
Future Developments	38
V. FUNDING PROCEDURES & COSTS	39
Funding	39
A Glance at Current Practices in Universities and Colleges	39
Conflicting Agency Requirements	42
Classification of Kinds of Computer Use	44
Enunciation and Discussion of Possible Funding Procedures	46
Cost of Equipment, Staff, and Space	50
Summary	50
Computer Support for Scientific Research	51
Computer-Science Research	51
Education	53
Totals	53
Cost of Equipment	53
Historic Costs	53
Estimated Computer Needs in 1968	55
Cost of Staff	55
Cost of Space	60
Space for Equipment	60
Space for People	61
VI. USE IN RESEARCH	64
Computer Requirements for Research in the Physical Sciences	64
General	64
Physics	64
Nuclear Reactions	65
Nuclear Structure	65
Elementary Particles—Experimental	65
Cosmic Rays	66
Solid-State Physics	66
Particle-Scattering Calculations	66
Plasma Physics	66
Other Fields	67
Estimate by Research Institutions	67
Chemistry	68
Crystallography	68
Quantum Chemistry	68
Statistical Mechanics	68
Activation Analysis	69
Molecular Spectroscopy	69
Astronomy	69
Planet and Satellite Orbits	70
Astrophysics	70

Space Science	71
The Physics of Interplanetary Space	71
Orbit Computations	71
Experiments in Space	71
The Earth Sciences	71
Meteorology	72
Oceanography	72
Geophysics	72
Geology and Geochemistry	72
Estimates of Needs in 1966 and 1968	72
Abstracts of Examples of Computer Use in the	73
Physical Sciences	73
Bubble-Chamber and Spark Chamber Data Processing.	73
Analysis of Short-Term Fluctuations in the	76
Cosmic-Ray Intensity	76
Nucleon-Nucleon Scattering	76
Nuclear Absorption and Decay of Muons Bound to	76
Deformed Nuclei	76
Quasi-Elastic Proton-Proton Scattering	76
Plasma Diffusion	77
Quantum Chemistry	77
Combined X-Ray and Neutron-Diffraction Study of the	77
Ordering of Atoms in Transition-Group Alloys	78
Calculation of Approximate Stellar Models	78
Literal Development of Satellite Theory and Estimation	78
of the Accumulated Error by Numerical Integration	78
Long-Range Solar Weather Relationships	78
Analysis of Microseisms	78
Pattern Recognition and Interpretation in Geophysics.	78
Calculation of Invariant Co-ordinates for Data Points	78
from Explorer XI	78
Automated Observatories	79
Computers in Engineering Research	79
Introduction	79
General Characteristics of Computer Utilization in	79
Engineering	81
Improvements in Classical Computations	82
On-Line Computer Control	82
Simulation	83
Computer Utilization in the Aerospace Industry	84
Computer Optimizations and Automated Design	85
Computers in Engineering Education	86
Life Sciences and Medicine	86
Data Analysis	87
Analysis of Experiments	88
Information Retrieval	89
Analysis of Models	90
Diagnosis	91
Outlook	91
Behavioral Science	92
Economics	93
Business Administration	94
Psychology	95
Artificial Intelligence	95
Linguistics	95

Information Retrieval	95
Other Social Sciences	96
Summary	96
Illustrative Examples	97
Economics: Simulation of the Economy.	97
Economics: Technological Changes	97
Business Administration: Linear Programming.	97
Business Administration: Simulation	97
Psychology: Factor Analysis	98
Psychology: Simulation of Thought Processes.	98
Artificial Intelligence.	98
Linguistics.	98
Information Retrieval: Legal Searches	98
Content Analysis	99
Professional Applications of Computers	99
Introduction	99
Computers and the Law	100
The Uses of Modern Logic in Law	100
The Uses of Information Retrieval in Law	101
The Analysis of Legal Decision Making	101
The Relationships Between Science and Law	102
Summary	102
VII. THE COMPUTER SCIENCES	103
Introduction	103
The Development of Hardware	104
Increased Speed	105
Microwave Techniques	105
Tunnel Diodes	106
Superconductivity	106
Thin-Film Magnetics	106
Others	106
Increase in Memory Size	107
Thermoplastic Recording	107
Photoscopic Disks	107
Disk Files	107
Others	108
Organization	108
Time Sharing	108
Satellite Mode	108
Variable Structure	109
Highly Parallel Processing	109
Others	109
Subsystems	110
The Development of Software	110
Computer Languages	110
Executive Systems	112
Debugging	112
Future Prospects	113
Development of Procedures	113
Numerical Analysis	114
Information Storage and Retrieval	115
Artificial Intelligence	116
Teaching Machines	117
Outlook	118
General Considerations	118

Present Financial Support	118
Future Financial Support	119
Computation Laboratories	119
Purposes and Kinds	119
Some Historical Remarks	121
 VII. USE IN EDUCATION	 123
Summary	123
Discussion of Needs	123
Social Desirability	123
Training of Technicians	124
Undergraduate Education of Computer Users	126
Engineering, Mathematics, and the Physical Sciences	126
Business Administration and Commerce	128
Behavioral Sciences	129
Biological and Premedical Studies	130
Summary of Teaching of Computation as a Tool for Scientists	130
Education of Computer Scientists and Specialists	130
Undergraduate Education	131
Graduate Education	131
Curriculum Questions	132
Financial Support	133
Present Investment	133
Estimated Future Needs	134
 APPENDIX A	 137
Classification of Computers	137
 APPENDIX B	 140
Computers in Colleges and Universities—1963	140
 APPENDIX C	 151
Investment in Computers as of January 1963	151
 APPENDIX D	 153
The Programming Process	153
 APPENDIX E	 159
Updated Estimates	159
 APPENDIX F	 163
Analogue Computers	163
 GLOSSARY	 165
 INDEX	 173

Basic research in genetics uses computer-based techniques ranging from numerical integration occurring in a mathematical model for random fluctuation of mutation load⁵⁵ to the simulation of the effect of selection against phenotypic extremes⁵⁶ to segregation of genetic categories in the study of muscular dystrophy⁵⁷.

A model of genetic linkage in man has been made with the aid of a computer⁵⁸. Potentially, such studies can lead to the recognition of those individuals within families who are susceptible to particular diseases. Preventive care would be much facilitated by such information.

No satisfactory mathematical theory now exists which relates the behavior of assemblies of formal neurons to the experimental data of neurophysiology. Recent studies of networks of neurooids (whose properties resemble those observed in neurons) have exhibited a great number of striking similarities between certain measures of summated excitation in a model "sheet" of only 1296 elements and several aspects of the so-called "slow potentials" recorded by electrophysiologists⁵⁹.

These similarities are still mostly qualitative and depend to a large extent upon the values of parameters that describe the properties of the individual elements (such as time-constants), the topology of the nets (for instance, two-dimensional probability distributions with radial symmetry), and the degree of coupling within the net. Investigations of this model have required many hours of computer time on the largest available computers. Many questions remain to be asked and expansion of the model would be valuable, but these require computers of greater capacity than are now available.

Modern embryology has turned to the view that differentiation and the resulting organismic architecture are controlled by a system in which some particular cells at particular stages of development interact with their environment—perhaps a differing cell population—changing the course of their continuing development. In this view, an algorithmic description of the organism would be contained in the gene. Very simple structural algorithms have been devised to simulate organisms with minimal structure complexity. Many questions that need to be answered concerning the efficacy of this approach will require extensive use of very large scale computers.

In the life sciences, the study of models of a complexity requiring the use of computers appears destined for a substantial increase. The computers now available for such research are often capable of dealing only with models

that are grossly oversimplified. Anticipated benefits from the study of such models will depend on the availability of large computers and the development of new computers with even greater capacities.

Diagnosis

The possibility of using the highly efficient techniques of modern data processing to assist in diagnosis has been recognized for some time. Nevertheless, the results of the initial experimentation along this line are only now becoming available. In Sweden⁶⁰, clinical data are being analyzed with the aid of a computer to provide tentative diagnoses. The doctor makes a final diagnosis on the basis of the probabilistic machine diagnosis, his personal impression and, if necessary, additional medical examinations. The results of this procedure have been reported to be most satisfactory.

In this country, a number of studies are attempting to illustrate and to develop the use of computers in assisting the differential diag-

Notes

55. M. Kimura, T. Maruyama and James F. Crow, "The Mutation Load in Small Populations," Genetics, 1963, 48: 1303-12.
56. A.S. Fraser, "Simulation of Genetic Systems by Automatic Digital Computers," Australian Journal of Biological Sciences, 1960, 13: 344-350.
57. N.E. Morton and C.S. Chung, "Formal Genetics of Muscular Dystrophy," American Journal of Human Genetics, 1959, 11: 360-379.
58. V.A. McKusick and S.A. Talbot, "Analysis of Genetic Linkage in Man with Assistance of Digital Computer," Proceedings 1st IBM Medical Symposium, 1959, pp. 217-27.
59. B.G. Farley, "Similarities between a Neural Network Model and Electrophysical Experiment," Self-Organizing Systems, M.C. Yovits, G.T. Jacobi and G.D. Goldstein, eds., Spartan Books, 1962.
60. A. Walther, "The Spectrum of Information Processing," Proceedings of the IFIP Congress 1962, North Holland Publishing Company, p. 3.

nosis of disease⁶¹. Among the areas investigated have been hematologic disease⁶², coronary heart disease⁶³, electrocardiography⁶⁴, detection of placenta previa⁶⁵, detection of tuberculosis⁶⁶, detection of cancer⁶⁷, and others.

Even an incomplete summary of the literature on this subject leads one to conclude that there will be a great expansion of research into ways in which the computer analysis of medical data will aid the doctor in his diagnosis. It is not to be expected that the use of the computer in diagnostic procedures could ever supplant the human judgment of the skilled diagnostician, but it will help to organize the necessary data so that the doctor can make his own analysis more speedily and more surely.

Outlook

Because of the rapid changes and the extensive developments in the biological and medical sciences (many of which have not been discussed here), any predictions of future computer use must of necessity be based partly on guesswork rather than on extrapolation from current use. In the longer range, however, computer use in medical research and education seems destined to become substantial.

Within relatively few years, it is anticipated that half of the college population will see examples of computer use in the course of their undergraduate education. In scientific and technical areas, the majority of students in graduate schools will use computers in their research. Professional schools will include training in the use of computers as required in their professional fields.

Approximately one fourth of the doctoral degrees currently granted in the United States are in the biological sciences and in those areas of the social and behavioral sciences⁶⁸ in which quantitative methods and mathematical concepts are having an increasing effect. In addition, about 85 medical schools grant approximately 7,000 medical degrees each year. Thus, a large number of scientists will be entering these fields each year, and an increasing proportion of them will turn to computers in their research.

It is likely that requirements in mathematics and computer science in the training of medical students will increase rapidly in coming years as the role of the biomathematician in medical research becomes more fully established. In view of such developments, we anticipate the demand for some 34 Type-B computer equivalents by about 1970. A possible geographic distribution of these machines is shown in Table 22.

A summary outline of computer uses in the medical environment is shown in Table 23.

BEHAVIORAL SCIENCE

Economics, business administration, psychology, and linguistics have been using computers more and more extensively for almost a decade. In the past several years, professional fields like law and library science and such disciplines as anthropology, sociology, and demography have discovered important possibilities of computer applications to their problems.

Levels of national expenditure for research

Notes

61. (a) M.A. Woodbury and M. Lipkin, "Recent Developments in the Science of Diagnosis," Applications of Digital Computers, W. F. Fredberger and W. Prager, eds., Ginn and Co., Boston, 1963, pp. 28-32.

(b) L.B. Lusted, "The Mathematics of Medical Diagnosis," Proceedings 1961 Rochester Conference on Data Acquisition and Processing in Biology and Medicine, Pergamon Press, Inc., 1962, pp. 16-18.

62. M. Lipkin, R. Engle, B.J. Davis, V. Zworykin, R. Ewald, M. Sendrow and C. Berkeley, "Digital Computer as Aid to Differential Diagnosis," Archives of Internal Medicine, 1961, 108: 56.

63. L.D. Cady, Jr., M.M. Gertler and M.A. Woodbury, "Clues to the Development of Coronary Heart Disease," Geriatrics, February 1961, 16: 69-73.

64. M.A. Woodbury and L.D. Cady, Jr., "Mathematics of Pattern Recognition for the EEG," Proceedings of the Conference on Mathematical Problems in Biology, April 6, 1961.

65. T. Sterling, "Automated Placentography," Proceedings 1963 Rochester Conference on Data Acquisition and Processing in Biology and Medicine, to be published by Pergamon Press, Inc.

66. D.D. Rutstein, M. Eden and M.P. Schutzenberger, "Report on Mathematics in the Medical Sciences," New England Journal of Medicine, 265: 172-5.

67. E. Stern and W.J. Dixon, "Cancer of the Cervix—A Biometric Approach to Etiology," Cancer, 1961, 14: 153-60.

68. Agriculture and Psychology are included in this group.

UNIVERSITY OF CALIFORNIA, DAVIS

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO



SANTA BARBARA • SANTA CRUZ

KEARNEY FOUNDATION OF SOIL SCIENCE

265 HOAGLAND HALL
DAVIS, CALIFORNIA 95616

July 29, 1966

AUG 3 1966

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

Dear Professor Rogers:

This is to acknowledge your letter of July 25, 1966 and to thank you for the interest expressed.

I have a copy of "The Plant Sciences Now and in the Coming Decade" which I am sending to you on loan because of your special interest in the report.

The article which appeared in Bioscience was intended to supply American plant scientists with information which Dr. Price and I thought might be of interest to them but, it was in no way intended to reflect the views of the whole committee report which consists of 167 pages.

On page 138 of the latter, reference is made to "use of large scale computers by plant scientists". Roles for computers were brought up many times and debated from many points of view, both by the Committee of the Botany Section and by the Committee on Science and Public Policy as well.

From your letter, I gather that your visualization of computer needs by taxonomists may be considerably greater than has been implied by the panel.

The use of computers in science at large is blossoming out in all directions and I know that the National Academy of Sciences is cognizant of the situation. It is my feeling from conversations and discussions that I have heard on the subject, that projections of computer use for the next ten years are in a great state of flux. I have also gathered that there seems to be considerable aversion to having special computers for special projects as against having computer centers which will serve many functions. The debate is likely to continue for some time to come.

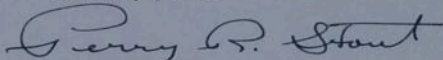
July 29, 1966

Chapter 8 of the report I am sending you is on "Taxonomy, Paleobotany, and Plant Geography". In the light of the full report, you will see that computers for plant sciences have been taken into account. If the real needs have been underestimated, at least the announcement has been made that computers are an important kind of instrumentation in plant sciences and that requests for their use will increase.

Perhaps the best way for you to accelerate recognition of the need in taxonomy would be for you to get in direct touch with fellow taxonomists and analyze their computer needs at fixed intervals. If growth rates turn out to be rapid, this fact could be called to the attention of the NSF, NIH and other agencies who would probably be instrumental in supplying necessary funds.

Please return the volume of "The Plant Sciences" when you have finished with it. I shall be glad to have further discussion with you on the subject but I doubt whether a paper from me would be nearly as effective as one from you since you are directly concerned with the problem and in a better position to assay the real need.

Sincerely yours,



Perry R. Stout
Director

PRS:vla

P.S. Copy of "The Plant Sciences" is being mailed under separate cover.

July 25, 1966

Dr. Perry R. Stout, Chairman
Dept. of Soils & Plant Nutrition
University of California
Davis, California

Dear Dr. Stout:

I am prompted to write after having read your paper in BioScience, "Plant science research strength in American universities, 1965-1975."

Inasmuch as the results of your work will probably have some rather fundamental effect on the thinking processes for the various granting agencies and policy within the National Academy of Sciences, it seemed to me that I should point out what I feel to be a fundamental omission with respect to plant taxonomy. Of course, I realize that you cannot be on top of the latest developments in all of the disciplines of botany, but there is one place where I feel that your sampling technique broke down, and that is that the newer endeavors in taxonomy were not at all reflected by those taxonomists who answered the inquiry.

Taxonomy, much as plant physiology and biochemistry could not begin to be a useful or fundamental contributor to modern biology until certain methodologies and techniques from chemistry came forward, could not contribute greatly until the development of certain pieces of equipment, particularly the computer. Although a survey at this date would not indicate more than a half dozen people who are using a computer in taxonomy, particularly plant taxonomy, I think you will discover that in the next decade the growth of the use of the computer for this field will be phenomenal. Although I am one of the few taxonomists who specifically devote their attention to the applications of computers to the complexities of plant classification at this stage of the game, I know that any number of young Ph.D.s and graduate students around the country are vitally interested in methodologies dealing with the computers.

Since this is the case, there should have been some reflection in your survey of a greatly increased need for computers, and therefore, a vast expenditure of funds to put computers in several taxonomic centers in the country. The types of computer hardware needed for plant classification is not the small core, small memory computer but the very largest computer with the greatest speed and storage capacity. These pieces of equipment must be primarily under the control of the taxonomist

July 25, 1966

rather than to be a university-centered computer, that is, the primary responsibility and control and use of these computers will be for taxonomy alone with only secondary utilization by such people as mathematicians, etc. The cost of the computer centers is astronomical as you know— (that is, astronomical in terms of biological expenditure). But their application is such that the quantum jumps to be made in taxonomy as a result of their use will certainly justify their cost.

It seems to me that it is critical to add an addendum to your paper, perhaps in a subsequent issue of BioScience which helps to show this need. I would be glad to work with you to help flesh out the information about the needs for computers and classification, and also to point out in such a paper the new direction that taxonomy inevitably has to take.

I am looking forward to having your comments on this letter.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR:chh

August 22, 1966

Dr. Daniel B. Ward
Herbarium
Institute of Food and Agricultural Sciences
University of Florida
312 Rolfs Hall
Gainesville, Florida

Dear Dan:

I am sending under separate cover three mimeographed or copied papers: 1) the orchid characters, 2) flow chart, and 3) a paper on characters. Items two and three are privy communications not for general distribution as they are both in press. Item two is the actual programmer's flow chart for the clustering process. This is a program that has been written for the IBM 7044. It is a later edition and we think a cleaner program than the one that we had running in New York on the CDC machine. The only part we do not like is the sort routine. You should probably insert a standard package sort routine.

Anyone wishing to use the CDC program would have been in very great difficulty because the thing was under the control of the programmer to such an extent that unless you were the programmer who did the work originally, you would have a difficult time knowing what the directions for the operation of the program were. The flow chart sent to you allows anyone with good programming skill the ability to write the program for the individual machine that you wish to use. This would almost be the case under any circumstances.

As you may know, no two computer installations, even though they have the same manufacturer and number designation, are identical in their configurations. And for each installation, one has to have a few special instructions. Therefore, the best thing to do is to start with the flow chart and follow it through. The flow chart is, incidentally, to be published in the SHARE Library soon.

The second mimeograph paper is a preprint that we have submitted to BioScience for publication. This is an essential before you can use the computer program. We have discovered that the most critical element in the whole process of the use of computers for classification of plant materials is the structuring of the input data. We, ourselves, have had a long learning period in order to be able to put out the kind of paper that we are sending to you. Unless one is willing to be very diligent in the preparation of one's data for a computer, no matter how good a

August 22, 1966

program be, you will have no results unless your data be prepared adequately for it.

We have had a number of people to collaborate with us in the use of our clustering methods, and it has been our experience that these individuals do their best job after they have had some personal instruction from us about their own data. This is in respect to how to set up the data if one is to appropriately use the machinery. It does not refer to the scientist's own ability to recognize pieces of information which will make good characters. One of the things that we have discovered is that the most critical element in any kind of work, whether he be using a computer or not, is a specialist's own skill in taxonomy. We are trying to discover how one may achieve this skill on a better than trial-and-error basis, but so far the *modus operandi* has not been completely worked out. If it had been worked out, you may be sure that you would have seen, for example, a monograph of the genus Manihot by now.

I will be very interested in your comments after you have had a chance to read through both of the documents that I am sending to you. I would like to know in what manner you would like to utilize these programs, and I would further be very pleased if you would let us know of any problems that you have. We will try our best to answer specific questions.

Good luck and keep it up.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

Enc.

UNIVERSITY OF FLORIDA
INSTITUTE OF FOOD AND AGRICULTURAL SCIENCES



COLLEGE OF AGRICULTURE AGRICULTURAL EXPERIMENT STATIONS AGRICULTURAL EXTENSION SERVICE SCHOOL OF FORESTRY

HERBARIUM
312 ROLFS HALL
GAINESVILLE, FLORIDA

August 9, 1966

Dr. David J. Rogers
Department of Botany
Colorado State University
Fort Collins, Colorado 80521

Dear Dave:

In addition to your other talents you are ^{prescient!} ~~precendent~~. This morning I was planning to write you for a copy of your orchid paper, when, lo, the mail brought me one.

I did, however, want to ask also for two other things. First, I would appreciate a copy of your C. D. C. program for your clustering process. Second, and particularly at the request of John Beckner, who is as interested in the orchids as I am in the math, could you send a list of the 120--odd characteristics you used for the gathering of your raw data? I am glad you are continuing to work with "taximetry". Still, it is startling to find a paper on orchids in the Systematic Zoology.

Sincerely,

Don

Daniel B. Ward
Curator

DBW/cjd
cc

*Dept. Biol.
Mo. Valley College
Marshall, Mo.*

- Taxonomy Laboratory

July 6, 1966

Mr. George F. Freytag
Head, Agronomy Department
Escuela Agricola Panamericana
Tegucigalpa, Honduras

Dear Mr. Freytag:

Dr. Rogers left for Mexico July 1 in search of more Manihot.
He expects to return around the end of July and will answer your
letter of June 30, 1966 as soon as possible thereafter.

Sincerely,

(Mrs.) Cleo Hepworth
Secretary

No reply needed
DGR 7/21



30 June 1966

JUL 5 1966

Dr. Dave Rogers
Taxonomy Laboratory
Department of Botany
Colorado State University
Fort Collins, Colorado 80521

Dear Dave:

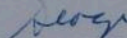
Thanks a million for your letter. Damn sorry you haven't gotten this all worked out yet! I was really hoping to get out of a lot of work. And now...

Could you do me a favor and tell Dr. Edwin James that I will be writing him one of these days? Nobody of importance in the Rockefeller Foundation was at the meetings so couldn't do any good, but am expecting a visit any day now from the Turrialba people and expect to get going on the coming year's program, which by the way, is to be in Costa Rica instead of hear. The new Minister of Agriculture there is an old frijolero friend of the group and insisted that they hold the meetings in Costa Rica while should yet be in office. Would you give him this information please?

I am writing Dr. Dietz and IBM as you suggested. (Is this 'the' Dietz that studied with Dr. Anderson?) I do hope they have some relatively simple and inexpensive card system which can be used or adapted to our needs. You will probably be hearing from me again.

Best wishes,

Sincerely yours,


George F. Freytag
Agronomy



17 June 1966

Dr. Dave Rogers
Taximetrics Group
Department of Botany
Colorado State University
Ft. Collins, Colorado

Dear Dave:

Sorry to take so long to get around to writing you. Usually when a person has a need...

First, let me tell you how glad I was to have run into you after all these years. It was quite a surprise to me, and only sorry I didn't have time to renew acquaintance with the whole family.

For a number of years now I have been working towards getting a good seed storage facility, and finally got the money for it last year. We are now well on the way to seeing the storage rooms take shape and hope to have them operating in a month or two. Since these have been undertaken with AID money we have offered to share them with the Central American Common Market Countries and have been asked by their general committee to submit a project for this. In a way I was unprepared to be taken up on the offer so soon, being accustomed as I am to the latin way of life, so that I submitted a program outline and budget without really having adequate time to plan it.

Nevertheless I have had the faint idea that the crop seed so stored should be filed systematically and that all seed should have a file on it with all available data. I have in mind the collection data and subsequent testing (yield, disease resistance, etc.) as well as the status quo (amount and viability, etc.). All of this should be available in a system better than a notebook presents and I had in mind something like an IBM card file, but having almost no experience in these things (except in the form of checks and even then have to read the thing) I really don't know if it's feasible. So...

Have you ever thought of having a system such as this? perhaps as a card file to an herbarium as a library has? Could you help me devise such a system for a Genetic Bank (or better yet do you have a system ready to go?) As I am innocent I placed in my budget 3,000 dollars for a punch, sorter, and filing cases. Certainly IBM makes something simple for 110 v. 60 cy. to work out in the boondocks. Can you help me?

Sincerely yours,

George F. Freytag
George F. Freytag
Head, Agronomy Dept.

- Taxonomy Laboratory

June 20, 1966

Mr. George F. Freytag
Head, Agronomy Dept.
Escuela Agrícola Panamericana
Tegucigalpa, Honduras

Dear George:

In response to your letter of June 17 on the setup of an information system for your seed storage activities, I have the following comments.

First of all, I myself do not have an information system going. There are several reasons for this. One, we find that we have to do considerable amount of research before we can adequately establish a big and workable generalized system. This requires a considerable amount of effort, not only from the biological input standpoint but from the organization and development of sufficient and adequate models.

If I were you and if you plan to have an IBM punch and sorter, I would write directly to IBM and ask them for all the publications that they have available on information retrieval systems using a punch and sorter. I would send my inquiry into IBM to their offices on East Post Road, White Plains, N.Y., suggesting your problem, the nature of your problem, the kind of hardware or equipment you have and ask them for a publication of all sorts which will help you to get started.

You can also gain assistance in this respect by writing to Dr. S. M. Dietz, Regional Plant Introduction Station, Johnson Hall 59, Washington State University, Pullman, Washington 99163. From Dietz' address you can see that he is in the introduction business for USDA. He has been attempting for sometime and has done quite a bit of work towards some sort of a punchcard retrieval system for entries or introductions and has written a number of pamphlets which may be useful to you in trying to sort yourself out for which way you want to go. I just talked to Dr. James who is the head of the National Seed Storage here at Colorado State University. I understand that you have corresponded with him previously. I asked him what sort of system they used for their work and discovered that he uses an edge punchcard, I believe the Royal McBee cards. In his conversation it is pretty well indicated that his system is too small to do a really bang-up job of automatic data retrieval, but he seems to be satisfied with what

June 20, 1966

he's got. I would suggest that you write to him asking him about how much information about each of your accessions you should have and how to establish that.

Frankly, the most important part of an information retrieval system is for you yourself to decide how big the thing is going to eventually be. You have to be the one who directs how much information will be kept. I must, at the same time, try to warn you that once you get into one of these systems using any kind of IBM equipment, you open up Pandora's box, because the methods and procedures used are frantic in numbers and apparently none of them entirely satisfactory. However, as a biologist you should stick to your guns primarily and not try to delve into the business of how to set up your cards and so forth. You can get plenty of information about that sort of thing from the IBM literature.

Whatever you decide on, you are going to have to do an awful lot of homework in order to establish this thing in a way which will be meaningful to you. As I indicated earlier, we are ourselves beginning the study of information retrieval systems for biology. As I also indicated, this is something which is requiring a considerable amount of investigation in the procedures themselves, not on the data but on how to take care of the frantic amount of information that biologists want to stash away and then find again some day. I hope that we can get funds ourselves from somebody to make a big investigation along these lines. Our investigations, of course, deal not with just the punch machine and the sorter but with the main frame of the computer itself. These devices, the computers, will be the only answer to a large information system. This is the problem that we work on.

Good luck on your project and let me know how you are proceeding. I know I haven't answered many of your questions, but I hope you have gotten some leads from where I've started.

Sincerely,

David J. Rogers
Professor of Botany

DJR:ch

- Taxonomy Laboratory

April 19, 1966

Mr. Hal Williams
Graduate School Computer Center
University of Colorado
Boulder, Colorado

Dear Mr. Williams:

A situation arose on the enclosed job 12409 which seems to be exactly like the one encountered recently on job 11069--that is, the run was terminated as the result of what seems to be an operator error in positioning tape 4.

I refer you to the enclosed Memory Map from job 12409. We are in the process of checking out decks PINE through ELM. We run decks CALDEK through TEMP in order to generate input to PINE. You may remember checking these source programs last time this occurred to see if tape 4 was properly handled and you concluded that all was well with the program, especially since it has run a considerable number of times with this data. Your conclusion was that the operator did not reset the tape after a read error and consequently you did not charge us for the job. I think you will probably draw the same conclusion again and I hope that you will handle it the same way.

If you would like to see the source program listings for decks CALDEK through TEMP or any other documents that I can provide, I will be glad to send them to you.

Sincerely,

Robert C. Brill

RCB:chh

Enc.

cc: Dr. D. J. Rogers ✓

- Taxonomy Laboratory

March 4, 1966

Dr. B. L. Shaffer
Continental Oil Company
P. O. Box 1506
Shreveport, Louisiana 71102

Dear Dr. Shaffer:

Your request for reprints and information on our computer programs was forwarded to me at the above address. We have moved our operations to this institution where we are continuing investigations on computer-aided classificatory procedures. At present, my group consists of two biologists, a mathematician, and a programmer.

The reason I explain this is to let you know that, if you are interested, we can provide competent advice on a wide range of related endeavors. Also, we are interested in a varied set of applications of our more recent programs, some of which you may find interesting. Your endeavors sound interesting to us, and different from other applications we have had

One of our programs, devised since the publication of the paper in BioScience, is described in a paper to be published in the next issue of Systematic Zoology. The mathematical basis for this particular program is in press in the Jour. of Theoretical Biol. The date of appearance of the latter paper is not certain. However, the program described in these two publications has largely replaced (in our lab.) the procedures we described in BioScience. While the methodologies of the information theory model have much to recommend them, we found a sufficient number of fundamental errors in them to cause us to shelve that program (the program decks are still available) in favor of one based on an entirely new mathematical model. Having tested this new program on a number of different problems, we find ourselves much better satisfied with it than the other. If you are interested, program decks of this new model can be made available for FORTRAN 63, CDC 1604, and will shortly be reprogrammed in FORTRAN IV for the IBM 7094 (which we are using here).

While the above described program is in operation, using as

March 4, 1966

input much the same types of data (formats are almost the same as those described in BioScience), we have, within our own group, learned to be much more sophisticated in the preparation of the data. We are now in the process of describing some of these new sophistications and modifications of data preparation. The ms. is through one rough-draft. This means, of course, that the paper probably won't appear for at least eight to ten months. If some of these newer ideas are incorporated, one's work has an even higher chance of giving the desired results. If these sophistications and modifications interest you sufficiently, we could tell you about them at some mutually convenient time.

Since I am not certain of your desired application, and whether or not you are interested in other pieces of the classificatory process, I will not mention other, equally interesting procedures which are now available, or soon will be. I shall be interested in your response.

Sincerely,

David J. Rogers
Professor of Botany

DJR/pam

Encl. - reprint

MAR 3 1966



CONTINENTAL OIL COMPANY

P.O. BOX 1506

SHREVEPORT, LOUISIANA 71102

Feb. 23, 1966

Dear Dr. Rogers:

At the present time I am in the formulative stages of applying a computer program for the classification of fossil spores, pollen, and microphytoplankton as well as various geologic data, and am finding your ideas and information most useful. I would greatly appreciate receiving a reprint of the following paper:

Rogers & Fleming, 1964

"A Computer Program For Classifying Plants - II A Numerical Handling of Non-Numerical Data" Bioscience, V. 14, #9, p. 15-28.

If you have any other available reprints on this or related topics, I would be honored to receive copies of them also.

Sincerely,

Bernard Shaffer

Dr. Bernard L. Shaffer