

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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD. 20014
AREA CODE 301 TEL: 656-4000

January 8, 1969

In reply refer to:
LM 00724-01

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

Dear Dave:

Many thanks for your candid comments about the Hersey application, LM 00724-01. I think you hit the nail on the head in remarking about his lack of homework. It is always discouraging to see such an approach when there may be a kernel of good imbedded somewhere. Thanks for taking the time to read and comment. Your remarks will be very useful to the Study Section members when they meet later this month to consider the application.

My best wishes for the New Year.

Sincerely,

Ileen E. Stewart
Executive Secretary
Biomedical Communications
Study Section
Division of Research Grants

Digitized by the Hunt Institute for Botanical Documentation

3 December 1968

Mrs. Ileen B. Stewart
Biomedical Communications Study Section
National Institutes of Health
Bethesda, Md. 20014

Division of Research Grants
Westwood Building, Room 2A-14

Dear Ileen:

Sorry we waited so long to get this review back to you. My opinion on the thing is that Hersey has not done his homework before coming to you and asking for 65 Gs. I think he should definitely go back and re-write this thing and come back in with a great deal more detail. Evaluations such as his should have an economist on the staff and I don't believe I see one. I am sure they have something in mind which would be useful, but frankly it is not written down here. I hope you will encourage Hersey to go into more effort to say what he is going to do, because I would like to see such a thing supported.

Sincerely,

David B. Rogers
Professor of Biology

DJR:qm

Digitized by the Hunt Institute for Botanical Documentation

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

NATIONAL INSTITUTES OF HEALTH
BETHESDA, MD. 20014
AREA CODE 301 TEL: 656-4000

In reply refer to:
LM 00724-01

November 13, 1968

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

Dear Dave:

Grant applications received by the U. S. Public Health Service in the area of scientific publications and communications research are evaluated by the Biomedical Communications Study Section. Although the members of this Study Section bring a wide diversity of expertise to bear upon proposals, from time to time the assistance of ad hoc reviewers becomes necessary. We hope that members of the scientific community will be willing to provide the occasional evaluation so that meritorious research can receive support.

I am enclosing a copy of grant application LM 00724-01, David F. Hersey, and a brief list of items on which we would appreciate comments. I hope that you will be willing to examine this application and give the Study Section the benefit of your evaluation. A form on which you may submit your comments is enclosed; a franked, addressed envelope is also included. It would be helpful if your reply could reach us no later than November 29, 1968.

Inasmuch as this application and the attachments are privileged communications, would you return all materials with your comments. Your reply will also be considered confidential and made available only to Study Section members and staff. If any questions arise about this application, please feel free to call me on area code 301, 496-7581.

Sincerely yours,

Ileen E. Stewart

Ileen E. Stewart
Executive Secretary
Biomedical Communications
Study Section
Division of Research Grants

CE

Enclosures

Digitized by the Hunt Institute for Botanical Documentation

December 31, 1968

Mr. Harold A. Olsen
Office of Science Information Service
National Science Foundation
Washington, D.C. 20550

Dear Harold:

I enclose my response to your request for evaluation of the Flora North America project. I have followed the outline given in your letter of Dec. 23, to answer the questions. The evaluation is not signed, in case you want some sort of anonymity, but I personally do not mind if you use my name as the author of it.

Hope the answers are pertinent, and clearly, I'm willing to answer any more, if you have need.

Perhaps I'll be in some time toward the end of January, to give a seminar to some NSF group. Ernie Sohns is organizing it, I think. If I do come in, I'll be sure to look you up.

Sincerely,

David J. Rogers
Professor of Biology
Project Director

DJR:sw

Digitized by the Hunt Institute for Botanical Documentation

Comments on Proposal M-3155, "An Information System for Flora North America (FNA): Planning Stage." Responses given in same order as those in cover letter from Harold A. Olsen, dated December 23, 1968.

(1) Without question, the objectives of this effort are useful to a large number of life scientists. Inasmuch as the work of systematic botanists is largely in information retrieval, their efforts should be considered as fundamental procedures in all the life sciences. Given the addition of computing machines and good programs for them, a large body of workers will be well-served. That is, the concepts of biological information are well preserved in the workings of taxonomists, and the biological fraternity uses taxonomic output and methodologies, whether they recognize it or not. Therefore, there should be produced a very useful set of methodologies by funding this proposal.

As this proposal is conceived, there should be beneficial effects for other groups active in this area by making the separate groups more aware of what types of programs are being produced in the several places in this country and abroad, and perhaps foster more cooperative effort with less duplication. Clearly, this type of thinking is a two-way street, with as much going into the FNA effort as is coming out. Cooperation must be the main theme fostered by the FNA in its dealing with the numerous producers of programs which can be useful in various ways for the total FNA effort.

(2) It is difficult to conceive of any but adequate consideration to the current state-of-the-art in this proposal. After all, the funds are largely requested for a Systems Development Manager, and not for actual programming activities. As I see it, the SDM will gather together useful programs from around the country, survey them and incorporate them or ask the producers to modify in certain ways to meet some specific programming need for FNA. Clearly, the SDM must be above average if he is to evaluate this or that program for its generality, power, elegance, sophistication in hardware use, etc. He will certainly need to be educated by those of us who have had rather intimate contact with the types of problems in FNA.

(3) This is the weakest part of the proposal, in my estimation. I would require the principal investigator to name those in the botanical profession who will form the most knowledgeable back-up for the SDM. There is quite a bit of sophistication in our laboratory, and at one or two other places in the country. These people should be clearly identified by the principal investigator so that you can check the relevance of his advisors. The fact that I am a member of the Advisory Committee for FNA does not guarantee that the principal investigator will seek my advice. While I have no reason to believe that Shetler will not ask for my help, I am not assured that he will.

Digitized by the Hunt Institute for Botanical Documentation

To achieve fruitful cooperation, I recommend that a standing sub-committee be established to aid in this aspect of the work of the SDM. I am thinking here of identifiable botanists whose efforts have already indicated some value for FNA. This committee should be formed after the appointment of the SDM, and formed by sending questionaires to potentially interested botanists. The questionaire would ask if the individual or group wished to serve, and in what capacity the individual or group could function. Then, upon forming of a committee be certain, someway, that the SDM works with this committee.

I would further ask Shetler to prepare, with the idea that the preparation is preliminary only, a flow chart of the workings of the FNA as these impinge on IR systems. What will be the various types of input and output requirements of an IR system? What will the SDM finally devise? I think it possible to identify in advance much of the work the SDM has to do, and it would probably be advantageous at the point of hiring the man if he could be shown what types of programs will be useful to FNA.

One final bit of advice which I think valuable is that some sort of cost/effectiveness study be made. This isn't given in the proposal. I do not think that there need be a large budgetary factory to fulfill this type of task. We already have a pretty well worked-out system for making such an analysis, and only a few hours will be required of skilled time to direct such studies. I think that FNA and its SDM would be very much pleased to see the results of such studies. They are very enlightening.

But above all, I recommend the funding of this effort. I haven't seen as much cooperative effort exhibited by taxonomists ever before in this country. They seldom tell you why their work is good and useful, (and perhaps I haven't been too explicit myself) but let it be said that there is a fundamental importance to knowing what plants we have, where they are, and how do you identify or locate them. This will be the minimum value derived from the FNA effort, and the maximum is not known by any of us.

Digitized by the Hunt Institute for Botanical Documentation

INFORMATION SYSTEMS PROGRAM

Review of Proposals

Administrative

A proposed effort qualifies administratively for support if

1. the applicant's underlying purposes lie within the authorized scope of the OSIS (e.g., "a more effective dissemination of scientific information" and "new or improved methods of making scientific information available").

2. the applicant represents (i.e., is accountable to) a nationally significant scientific community and has accepted the responsibility for operation and development of that community's specialized information system.

Technical

A proposed effort qualifies technically for support if its expected value exceeds its cost. Accordingly, the review considers the potential value of the specific objectives to be pursued, the probability that they will be achieved, and the cost of achieving them.

3. Proposed objectives (e.g., extensions in the scope of an information system; improvements in its performance, efficiency, and dependability; and advances in an information system program) are weighed in relation to the system program in which they are embedded. Their valuation by a responsible applicant is ordinarily accepted as reasonable.

The probability of achieving the stated objectives depends largely upon the applicant's readiness to pursue them. His readiness is assessed in consideration of the technical approach which he proposes and of his ability to carry out that approach.

4. The proposed technical approach is reviewed to determine whether (a) it takes adequate account of the state of the art, related work, and available resources; (b) the component tasks are individually necessary and collectively sufficient to achieve the stated objectives; and (c) the planning for each task is realistic as to initial conditions, duration, and the application of resources. A responsible applicant's choice of approach is ordinarily accepted as reasonable.

5. The applicant's facilities, staff, and organization are reviewed in relation to his proposed technical approach. Here again, a responsible applicant's judgment as to his own adequacy is ordinarily accepted as reasonable.

Digitized by the Hunt Institute for Botanical Documentation

The only remaining technical consideration is that of cost.

6. The cost of achieving the stated objectives is estimated on the basis of the resources to be applied in the performance of the specific tasks proposed.

Priority

When funds are insufficient for the support of all the qualified efforts proposed, competing efforts are funded in order of priority.

7. Within a competitive class, higher priority is given to those proposed efforts whose expected value in relation to cost is higher.

HEBamford:dm
OSIS/ISP
5/28/68

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION

WASHINGTON, D.C. 20550

Office of Science Information Service

December 23, 1968

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

Dear Dr. Rogers:

We request your assistance in reviewing the enclosed proposal, N-3185, "An Information System for Flora North America (FNA): Planning Stage," which we are currently considering for support though at a reduced budget level. We would like your review to address the following specific questions:

(1) Do you regard the achievement of the objectives of this effort as a useful contribution to information systems development efforts in the life sciences generally? Will the outputs of this effort provide any potential benefits to other groups active in this area?

(2) Does the proposed approach promise adequate consideration to the current state-of-the-art and to the work of other related efforts active in this subject area?

(3) What views have you concerning possible cooperative relationships between the FNA project and other groups with similar interests? What conditions do you believe necessary for fruitful cooperation to occur? What steps do you recommend to achieve such cooperation?

We would appreciate your response by January 15, 1969. Thank you for your assistance.

Sincerely,

Hal Olsen

Harold A. Olsen
Assistant Program Director
for Information Systems

Enclosure

Thank you again for your informal counsel on this. If it is not too much of a further burden or imposition, I would appreciate some comments for the record.

Digitized by the Hunt Institute for Botanical Documentation

PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

TITLE:

"An Information System for Flora North America (FNA): Planning Stage"

SUBMITTED BY:

American Institute of Biological Sciences (AIBS)
3900 Wisconsin Avenue, N.W.
Washington, D.C. 20016

ADMINISTRATOR:

John R. Olive

John R. Olive, Director
American Institute of Biological Sciences
Telephone 362-6733

PROGRAM DIRECTOR:

Peter H. Raven

Peter H. Raven, Associate Professor
Department of Biological Sciences
Stanford University
Stanford, California 94305 Telephone 415-321-2300 X3168

SECRETARY:

Stanwyn G. Shetler

Stanwyn G. Shetler, Associate Curator
Department of Botany
Smithsonian Institution
Washington, D. C. 20560 Telephone 381-5801

DURATION OF EFFORT:

30 months, beginning when grant is awarded

TOTAL FUNDS REQUESTED:

$187,869

DATE SUBMITTED:

30 September 1968

Digitized by the Hunt Institute for Botanical Documentation

SUMMARY

The overall purpose of the Flora North America (FNA) Project is to produce a new synthesis of knowledge about the vascular plants of North America north of Mexico through a massive, 15-year cooperative effort supported by an information system based on the latest computer science and technology.

The objectives of the specific effort proposed here are (1) to develop the concept for the FNA information system, and (2) to produce the program plan for designing and implementing the system conceived. Insofar as the information system is concerned, therefore, this is the initial planning stage of the FNA Project, and it should result in a full program definition.

To accomplish these objectives the AIBS proposes to take the following steps over a period of 30 months:

(1) Employ an executive-level Systems Development Manager (SDM) for two years, starting as soon as possible but no later than six months after the grant is awarded, who will be responsible for developing the goals, concept, and program plan for designing and implementing the FNA information system during the next stage of the FNA Project. The SDM will consider all aspects of data input, processing and storage, output, and communication.

(2) Impanel a small group of experts immediately who, serving as an ad hoc SDM Task Force, can assist the AIBS in defining the tasks and qualifications of the SDM, in recruiting the right man, and in guiding both the SDM and the AIBS while the SDM is conducting his planning study. After the SDM is appointed, the Task Force may be enlarged, because it will become more of an advisory body and less of a task force for the remainder of the grant period.

(3) Convene the AIBS-FNA botanical committees and other botanists and biologists as often as necessary to advise the SDM Task Force and the SDM on the botanical objectives of the FNA Project and to consult regularly on the development of the information system. The financial resources to accomplish this are mostly requested in the main FNA proposal to the NSF, submitted to the Systematic Biology Program in December 1967.

(4) Submit the SDM's final concept and program plan to the appropriate bodies of the AIBS for discussion and approval.

The SDM will be responsible to the Director of the AIBS, as are all members of the AIBS-FNA committees, and he will become an integral member of the FNA Project's working group, the AIBS-FNA Editorial Committee. Physically and administratively he will be housed for the present in the FNA Secretariat, and on a daily basis he will be expected to work effectively with and through the FNA Secretary, who has the responsibility for the overall coordination of the FNA Project.

The funds requested here are, in the opinion of the AIBS, both sufficient and necessary to achieve the objectives outlined.

Digitized by the Hunt Institute for Botanical Documentation

TABLE OF CONTENTS

Introduction: Flora North America Project 1
Overview 1
Objectives 2
Reasons for objectives 2
Users 5
Institutional responsibility 5
Stages of development 6
Proposed Planning Effort 7
Objectives 7
Resultant program plan 7
Work plan 8
Approach 8
System considerations 9
Timetable 11
Personnel 12
Ad hoc SDM Task Force 12
Systems development manager 13
Secretary 13
Facilities 14
Budget justification 15
A. Salaries 15
B. Permanent equipment 15
C. Expendable office supplies 15
D. Reproduction 15
E. Communications 16
F. Printing and production 16
G. Travel 16
H. Honoraria 17
I. Miscellaneous other costs 18
Budget summary 20
References 23

Fig. 1. Organizational Chart of AIBS-FNA Project 24

Appendix I. Brief Biographies of SDM Task Force Members 25
Appendix II. List of Members of AIBS-FNA Steering and Editorial Committees 27

Digitized by the Hunt Institute for Botanical Documentation

INTRODUCTION

Flora North America Project

Overview

Flora North America (FNA), initiated in 1966 by the American Society of Plant Taxonomists (ASPT) after appropriate feasibility study, is an international project conceived and organized to produce a new synthesis of knowledge about the vascular plants of North America north of Mexico during the next 15 years. In the future the effort possibly can be extended to other plant groups and to the whole of North America.

The project is being administered by the American Institute of Biological Sciences (AIBS) through three permanent committees appointed by recommendation of the ASPT. The AIBS-FNA Editorial Committee, including eight American and Canadian botanists, has the chief responsibility for developing and executing the program; its Chairman serves as the Program Director, and its Secretary manages the Secretariat, the executive arm of the committee. The AIBS-FNA Steering Committee, including seven regular and three ex officio members from Canada and the United States, provides policy guidance to the Executive Committee of the AIBS. The AIBS-FNA Advisory Council, comprising 30-50 leaders in the botanical and biological communities of Europe and Mexico as well as of Canada and the United States, is, as its name implies, an advisory body that consults and guides the other committees as needed. This advice may be offered by the members individually or as a body. The Advisory Council has no executive function. An organizational chart is presented in the appended Figure 1.

The Smithsonian Institution is host to the Secretariat at least for the present, but the approximately 50 committee members are located in almost as many different institutions in North America and abroad. As the project develops, significant tasks or parts of the work will be performed at many of these and other institutions. Furthermore, it is possible that the project will require its own facility eventually, particularly if an information center of some type is established.

The comprehensive proposal for the 15-year effort, entitled "Flora North America Project," was submitted by the AIBS to the Systematic Biology Program of the Division of Biological and Medical Sciences of the NSF in late December 1967 (see references). This main proposal sets forth in detail the "General Long-Range Plan" and the "Implementation Plan," and it documents "Current and Pending Support" as of November 1967. Reference should be made to the main proposal for full elaboration of the principal program objectives, past and present justification of goals, feasibility

Digitized by the Hunt Institute for Botanical Documentation

planning, AIBS-FNA organization (committees, etc.), writing and editorial procedures, educational objectives, overall timetable, and budgetary projections for the 15 years. All committee members are listed in the main proposal, and brief biographies are given for the members of the Editorial and Steering Committees. Details are not repeated in the present proposal, but cross-reference is made at appropriate places in the following text.

The proposal herewith submitted represents a complete reformulation of those parts of the main proposal (pp. 38-45, 54-62, 85-86) which deal with "automatic data processing" ("ADP"); therefore, with respect to ADP this proposal supplements and supersedes the main proposal. (Hereafter the more widely used term "electronic data processing" ("EDP") will be used.) This proposal is written to stand by itself, however, so that recourse to the main proposal is unnecessary except for background.

Objectives

(1) The program goal of the FNA Project is to prepare for the first time a concise, diagnostic treatise or "flora" of all the vascular plants of North America north of Mexico, based on the existing literature and botanical collections and on appropriate new collecting and research carried out by the botanical taxonomic fraternity at large over the next 15 years.

(2) The supporting goal is to develop and implement a comprehensive information system based on the most advanced EDP science and technology that will result in both a taxonomically-structured data bank and an innovative "kit of computer tools" or flora software to serve the FNA investigators for the duration of their program and ultimately to serve the whole botanical if not biological community for the indefinite future.

(3) An intangible goal is to stimulate new interest in taxonomy and the study of the higher levels of biological integration--organisms, species, populations, communities, floras--especially among the coming generations of students who should find in the FNA Project the rationale, unifying and organizing genius, and advanced technical resources to conceive and carry out taxonomic researches on a scale hitherto impossible.

Reasons for Objectives

The need for a modern synthetic treatise of the type projected has long been recognized. Now that Russian botanists have completed such a treatise for the USSR and European botanists are at work on one for all Europe, it becomes the more urgent for North American botanists to take up the task of closing the circumpolar floristic gap (see main FNA proposal, pp. 2-12). Much less evident, perhaps, is the absolute necessity

Digitized by the Hunt Institute for Botanical Documentation

of the second objective, namely, to place floristic-taxonomic research on a modern, computerized, information-system basis.

In a sense, taxonomists, plant or animal, have always been the keepers of the biological information system. In the plant sciences the floras and systematic monographs are timed-honored, if non-automated, storage and retrieval systems which through the years the plant taxonomist has produced and revised, as both the means and the ends of his own research, for the benefit of all who need to know about the kinds of plants in the world. The hierarchical classification of the organisms, with its binomial nomenclature providing unique addresses for parceling and storing information, has afforded the ideal means of structuring the data file.

Today the taxonomist no longer is able to cope with his historic responsibilities. Published knowledge has accumulated to the point where it no longer can be synthesized rapidly and concisely by conventional methods and media--not even to the extent of keeping the classificatory system itself up to date, to say nothing of handling new information within this system. Likewise, botanical collections (herbaria) have grown to the size where they no longer are comprehensible to the individual scientist, and thus a great wealth of information often goes untapped for want of sufficient manhours to search and retrieve. Even the seemingly simple task of supplying the correct name or an accurate description for a native or cultivated species of plant may require hours or even days of valuable professional attention, owing to unstandardized terminology, outdated nomenclature, or dispersed and unwieldy sources of information. Yet in this day of molecular biology and ecosystem ecology, when it often is a matter of national interest to be able to correlate many precise, authoritative facts about our plant resources in the shortest possible time, biologists, public leaders, and laymen are depending on the taxonomist more than ever, though perhaps unwittingly, even while his power to deliver is diminishing ever faster.

The time is overdue for plant taxonomists to design and implement a modern information system, using the latest methods of electronic data processing and communication, so that they can regain their historic control of the biological data bank. The core of this system must be a dynamic, multi-access, taxonomically-structured electronic data bank, designed for rapid and flexible input, processing, and output. Centralized or decentralized, it must be accessible to all potential users by means of the most advanced communications network that is economically feasible. We propose, therefore, that Flora North America be developed as such a dynamic, computerized data bank and that it not be developed merely as a "one-shot," static publication in which, relatively speaking, the state of knowledge would be rigidly embalmed. Many books and other forms of output should be possible as more or less automatic by-products of the data bank, and instant update, revision, and cumulation would be cardinal attributes of the information system. This is not to deny the initial and continuing goal of producing a flora of conventional book form, perhaps in 4-5 volumes, at some stage. For the foreseeable future this seems

Digitized by the Hunt Institute for Botanical Documentation

likely to remain the most practical and economical form for everyday use in identifying plants and retrieving distilled information about their morphology, ecology, and geography. However, the production of such a book cannot be set as our only goal for FNA, but rather as our minimum goal.

There is another aspect to the development of an FNA information system. The preparation of any flora or taxonomic monograph entails, in addition to the new research and genuine intellectual input, a large element of purely routine work--searching, abstracting, compiling, comparing, editing, revising, standardizing, and formatting data. In fact, it is reasonable to estimate that upwards of two-thirds of much taxonomic scholarship is concerned with such tedious, non-creative tasks, while only about one-third is devoted to true scientific endeavor. Thus the taxonomist often feels more like a scribe than a scholar. Much of the compilation involves basically repetitive information, which has been copied from generation to generation and book to book, thereby compounding errors indefinitely. Yet the various tasks of compilation, editing, etc., are so encumbered by tradition, subjectivity, and, above all, force of habit, that they have remained embedded in the scientific process itself, so that only a trained and experienced taxonomist can perform them, even though many of these tasks could have been reduced long ago to objective routines capable of being performed more consistently, efficiently, and rapidly by machine.

The individual taxonomist, regardless of his desire to update traditional procedures, has neither the authority nor the resources at his command to bring about the kind of innovative automation needed. On the other hand, such innovation is possible when taxonomists cooperate on the scale proposed for the FNA Project. Flora North America will be such a massive undertaking that those responsible for its success cannot afford to begin in the same old way, especially given the advanced and rapidly advancing state of computer science and technology. The opportunity must be seized to analyze the whole floristic-taxonomic research process from the point of view of EDP and then to develop the means for automating as many of the routine operations as possible.

A major effort will be made, therefore, to develop a package of EDP programs ("flora software") that, in cookbook fashion, can be used by any FNA or other taxonomist as a sort of "black box" to perform automatically many of the purely routine chores involved in compiling, correlating, and editing data for a flora or taxonomic monograph. This development should not only reduce the investigator's errors and increase his efficiency greatly, but also provide the Editorial Committee of FNA with a powerful means of quality control and standardization, to say nothing of the enormous amount of routine work it should spare the committee. Some of the programs should enable the user to construct "taxa X character" matrices and then to construct identification keys, descriptions, maps, etc., from these matrices completely by machine. Existing numerical taxonomic programs can then be linked to the flora-writing programs so as to add

Digitized by the Hunt Institute for Botanical Documentation

decision-making aids to the proposed "kit of tools" for the taxonomist and floristician.

In brief, a major thrust of the FNA Project will be to reduce the drudgery of floristic and monographic work to the bare minimum by applying machines to remove the human element wherever possible. This should result in reducing the taxonomist's work to its basic intellectual content, making him more scholar and less curator.

Users

The immediate users of FNA will be the authors, editors, and advisers of the project who produce the flora and information system and any other plant taxonomists who are concerned in some way with North American plants. Eventually, this means all plant taxonomists, and this primary community of users is represented on the international level by the International Association for Plant Taxonomy, on the national level in the United States and Canada by the American Society of Plant Taxonomists and the systematic sections of the Botanical Society of America and the Canadian Botanical Association, and on the regional and state or province level by many smaller professional societies.

The secondary community of users includes the medical, biological, and agricultural scientists at large, but especially the ecologists, who, in organizing the International Biological Program (IBP), are generating insatiable taxonomic demands.

The ultimate users are the members of society who must depend in some manner on the existence of accurate, up-to-date knowledge about the kinds of plants in the world. This fact is becoming more obvious as society is becoming more conscious of its natural environment and the need for ecological insight when using natural resources.

Institutional Responsibility

The American Institute of Biological Sciences is an independent, nonprofit organization representing and serving nearly all branches of the medical, biological, and agricultural sciences. It coordinates or administers a broad range of activities aimed at maintaining and improving communication among scientists in these fields. The AIBS represents, therefore, the ideal home for the FNA Project, particularly in view of the possibility that the proposed FNA information system will be adaptable directly to areas of biology other than floristics and taxonomy. Any general biological center or system logically should be developed within such a broadly based organization as the AIBS.

The Smithsonian Institution's offer to serve as the working headquarters for the Secretariat at least during the initial planning and development was accepted without hesitation by the AIBS, so that the FNA Project would have

Digitized by the Hunt Institute for Botanical Documentation

the necessary facilities and research environment to begin at once. The Smithsonian was a natural choice, because it maintains the U.S. National Herbarium, one of the three largest in the New World, and has been active in floristic and monographic work since the time of Asa Gray. The recent establishment of an Information Systems Division with the responsibility to develop systems for handling collection-based information has further enhanced the environment at the Smithsonian for FNA. Both the AIBS and the Smithsonian enjoy excellent international relations among biologists, a fact that bodes well for FNA, as it cannot realize its full potential without a large measure of international participation. In fact, the Smithsonian has been deeply committed to international activities in systematics from its inception, and in this tradition it plays host at present to some part of such other international botanical projects as Flora Neotropica and the Index Nominum Genericorum.

It must be stressed, however, that FNA is not a Smithsonian project, despite the Smithsonian's keen interest and early, "pump-priming" investment. FNA is the responsibility of all cooperating institutions, and the Smithsonian is only the major off-campus site at present for this AIBS effort. The Smithsonian does not have the financial resources to conduct the FNA Project on the scale proposed by the AIBS-FNA committees, nor is it prepared to absorb all of the indirect costs of operating the Secretariat and these committees. Accordingly, equipment and related costs have been budgeted in this proposal, and the equipment will become the property of the AIBS. The AIBS wishes to maintain the project's functional integrity and independence so that the working headquarters can always be moved or decentralized if circumstances demand.

Stages of Development

The FNA Project is proceeding in several stages and substages, and the following are the major stages:

1. Feasibility Stage, 1965-66.--This stage, which began with the appointment of an ad hoc ASPT Feasibility Committee in August 1965, ended when the ASPT Council accepted the report of the Feasibility Committee in August 1966. The report concluded that a project of the scope and objectives of FNA is both feasible and desirable at this time (Shetler 1966a,b).

2. Organizational Stage, 1966-67.--Following the acceptance of the Feasibility Committee's report, the ASPT Council immediately appointed an ad hoc ASPT Organization Committee, which established the committee structure for FNA and ended its work in January 1967 (Mathias 1967). This stage ended in August 1967 when the AIBS officially adopted the FNA Project, including the committee structure set up by the ASPT Organization Committee.

3. Planning Stage, 1968-71.--Preliminary planning began in late January 1967 when the newly appointed Editorial Committee held its first

Digitized by the Hunt Institute for Botanical Documentation

meeting. Since then three other meetings have been held (Texas, Missouri, Ontario), and the Steering Committee has held two meetings (Texas, Ohio). Planning to date has been concerned largely with refining the botanical concept and policy of the FNA Project (program definition) and only to a limited degree with the concept of an information system. The Secretariat has been organized on a rudimentary scale and the Editorial Committee has met, thanks mainly to funds from the Smithsonian Institution totalling about $50,000 over three fiscal years. The most recent Editorial Committee meeting was financed by the National Museum of Canada in Ottawa.

Formal planning cannot begin until sufficient funds become available to permit regular functioning of the committees and the Secretariat, and to permit the appointment of a person who can develop the concept and plan for the information system. The present proposal is directed toward the planning requirements for the information system, while the main FNA proposal, submitted in December 1967 (see "Overview"), is directed toward the botanical planning requirements. It is expected that about two and one-half years will be needed to complete the planning stage, although the FNA committees certainly will implement the program sooner if possible. In fact, no sharp distinction can be drawn between planning and implementation insofar as the botanical aspects are concerned, because the many background materials that must be prepared during the planning stage represent also the first steps on the road to implementation.

5. Design and Implementation Stage, 1971-74.--When the planning is completed the committees will begin at once with the tasks of preparing Flora North America. By then staff should be available both for writing and editing taxonomic manuscript, and for designing and implementing the various elements of the planned information system. It is assumed that the necessary equipment and other resources will be available as needed at this stage to implement the information system. An estimated two and one-half years will be required for this stage, which means that by January 1974 at the very latest--five years from the start of formal planning--the program should be fully implemented, though not necessarily in its most advanced form.

6. Operational Stage, 1974- .--All aspects of the FNA Project should be fully operational by 1974, and it is expected that about 10 years will be required thereafter to complete Flora North America. However, the FNA information system or its outgrowth and various other spin-offs from the project should continue indefinitely past the estimated target date of 1984.

PROPOSED PLANNING EFFORT

Objectives

(1) To develop the concept for the FNA information system, based on the latest EDP technology.

Digitized by the Hunt Institute for Botanical Documentation

(2) To produce the program plan for designing and implementing the system conceived.

Although the proposed effort will be part of the third stage in the development of the overall FNA Project (see above), with respect to the development of an information system it will constitute only the first or program definition stage.

Resultant Program Plan

The program plan developed should define and specify the requirements and costs for: design engineering; purchasing or renting and installing facilities, hardware, and software; technical and managerial staffing; testing and evaluating; system management and operation; and user communication--to name the most obvious elements. The completed plan should include a PERT analysis. Attention must be paid to defining as exactly as possible the immediate and larger user communities and to specifying the indexing requirements and tools, especially from the point of view of the botanist who must provide the input for any information system.

The design and implementation stage will doubtless have a short-range and a long-range phase, and the program plan should include subplans for these two phases. The long-range needs of FNA probably will call for the development of some type of information center or network. Meanwhile, the FNA workers will need interim use of EDP to carry on some of their background compiling and editing while the full system is being designed and built.

Work Plan

Approach

The AIBS does not have its own information systems expert(s) who could conduct the necessary program definition for the FNA Project. Therefore, the AIBS proposes to employ a high-level Systems Development Manager (SDM) at the earliest possible time after the grant is awarded, and to provide him with all necessary resources to conduct the program definition outlined above.

To aid the existing FNA committees in defining and recruiting the SDM, the AIBS further proposes to impanel an ad hoc SDM Task Force who will meet as often as necessary following the award of the grant to fulfill this responsibility. Initially, the Task Force will need to meet with FNA botanists to refine its own concept of the FNA Project, looking at the background, progress, and overall program goals. Once the Task Force has gained a thorough understanding of the FNA Project from a botanical standpoint, then it will be able to determine the needs for, and

Digitized by the Hunt Institute for Botanical Documentation

scope of, a supporting information system, giving attention first to the project's immediate EDP needs and second to the long-range needs.

After the Task Force has developed the criteria and responsibilities of an SDM it can proceed quickly to recruit such a man. To recruit, the Task Force may choose to work through a recruitment consultant or agency so as to obtain the best man possible in the shortest time possible, and the AIBS proposes to have the necessary resources available to permit this option.

When the SDM is appointed the Task Force is expected to play an important role in introducing him to the information problems of the FNA Project and then in providing continuing help and advice as needed by the SDM for the duration of the planning stage. Additional members may be added to the Task Force as time goes on if this is deemed necessary to provide a broader consultative base.

The ad hoc SDM Task Force will be directly responsible to the AIBS Executive Committee (see Fig. 1) and will be expected to provide monitoring advice as needed while the planning is underway. It must be stressed, however, that this committee's role will be purely advisory in all respects, and the actual selection of the SDM from recruited candidates will be made by the AIBS Director in consultation with the FNA botanists. Furthermore, the SDM will be given full responsibility and authority to develop the plan for the FNA information system as he sees fit, though the decision to implement it will rest with the AIBS.

The AIBS will expect to submit a new proposal to the NSF for designing and implementing the information system as soon as the SDM has prepared an acceptable definition and plan, but no later than two and one-half years after the proposed grant is awarded.

System Considerations

The Systems Development Manager must consider all aspects of data input, processing and storage, and output with respect to the botanical goals of the FNA Project, giving special attention from the very start of his planning to the nature, form, and means of output.

Ultimately, there are three sources of information from which FNA must be produced: (1) the literature, (2) the botanical collections, and (3) new, unpublished research. The plan for implementing an EDP-based information system must take into account not only the rapidly accumulating new information, what we may call the "forelog," but also the massive backlog of data files. Today taxonomists are still compiling their most important research indices by hand, e.g., the Gray Card Index, Index Kewensis, Index to Plant Chromosome Numbers. Some of these efforts are endangered at present for lack of funds or of dedicated specialists who are willing to devote their time and energy to the tedious, manual compiling.

Digitized by the Hunt Institute for Botanical Documentation

This is a propitious moment, therefore, to be examining the information problems of plant taxonomy, and first attention must be given to the question of incorporating these and similar, existing data files into an overall information system.

In the latter connection it should be noted that the International Bureau for Plant Taxonomy and Nomenclature (Utrecht), currently exploring the feasibility of computerizing some of its operations in compiling and documenting taxonomic data (e.g., Index Nominum Genericorum), has expressed much interest in this FNA proposal and its objectives and hopes to present some of the Bureau's problems to the SDM when employed.

The problem of automating data in botanical collections (herbaria) is an immense one that needs thorough analysis. Several institutions have already made some progress on this matter (National Museum of Canada, U.S. National Herbarium, Herbarium of National University of Mexico), and their efforts must be studied by the SDM. There is an acute need for an economically feasible system for automatically retrieving geographic data from herbaria and plotting distribution maps. Today there are well over 125 million herbarium specimens in the world (Index Herbariorum, 1964), so that the physical problem of handling specimens in order to retrieve any data is staggering.

The imminent practical tasks of the AIBS-FNA Editorial Committee concern the preparation of the concise diagnostic manual of North American vascular plants, because this effort will involve editing and revising thousands of pages of manuscripts submitted by taxonomists at large. The SDM should give his most urgent attention to the latest EDP systems for manuscript editing, revising, and storing, so that the work of the Editorial Committee will not be stymied indefinitely while a perfect information system is being conceived, designed, and implemented. Closely related here is the problem of handling dichotomous identification keys by computer and the need, discussed earlier, for developing data-matrix methods for automating identification and the whole key- and description-writing process. These aspects should receive early attention and be incorporated into the short-range subplan proposed (see "Approach"). Some progress has already been made on automating identification and key-writing (e.g., Rypka et al. 1967, Boughey et al. 1968, Morse et al. 1968, Morse 1968).

Some of the kinds of output that might be needed are the following: floras, distribution maps, term glossaries, data file indices, gazeteers, taxonomic descriptions, ecological summaries, statistical calculations, phenetic diagrams, research reports. Sophisticated search and retrieval programs for answering specific inquiries or for performing simultaneous multi-character identifications should be coupled with indexing tools that will enable the FNA botanists to handle some of the information requests with periodically updated dictionaries and gazeteers. The taxon-specific data matrices should be expandable to include descriptive, biochemical, pharmacological, and other data as the users demand.

Digitized by the Hunt Institute for Botanical Documentation

Finally, no small part of the SDM's responsibility will concern the means of communication. Special attention must be given to man-machine interactive systems, using remote terminals to access the FNA data bank and operating programs. Remote access to the data bank should be made a long-term goal of the advanced information system development, even if such access is economically prohibitive today for the kind of information and user network anticipated. It is important to take note that certain taxonomic functions (e.g., identification) can never be automated successfully, in terms of the user, until remote, real-time access is routinely possible.

Timetable

Event or Task	Max. No. Months from Grant Award
1. Ad Hoc SDM Task Force impaneled 2. Task Force defines SDM responsibilities and qualifications, meeting 2-4 times 3. Secretary employed for Task Force and SDM	3
4. SDM recruited by Task Force, perhaps with aid of professional recruitment agency or consultant; 2-4 meetings held 5. SDM appointed by the AIBS	6
6. SDM briefed by Task Force and other AIBS-FNA committees in 1-several meetings 7. SDM begins consultation and study 8. Task Force may be enlarged	9
9. SDM presents first progress report to the AIBS through Task Force and botanical committees	12
10. SDM presents second progress report	18
11. SDM presents preliminary report on proposed information system for FNA, giving special attention to short-range EDP needs at this stage	24
12. SDM completes concept and program plan for FNA information system and presents it to FNA botanists and appropriate other bodies of the AIBS	30

Digitized by the Hunt Institute for Botanical Documentation

It is to be understood clearly that this represents the most expanded schedule that will be permitted by the AIBS, and every effort will be made to shorten the spans of time between the several events and tasks.

Personnel

Ad Hoc SDM Task Force

The purpose of this committee has already been explained (see "Approach"), namely, to help the AIBS-FNA botanical committees in selecting the best qualified man for the project's Systems Development Manager.

The Task Force should be relatively small if it is to work effectively. The following four persons have agreed tentatively to serve if appointed by the AIBS, and brief biographies of them are attached to the proposal (Appendix I):

1. Charles T. Meadow, Chief, Systems Research and Development Division, Center for Computer Sciences and Technology, National Bureau of Standards, Washington, D.C. 20234

2. Jack Minker, Scientific Consultant, Washington Facility, Auerbach Corporation, 1501 Wilson Boulevard, Arlington, Virginia 22209, Associate Professor of Computer Science, University of Maryland, College Park, Maryland 20742

3. Norman J. Ream, former (until September 1968) Special Assistant to the Secretary of the Navy for Information Systems and Member of the Secretariat, Department of the Navy, Pentagon, Washington, D.C. 20530

4. Nicholas J. Suszynski, Jr., Director, Information Systems Division, Smithsonian Institution, Washington, D.C. 20560

This Task Force will serve for the duration of the proposed grant period, i.e., until the SDM has presented his program plan for an FNA information system. However, its major task will come during the first 6-9 months. The Task Force should be free to co-opt another member or two if deemed essential. Likewise the AIBS reserves the right to name an additional member or two if necessary to insure the success of the Task Force. After the SDM has been appointed the AIBS will look upon this committee more as an advisory body than a Task Force, and at this time it is likely that as many as 4-6 new members will be added to enrich the advisory capacity of the Task Force.

It is expected that the AIBS will impanel some type of standing advisory committee on information systems once an FNA information system is implemented in the next stage, and it is hoped that several or all of the original members of the ad hoc Task Force will be available then to serve

Digitized by the Hunt Institute for Botanical Documentation

on the standing committee, thereby providing the AIBS with continuity in the development of an FNA system.

The pace of meetings is projected for the SDM Task Force under the "Budget" (see "Travel").

Systems Development Manager (SDM)

The role and responsibilities of the SDM have already been discussed (see "Approach). Broadly speaking, the SDM will be charged with developing the goals, concept, and program plan for designing and implementing the proposed FNA information system; therefore, he must be a person who is highly trained and experienced in managing information systems as well as in electronic data processing, and he must have a record for getting things done. Although the exact qualifications and responsibilities are yet to be determined, as explained above, it is obvious that he must be a high-level person in the executive managerial group.

The SDM will be directly responsible to the Director of the AIBS, as are all members of the AIBS-FNA committees, and he will become an integral member of the FNA Project's working group, namely, the AIBS-FNA Editorial Committee, participating in all of its activities and deliberations insofar as they are relevant to his work. He will be housed physically and administratively in the FNA Secretariat and on a daily basis must work effectively with and through the Secretary, who has the responsibility for the overall coordination of the FNA Project (see Fig. 1). He will have a secretary directly available to him (see below), and it is assumed that when the next stage comes he will be asked to manage the designing and implementation of the FNA information system, having all necessary staff available and reporting to him.

A two-year appointment, beginning about six months after the grant is awarded or as soon as the SDM can be recruited, is proposed because a properly qualified individual is not likely to accept such an assignment on a shorter-term basis. Furthermore, the complexity of the assignment may demand this much planning time. Nevertheless, as previously indicated, the SDM will be given every inducement to complete a concept and program plan sooner, preferably within a year after appointment, so that resources can be sought earlier to proceed with the designing and implementation of the proposed system. Possibly certain modules or subsystems of the total system will be ready for designing and implementation before others. Moreover, the possibility is not precluded that the total system will be developed, approved, and implemented in stages rather than as a whole.

Secretary

The SDM will need a secretary who in addition to being capable of performing the usual clerical duties should be qualified for a more responsible role, including some research assistance. Inasmuch as the

Digitized by the Hunt Institute for Botanical Documentation

SDM will work in the Secretariat, his secretary should work under the general supervision and coordination of the administrative secretary (or assistant) who is to be employed for the FNA Project with funds requested in the main proposal (see pp. 64, 83 in that proposal). Therefore, in the interest of efficiency and economy at this stage in the project's development, it is proposed that when her services are not required by the SDM she perform whatever duties may be assigned in connection with the general activities of the FNA Project. With this approach in mind it is further proposed that the SDM's secretary be hired shortly after the grant is awarded, but no later than three months after (see "Timetable"), so that she can assist the Secretariat in administering the work of the SDM Task Force and be, in effect, secretary to the Task Force.

Facilities

The SDM and his secretary will be provided office space at the Smithsonian Institution where the rest of the Secretariat is located at present (see also "Institutional Responsibility").

Digitized by the Hunt Institute for Botanical Documentation

Budget Justification

The "Budget Summary" follows this section (p. 20). The justification and summary are presented according to the expense categories used by the AIBS.

A. Salaries

Salary is requested to employ a Systems Development Manager for two years, beginning about six months after award of the grant. Salary for the SDM's secretary is requested for the full two-and-one-half-year grant period for reasons explained above (see "Secretary"). Annual increases are included for the secretary, but not for the SDM because of the executive scale of his salary.

B. Permanent Equipment

Certain pieces of equipment will be needed, as indicated in the budget summary, to support the SDM's work. This equipment will become the property of the AIBS even though it is used at the Secretariat (see "Institutional Responsibility"). All other office equipment needed for the SDM and his secretary will be provided by the Smithsonian Institution as long as the Secretariat remains here.

C. Expendable Office Supplies

This category covers all ordinary office supplies, including expendable equipment (e.g., waste baskets) and the printing of necessary stationery and office forms, which will be needed to support the work of the SDM. Also included here are essential dictionaries and other specialized reference books and subscriptions to technical journals.

D. Reproduction

The SDM will need to circulate working papers and progress reports to many individuals, and it is expected that a large number of copies of the final report, detailing his system concept and design-implementation plan, will be needed. It is estimated, therefore, that he will be distributing the equivalent of 100 pages/year to 100 persons (10,000 pages/year) during his tenure, and the cost of this reproduction is estimated at $0.10/page. The AIBS proposes to use the money for renting a Xerox machine or similar copier if that proves cheaper than paying by the page.

The SDM likely will need modest photographic services during his study for illustrating reports and technical talks that he may present, and an estimate of these expenses is included in this category.

Digitized by the Hunt Institute for Botanical Documentation

E. Communications

Self-explanatory in "Budget Summary."

F. Printing and Production

Funds are requested in this category for the preparation of charts and graphs by a professional illustrator. These will be needed to illustrate the SDM's written and oral reports.

G. Travel

Travel is a major item in the budget and is absolutely essential to the success of the proposed effort because of the cooperative, international character of the FNA Project. Without adequate travel funds the proposed consultative-advisory mechanism (see "Approach") will not succeed, and the SDM will be unable to carry out his assignment.

SDM Task Force.--As presently constituted, the Task Force comprises only four persons who all live in the Washington, D.C., area, so that transportation costs should not average more than about $5/man/day (i.e., per man-day) for every day that the Task Force convenes. Subsistence expenses may average another $5/man-day. During the full period of the proposed grant the members of the Task Force are expected to spend about 20 days each on FNA business, of which about 15 days each will be spent in committee. Thus, expenses will be required for about 60 man-days (4 men X 15 days).

As explained earlier (p. 12), the initial Task Force may wish to co-opt an additional member, and AIBS is likely to want an expanded representation on this committee after the SDM has been recruited. Such additional members are likely to be selected from outside of the Washington area. It is proposed, therefore, that funds be available to support the travel of one out-of-town member to all of the substantive meetings of the Task Force, estimated to be about 10, and to support the travel of three more out-of-town members to the last 4 of the 10 meetings during the grant period. This is a total of 22 trips (1 man X 10 trips + 3 men X 4 trips). Subsistence is computed at the rate of 3 days/man/trip, assuming that a day or more will be consumed in travel for each trip, depending on point of origin, and another 1-2 days in meeting.

Recruitment of SDM.--Travel funds will be needed to recruit the SDM. It may be desirable to interview as many as four out-of-town candidates before a choice is made. Subsistence is computed on the basis of 2 days/interview.

Systems Development Manager.--The SDM must be able to travel frequently to study existing information systems, to consult with colleagues and advisers concerning his assignment, and to attend technical meetings. He will be expected to attend all meetings of the AIBS-FNA Editorial and

Digitized by the Hunt Institute for Botanical Documentation

Steering Committees, which combined are expected to average four a year (see main proposal) and which will be held at different places in the United States and Canada. It is estimated, therefore, that he will need to take about 10 trips/year in the conduct of business, averaging 3 days/trip. This will be domestic travel primarily in the United States, but including trips to Canada and possibly Mexico.

One 10-day European trip will be required early in the SDM's tenure so that he can acquaint himself with the work of the Royal Botanic Gardens at Kew (near London) on Index Kewensis, the botanical data processing activities at the Monks' Wood Experimental Station near London, and the documentational work of the International Bureau for Plant Taxonomy and Nomenclature at Utrecht, the Netherlands. If possible, this trip should be carried out in September 1969 to enable the SDM to attend the conference planned for that date on "Advances in Methods of Data Handling in Biology and Geology," to be held in Cambridge, England.

Other AIBS-FNA Personnel.--Support for the botanical committees and personnel of the FNA Project is being sought in the main grant proposal, as stated in the "Overview," but certain safeguards must be built into the present proposal to insure that the planning of the information system can begin at once if this grant is awarded. These safeguards include providing the means (1) for the AIBS-FNA Editorial Committee to convene with the SDM Task Force immediately to brief the Task Force on the botanical goals of the project before it sets out to define and recruit an SDM, and (2) for bringing the Program Director, Dr. Raven, who resides in Stanford, California, to all substantive meetings (ca. 10) of the SDM Task Force. For the 7 members of the Editorial Committee (including Chairman of Steering Committee) to attend a meeting in Washington, D. C., experience has proven that the cost averages about $150/member. For the Program Director, transportation to Washington costs about $300/trip.

The work of the SDM Task Force and the SDM will require that the FNA Secretary and/or the Director of the AIBS or his representative travel on their behalf or with the SDM from time to time. Therefore, it is estimated that about one such trip per quarter, or 10 trips, will need to be made during the 30-month grant period.

Consultants for SDM.--The SDM must be able to obtain all necessary advice while he is carrying out his assignment; therefore, he must have the resources to bring a consultant to Washington, D.C., from anywhere in North America and, if vital, from abroad. It is estimated that he may need to bring in North American (U.S., Canada, Mexico) consultants an average of 5 times each year. In addition, he may need to bring in a European consultant as many as 3 times during the two years.

H. Honoraria

SDM Task Force.--The members of this ad hoc committee will be expected to devote up to 20 days of their time to FNA business during the grant period and to perform significant tasks on schedule. It may be appropriate if not necessary, therefore, to pay honoraria, the amounts varying with the

Digitized by the Hunt Institute for Botanical Documentation

circumstances. The AIBS is zealous to maintain a balance on the Task Force between experts who are in and out of government, realizing that the commercial cost of retaining the services of members from industry will be relatively high. The proposed number of meetings and plan for enlarging the Task Force after the SDM is selected are explained under "Travel." Based on those calculations, it is estimated that funds will be needed to pay honoraria approximately as follows: 20 days at $200/day; 10 days at $150/day; 25 days at $100/day; and 45 days at $50/day. This is 100 days at an average of about $102/day, which has been rounded to $100/day.

Consultants for SDM.--As indicated under "Travel," the SDM must be able to retain consultants in biology or information science and technology on a daily basis when needed. These may be drawn from government, industry, or the academic world. It is estimated, therefore, that he will need to pay consulting fees for at least 12 days/year.

I. Miscellaneous Other Costs

Page Charges for Publication.--The work of the SDM and the Secretary or other members of the AIBS-FNA Editorial Committee concerning an FNA information system is certain to result in one or more publications each year. In particular, it is likely that some form of the final report of the SDM will be published. Therefore, it is estimated that 25 pages will be published during the first 18 months and another 50 will be forthcoming during the last 12 months.

Reprint Costs.--The Secretariat will have the responsibility to see that published information on the work of the SDM and related personnel is widely disseminated. Accordingly, a modest sum is being requested to purchase reprints of the papers that are expected to be published.

Recruitment Expenses.--The AIBS does not maintain its own personnel file and service for the type of recruitment that will be required to secure a competent SDM. The AIBS proposes, therefore, to retain a recruitment consultant or agency to perform the necessary searching and screening if this appears to be the most effective and economical way to proceed after the SDM Task Force has defined the position. The usual service charge for conducting an executive search is about 25% of the first year's salary, plus expenses. The expenses of bringing recruits to Washington, D.C., for an interview have been included under "Travel." Recruitment of the SDM probably represents the most important single task that must be performed during the proposed effort, and adequate resources must be available to allow some flexibility in carrying out this task. If the SDM is hastily and poorly chosen, the whole effort is likely to fail.

Relocation of SDM.--The AIBS proposes to cover the full cost of relocating the SDM and his family. Inasmuch as the candidate is as likely to originate from the West Coast as from the East Coast of the United States, a fairly substantial relocation fund is requested.

Digitized by the Hunt Institute for Botanical Documentation

Cost of SDM Task Force and SDM.--The ad hoc Task Force and the SDM, after he is hired, are certain to incur unforeseen expenses in the conduct of their assignments. Given the expense to which the AIBS proposes to go in conducting this planning stage and the high salary it proposes to pay the SDM, it seems reasonable to take every precaution to avoid a situation in which the Task Force or SDM cound not function properly for lack of foresight on the part of the AIBS in preparing this grant proposal. Accordingly, a modest contingency fund is requested under this heading, to be used at the discretion of the AIBS to facilitate the work of the Task Force and SDM when funds are not otherwise budgeted.

Digitized by the Hunt Institute for Botanical Documentation

BUDGET SUMMARY
(30 months)

	First 18 Months	Last 12 Months	Total
A. SALARIES			
Senior Personnel			
Systems Development Manager (SDM)	$30,000	$30,000	$60,000
Other Personnel			
Secretary (starting at $7,000/year)	10,750	7,750	18,500
SUBTOTAL	$40,750	$37,750	$78,500
Employee Benefits (15% of gross salary)	6,113	5,662	11,775
TOTAL A	$46,863	$43,412	$90,275
B. PERMANENT EQUIPMENT			
Typewriter, electric (special keyboard)	$600		$600
Dictating-transcribing machine & accessories	400		400
Desk, executive	400		400
Chair, executive	100		100
Cabinet, filing	150		150
Bookcase, metal, 2 units	150		150
TOTAL B	$1,800		$1,800
C. EXPENDABLE OFFICE SUPPLIES	$1,500	$1,500	$3,000
D. REPRODUCTION			
Duplication services	$1,000	$1,000	$2,000
Photographic services	250	250	500
TOTAL D	$1,250	$1,250	$2,500

Digitized by the Hunt Institute for Botanical Documentation

BUDGET SUMMARY--cont.
(30 months)

	First 18 Months	Last 12 Months	Total
E. COMMUNICATIONS			
Postage, express, freight	$350	$250	$600
Telephone and telegraph ($75/month)	1,350	900	2,250
TOTAL E	$1,700	$1,150	$2,850
F. PRINTING AND PRODUCTION			
Artwork (15 charts and graphs at $30/illustration)	$150	$300	$450
G. TRAVEL			
SDM Task Force			
Washington, D.C., members (local expenses of convening, 48 + 12 days at $10/day)	$480	$120	$600
Out-of-town members			
Transportation (12 + 10 trips at $150/trip)	1,800	1,500	3,300
Subsistence (36 + 30 days at $20/day)	720	600	1,320
Recruitment of SDM (4 interviews)			
Transportation ($200/interview)	800		800
Subsistence (2 days/interview at $20/day)	160		160
Systems Development Manager			
Domestic travel (U.S., Canada, Mexico)			
Transportation (10 + 10 trips at $150/trip)	1,500	1,500	3,000
Subsistence (3 days/trip at $20/day)	600	600	1,200
Foreign travel			
Transportation (1 trip to London & Amsterdam)	600		600
Subsistence (10 days at $25/day)	250		250
Other AIBS-FNA Personnel			
Transportation (12 + 5 trips at $150/trip & 6 + 4 trips at $300/trip)	3,600	1,950	5,550
Subsistence (54 + 27 days at $20/day)	1,080	540	1,620

Digitized by the Hunt Institute for Botanical Documentation

BUDGET SUMMARY--cont.
(30 months)

	First 18 Months	Last 12 Months	Total
G. TRAVEL--cont.			
Consultants for SDM			
North American			
Transportation (5 + 5 visits at $150/visit)	$750	$750	$1,500
Subsistence (2 days/visit at $20/day)	200	200	400
European			
Transportation (2 + 1 visits at $600/visit)	1,200	600	1,800
Subsistence (6 days/visit at $20/day)	240	120	360
TOTAL G	$13,980	$8,480	$22,460
H. HONORARIA			
SDM Task Force (60 + 40 days at $100/day, see justification)	$6,000	$4,000	$10,000
Consultants for SDM (12 + 12 days at $150/day)	1,800	1,800	3,600
TOTAL H	$7,800	$5,800	$13,600
I. MISCELLANEOUS OTHER COSTS			
Page charges for publication (25 + 50 pages at $25/page)	$625	$1,250	$1,875
Reprint charges	150	300	450
Recruitment expenses	9,000		9,000
Relocation of SDM	3,000		3,000
Cost of SDM Task Force and SDM	1,000	1,000	2,000
TOTAL I	$13,775	$2,550	$16,325
TOTAL DIRECT COSTS	$88,818	$64,442	$153,260
INDIRECT COSTS (22.85% on everything but "Permanent Equipment")*	19,884	14,725	34,609
TOTAL COSTS	$108,702	$79,167	$187,869

*Indirect cost rate as established by the Indirect Cost (Rate) Determination Office of the NSF, presented to the AIBS in a letter dated 19 September 1968. This rate applies to all budget items but equipment.

Digitized by the Hunt Institute for Botanical Documentation

REFERENCES

1. American Institute of Biological Sciences. 1967. Flora North America Project. Proposal to the National Science Foundation, November 1967. 91 p. + Appendices I-V. (Submitted to the NSF in December 1967 by John R. Olive, Director of the AIBS.)

2. Boughey, A. S., K. W. Bridges, and A. G. Ikeda. 1968. An automated biological identification key. Univ. Calif., Irvine, Museum of Systematic Biology, Research Series No. 2, 36 + xix p.

3. Lanjouw, J., and F. A. Stafleu (compilers). 1964. Index Herbariorum. Part I: The herbaria of the world. Ed. 5. Regnum Vegetabile 31, 251 p. (Utrecht)

4. Mathias, Mildred E. (chairman). Organization Committee. 1967. Final report of Organization Committee, Flora North America. 4 p. Duplicated. (Available from FNA Secretary, Smithsonian Institution.)

5. Morse, L. E. 1968. Construction of identification keys by computer. Amer. J. Bot. 55(6, pt. 2): 737. (Abstract)

6. Morse, Larry E., John H. Beaman, and Stanwyn G. Shetler. 1968. A computer system for editing diagnostic keys for Flora North America. Taxon 17(5): In press.

7. Rypka, E. W., W. E. Clapper, I. G. Bowen, and R. Babb. 1967. A model for the identification of bacteria. J. Gen. Microbiol. 46: 407-424.

8. Shetler, Stanwyn G. (secretary). Feasibility Committee. 1966a. Report of the Flora of North America Committee. Smithsonian Institution, Washington, D.C. 25 p. Duplicated.

9. _____. 1966b. Meeting of the Flora of North America Committee. Taxon 15: 255-257.

Digitized by the Hunt Institute for Botanical Documentation

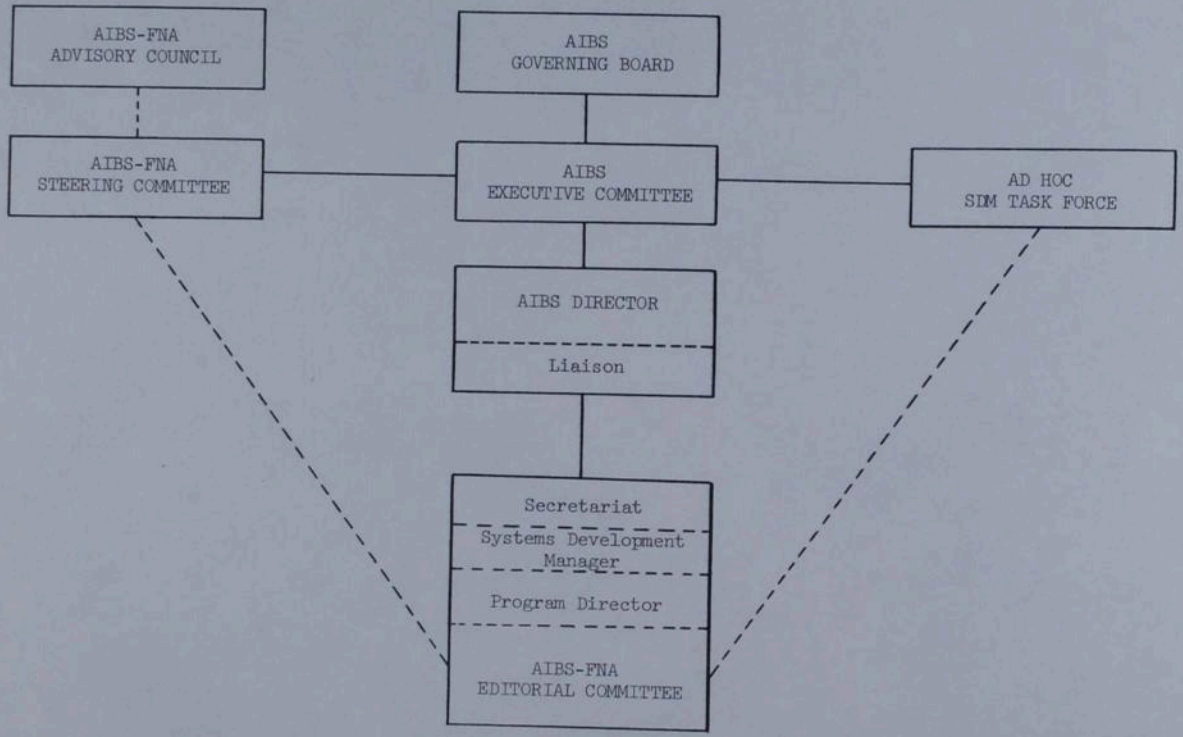


Fig. 1.--Organizational Chart of AIBS-FNA Project. Solid lines: formal lines of communication and responsibility; dashed lines: informal lines of communication and responsibility.

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX I

Brief Biographies of
SDM Task Force Members

1. Charles T. Meadow

Chief, Systems Research and Development Division, Center for Computer Sciences and Technology, National Bureau of Standards, 1968-

Born 16 December 1929 at Paterson, N.J. A.B., Univ. Rochester, 1951, M.S., Rutgers University, 1954, both in mathematics. Mathematician, David Taylor Model Basin, U.S. Dept. Navy, 1954-55; assistant mathematician, RAND Corp., 1955-56; unit manager, General Electric Co., 1956-59, consulting analyst, 1959-60; development mathematician, IBM, 1960-62, senior programmer, 1962-68. Lecturer, Arizona State Univ. 1957, Adjunct Lecturer, Graduate School of Library and Information Services, Univ. of Maryland, 1968- . Member of Assoc. Computing Machinery, Operations Research Society. Author of various technical papers, including textbook: "The Analysis of Information Systems," John Wiley & Sons, New York. 1967. 301 p.

2. Jack Minker

Scientific Consultant, Washington Facility, Auerbach Corporation, and Associate Professor of Computer Science, University of Maryland

Born 4 July 1927 at Brooklyn, N.Y. B.A., Brooklyn College, 1949; M.A., Univ. Wisconsin, 1950; Ph.D., Univ. Pennsylvania, 1959; all in mathematics. Dynamics engineer, Bell Aircraft Corp., 1951-52; engineer, RCA, 1952-57, Manager, Information Technology Section, Data Systems Center, 1957-61; Manager, Washington Systems Section, Auerbach Corp., 1961- . Invited lecturer, Gordon Res. Conf. Info. Storage & Retrieval, 1960; Co-chm. National Conf. Info. Storage & Retrieval, Assoc. Comput. Mach., 1960. Member of Assoc. Computing Machinery, Institute of Mathematical Statistics, Amer. Mathematical Society, and Amer. Documentation Institute. Author and co-author of many technical articles on simulation, information storage and retrieval, and mathematics, including the recent article: "File Organization and Data Management," Annual Rev. Information Sci. Tech 2: 123-160. 1967. (With Jerome Sable.)

3. Norman J. Ream

Special Assistant to the Secretary of the Navy for Information Systems and Member of the Secretariat, Department of the Navy, 1966-68

Born at Aurora, Illinois. B.S., Univ. Illinois, 1934, in accounting. Member Comptroller's staff, Pure Oil Co., 1934-42; staff of Touche, Niven and Co., Public Accountants, then of George Fry & Associates, Management Consultants, 1942-47; Director of Accounting Research, IBM, 1947-50; Assistant to Treasurer, Lever Brothers Co., 1950-53; Financial Planning Advisor, then Corporate Director of Systems Planning, Lockheed

Digitized by the Hunt Institute for Botanical Documentation

3. Norman J. Ream--cont.

Aircraft Corp., Burbank, Calif., 1953-65; first Director, Center for Computer Sciences and Technology, National Bureau of Standards, 1965-66. Member United States Delegation to United Nations Conference on the Application of Sciences and Technology for the Benefit of the Less Developed Areas, Geneva, 1963; Life Member, Senior Planning Council of American Management Association; member, Executive Committee, Committee on Data Systems Languages (CODASYL); member, Executive Advisory Council of Florida Atlantic Univ.; member, Amer. Institute of Certified Public Accountants, Amer. Accounting Assoc., Institute of Electrical & Electronics Engineers, Calif. Soc. Cert. Pub. Account., and Institute of Management Sciences. Author of several technical articles.

4. Nicholas J. Suszynski, Jr.

Director, Information Systems Division, 1965- , and Research Associate of Museum of Natural History, 1968- , Smithsonian Institution

Born 29 March 1929 at Lomza, Poland; U.S. citizen. B.S.I.E., Syracuse Univ., 1956. Programmer, Sylvania Electric, 1956-57; Computer Systems Analyst-Mathematician, Sperry Rand Univac, 1957-58; Programming Supervisor, Minnesota Highway Dept., 1958-60; Manager Analysis and Programming, General Electric Co., 1960-65. Member, Assoc. Computing Machinery (organized N.Y. Southern Tier Chapter); member, Executive Board, and Program Chairman, Special Interest Group for Business Data Processing, ACM; Finance Co-chairman 1964 Spring Joint Computer Conference, Amer. Federation of Information Processing Societies. Author of several papers on simulation, teleprocessing, and computer installations.

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX II

List of Members of AIBS-FNA
Steering and Editorial Committees

AIBS-FNA Steering Committee

A. W. F. Banfield, Director, Museum of Natural History, National Museum of Canada.
Allan Chan, Director, Plant Research Institute, Canada Department of Agriculture.
Lincoln Constance, Director of the Herbarium and former Vice-Chancellor for Academic Affairs, University of California, Berkeley.
David M. Gates (Chairman), Director, Missouri Botanical Garden.
Rogers McVaugh, Professor of Botany and Curator of Vascular Plants, University of Michigan.
Albert E. Radford, Professor of Botany, University of North Carolina, Chapel Hill.
Peter H. Raven (ex officio - see Editorial Committee)
Reed C. Rollins, Asa Gray Professor of Systematic Botany, Director of Gray Herbarium, and Chairman of Institute of Plant Sciences, Harvard University.
Stanwyn G. Shetler (ex officio - see Editorial Committee)
William C. Steere, Professor of Botany, Columbia University, and Director, New York Botanical Garden.
Robert F. Thorne (ex officio) (Chairman of Advisory Council), Professor of Botany at Claremont Graduate School and University Center and Taxonomist at Rancho Santa Ana Botanic Garden.

AIBS-FNA Editorial Committee

John H. Beaman, Associate Professor of Botany and Curator of Beal-Darlington Herbarium, Michigan State University.
Robert Kral, Associate Professor of Biology, Vanderbilt University.
Walter H. Lewis, Director of the Herbarium at Missouri Botanical Garden, Associate Professor at Washington University, and Senior Fellow, Center for the Biology of Natural Systems, Washington University.
John T. Mickel, Associate Professor of Botany, Iowa State University.
Peter H. Raven (Chairman and Program Director), Associate Professor, Stanford Univ.
Stanwyn G. Shetler (Secretary of Editorial and Steering Committees), Associate Curator of Phanerogams, Smithsonian Institution.
Roy L. Taylor, Chief, Taxonomy Section, Plant Research Institute, Canada Department of Agriculture.
John H. Thomas, Curator of the Dudley Herbarium and Lecturer in Biological Sciences, Stanford University.

Digitized by the Hunt Institute for Botanical Documentation

DEC 10 1968

UNIVERSITY OF ILLINOIS · SCHOOL OF LIFE SCIENCES
COLLEGE OF LIBERAL ARTS AND SCIENCES, URBANA

Office of Director, Morrill Hall
Urbana, Illinois 61801
Area Code 217 — 333-3044

December 6, 1968

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

Dear Dave:

Thanks for your letter of November 22 with comments on my manuscript. You are quite right about the need for references to the work of the Taximetrics Laboratory. In rushing to get the information to print, I had not made much of a literature search. I have altered the paper somewhat, and included some of the references which you provided. They are indeed relevant to the things discussed in my manuscript. A copy of the latest version is enclosed. The manuscript was mailed to Systematic Zoology today. As you and George both suggest, the value of my paper is probably more in the discussion and suggestions for procedures than it is for the actual results presented. Nevertheless, I wanted to include results from some actual data in order to provide some experimental support for my suggestions. Again let me thank you for your time and I hope it will be possible for us to talk again before too long.

Sincerely,

Dave

David C. Eades
Assistant Prof. of Zoology

DCE:gs
enc.

Digitized by the Hunt Institute for Botanical Documentation

UNIVERSITY OF ILLINOIS · SCHOOL OF LIFE SCIENCES
COLLEGE OF LIBERAL ARTS AND SCIENCES, URBANA

Office of Director, Morrill Hall
Urbana, Illinois 61801
Area Code 217 — 333-3044

November 18, 1968

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

Dear Dr. Rogers:

It was good to have an opportunity to talk to you again at the Numerical Taxonomy Conference. The conference seemed to me most valuable in focusing attention on several critical areas.

Dr. Reino Kallio, Director of the School of Life Sciences at the University of Illinois, is writing to ask you for an evaluation of my manuscript on "Some procedures for numerical phyletics." I would apprciate it if you would give him your reaction to the scientific merit of the manuscript and of my article on The coefficient of correlation published in Systematic Zoology in 1965. I would also be most interested in any reaction you might have to the manuscript. I plan to submit it to Systematic Zoology in about a week. Enough of my ideas have been mentioned in the Numerical Taxonomy Conference and elsewhere that it is important to me to get these ideas published promptly. The manuscript has been retyped since the copy I left with you. If you have not already read the manuscript, you will find the enclosed copy more accurate and more readable than the earlier version. Thank you for your time.

Sincerely,

Dave

David C. Eades

DCE:gs
enc.

Digitized by the Hunt Institute for Botanical Documentation

22 November 1968

Dr. David C. Eades
School of Life Sciences
University of Illinois
Urbana, Ill. 61801

Dear Dave:

I had already replied to Dr. Kallio when your letter arrived yesterday with the new copy of your manuscript. I said all the nice things necessary, I hope, to make you look like at least a super full professor. I am all in favor of taxonomists getting their just due, particularly in your case.

Concerning the manuscript: I have some reservations about it as well as some pleasant things to say. First of all, it is disheartening to us to see no reference to the work we have done, particularly with respect to the contrasting attitude between us and Sokal with respect to taxonomy. Our papers, particularly the chapter in the book Evolutionary Biology, Vol. I, edited by Th. Dobzhansky, M. K. Hecht and Wm. C. Steere; George's papers in the Journal of Theoretical Biology, Taxon, and BioScience give some of our philosophy and working procedures. I am sending along copies of the above-listed papers. Secondly, I had started to discuss with you while here the business of testing the value of any particular program. I was objecting, as you recall, to taking some conventional taxonomic output and using this as a yardstick for the value of any other procedure. My objection runs something like this. Suppose you did not get 100% agreement between the output of a computer program and the conventional taxonomy - how would you explain the difference? Or even if you did get 100% agreement between your processes and the conventional taxonomist's decisions, what assurance do you have that the next set of data will provide the same result? It seems to me that the most important feature is not to worry about whether the cranks turn gives you some particular result or other but rather that you know precisely what you have done in your own program, starting with well defined rules which you justify in terms of your knowledge as a taxonomist. I would further want to know how you justify the mathematical procedures employed. It is a critical element to know what the model actually says. Personally, I am not convinced that statistics is the basis on which we can build a strong mathematical base for procedures in taxonomy; particularly since the statistical process grew out of classical analytical math which was (and is) not intended to serve the synthetic function required in taxonomy.

Digitized by the Hunt Institute for Botanical Documentation

Thirdly, I am not particularly convinced by your discussion on weighting - that still seems to be, in George's terms, "data diddling."

Now to lay on the pleasant side of the picture. I6 is extremely useful to try to take the sting out of Sokal in relationship to taxonomists of all types and it is particularly nice to see some recognition of the thought processes which have generated over considerable amount of time. I would hate to see Sokal in charge of a large museum, because I feel that he would do disservice to the community by forgetting the significance of "cruddy old collections." I do not believe that Sokal has ever had the need to work with the natural world as expressed in museums, but if he had I suspect that his disdain for the material kept would be considerably less than it presently is. So anything you can do to set to rights the place we work as persons interested in more precise methodology and relationship to taxonomists of all sizes, shapes and dispositions the better we all are. Your paper goes a long way to do this.

The impressions above are mine without the reading by George or any of the others here, and I have given them the manuscript and asked them to give you an independent opinion on it. Perhaps they will say things defferently than I did.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

26 November 1968

Dr. David C. Eades
School of Life Sciences
University of Illinois
Urbana, Illinois 61801

Dear Dave:

I read with interest your recent manuscript "Some Procedures for Numerical Phyletics." I feel that you have succeeded in making some valuable points which need to be better understood by the systematic community. A few specific comments:

¶3. I question the overriding desirability of operationalism, and, for that matter, objectivity. This image of the scientists as a humanoid machine is "Maya" tempting us to confuse precision with correctness. Don't underrate "point 9"!

¶5. These discussions are very good. The concept of characters as real entities I think is misleading however. Characters are mental constructs; they exist in the minds of taxonomists, etc. To speak of "all the characters" is meaningful only in a very formal mathematical sense.

¶16. Discussions of the reasons for redundancy between characters isvery relevant. Redundancy is not summarily bad but in many cases the very thing which gives us "natural" taxonomic structure. Classifying redundancy into cause classes is an important concept which needs more attention. Good and bad redundancy needs to be distinguished.

I was much more impressed with your discussions than with your methods or procedures. It strikes me as weak to make an extant classification primary for your evaluation techniques. Mathematical methods should be recognized as additional tools for the professional systematist, not as alternatives to him. The procedures are not for the purpose of duplicating his work (and need not be diddled to that end) but rather to shed more light on the problem he has set out to study.

In deference to the excellent general discussions in your manuscript, it might better be entitled, "Some discussions of .. etc."

(over)

Digitized by the Hunt Institute for Botanical Documentation

I look forward to seeing you in Dallas.

Very truly yours,

George F. Estabrook

GFE:gm

Digitized by the Hunt Institute for Botanical Documentation

Some procedures for numerical phyletics

David C. Eades
School of Life Sciences
University of Illinois, Urbana, Illinois

Running head: Numerical phyletics

Digitized by the Hunt Institute for Botanical Documentation

Abstract

There is a possible common ground between conventional and numerical taxonomy involving computer analysis of many characters. It would utilize evolutionary concepts whenever possible to modify an otherwise phenetic classification. New algorithms are described for character analysis and for cluster analysis. Many combinations of computer analysis are tested on a set of 20 characters for 146 adult male grasshoppers belonging to 15 species. Results, judged by the effectiveness in showing greater similarity between individuals of the same species, were significantly improved by conclusion of computer procedures for adjusting correlated characters and for weighting characters.

Digitized by the Hunt Institute for Botanical Documentation

The split between conventional and numerical schools of taxonomic thought is too well known to require elaboration here. This polarization has been most troublesome for me because I have seen much virtue on both sides of the fence. It has seemed to me that numerical taxonomy has a great deal to offer if viewed as an improved method for certain operations in conventional taxonomy, but not as a substitute for the whole of conventional taxonomy. With this in mind, I have attempted to ~~assemble~~ think a theoretical framework assembled from the best of both methods. The result is an approach that could be called numerical phyletics. The enthusiastic reception of a talk by Throckmorton (1966) convinced me that many taxonomists are thinking along similar lines, and the time is ripe for someone to assemble and present such a philosophy.

In order to integrate conventional and numerical taxonomy, it is necessary to review the major points of dispute. Numerical taxonomy has challenged tradition with the following points: 1) the insistance on operationalism and therefor objectivity, 2) the need for simultaneous evaluation of large number of characters and therefor the need for computer analysis, 3) the restriction on the scope of taxonomic activity, and 4) the goal of a phenetic instead of a phyletic classification. A corollary of the fourth point is 5) a challenge to the significance of the biological species. Replies by those defending conventional taxonomy emphasize problems with numerical taxonomy: 6) the loss of information helpful in conventional taxonomy, 7) the greater difficulty in using numerical taxonomy, 8) the lack of agreement among numerical taxonomists, and 9) the loss of aesthetic values. Points 2, 7 and 8 should cause no great problem. Few would question the desirability of using many characters; the hindrance is practical, not theoretical. Point 7 becomes less significant as computers become increasingly accessible and as taxonomists become less afraid of statistics and computers. Point 8 is a reflection of the fact that numerical taxonomy is still in the experimental stage.

Points 1 and 6 form the real core of the problem. Taken individually, there seems little problem. Who would question that science should be objective or that scientists should incorporate as much relevant information as feasible before reaching conclusions? However, when these points are placed in juxtaposition, the conflict become apparent. We must often weigh the possible loss of information against the error or bias that might be introduced with the information. Numerical taxonomists have ~~taken~~ implied an extreme view that any information which cannot be handled operationally must be discarded. Such a position would have been far more palatable to experienced, practical taxonomists if there had been more acknowledgment of the possible value of such information. The

Digitized by the Hunt Institute for Botanical Documentation

views of Sokal and Sneath (1963) are not inconsistent with those of Sokal and Camin (1964); but the emphasis of the latter, if better understood, would do much to reconcile the different schools. Numerical taxonomists have attempted to program taxonomic procedures for use on computers. It was inevitable that programs developed in the early years would omit much of the highly complex subjective analysis done by conventional taxonomists. Now it is time for a thorough evaluation of the subjective procedures and the theoretical justifications. Some of the procedures should be discarded, some should be used in conjunction with numerical procedures, and some should be incorporated into computer programs more complex than those feasible in earlier years. If we consider the steps one by one instead of all or nothing computer adaptation, there should be reasonable common ground.

The matter of phyletic vs. phenetic classification has been argued at great length, but the issue is still no where near resolution. Let us begin with discussion of the two main purposes of classification. One is classification as pure science, a gratification to the artist-scientist who took chaos and created order. The other purpose is classification as applied science, a filing system to help those who want to understand nature and retreive information about it. Both of these purposes have merit; but if taxonomists hope to justify claims for support from tax dollars or for respect in centers of learning, they must make the latter purpose their primary one. If taxonomy is to be a pure science only, then we ~~should~~ could work toward phylogenies and dispense with classifications. Creating classifications should be an applied science even if the result is used primarily by pure scientists. Our understanding of evolution was largely developed by taxonomists, and I would encourage taxonomists to contribute to such other areas. Nevertheless, the classification per se should be judged on the basis of how useful it is to others, not on the extent of which it gratifies the taxonomist. We must then ask, "What sort of classification is most useful?" There are two main needs, a convenient filing system for information retreival, and maximum reliability in predicting characteristics of one taxon on the basis of information about related taxa. At this point we should rule out non-hierarchial systems above the species level. The suggestions of Ehlich (1961) and DePraw (1964) make sense below the species level, but lack of a hierarchy at higher levels could only create chaos in information retreival. I have made no mention of the value of understanding the pattern of variation to the greatest possible extent. This is a worthy goal, but it ~~is~~ should be handled separately from the classification. The classification, as a hierarchy, is a condensation that must lose much of the original data; there ~~is better ways to present the pattern of variation apart from~~

Digitized by the Hunt Institute for Botanical Documentation

Having established the primary purpose and needs for a classification we can now consider the phyletic vs. phenetic issue. If we had complete knowledge, we could construct an ideal phyletic classification and an ideal phenetic classification. A phyletic classification should be consistent with the true phylogeny as explained by Simpson (1961:113), but there are many such possible phyletic classifications. The ideal one would be as phenetic as possible without becoming inconsistent with the true phylogeny. The ideal phenetic classification would be based on maximum resemblance using all possible characters. If the ideal phyletic and phenetic classifications differ, in view of the reasoning in the preceding paragraph, I would be forced to choose the latter. However, we need not waste much time on this point because any difference between ideal phyletic and phenetic classifications is dwarfed by differences between ideal and actual classifications. In actual practice the phenetic approach to classification suffers from two serious weaknesses, non random samples of characters and insufficient independent data. The theoretical basis of phenetic numerical taxonomy rests on the concept of a random sample of characters, but this is unattainable. Because of practical limitations, most taxonomists use characters which are almost all based on morphology, usually external morphology. We should not casually assume that such characters are representative of all characters; natural selection can impose correlated changes which influence morphology quite differently from other characters. The actual practice in numerical taxonomy is often far worse than necessitated by the theoretical and practical limitations. When there is no attempt to evaluate and eliminate correlations between characters, the effective number of characters may be only a small fraction of the number used. The blithe assumption that we should include all character data without character evaluation is unjustifiable. If an ideal phenetic classification exists, then we may refer to "reliable" and "unreliable" characters. Reliable characters tend to reinforce the ideal classification while unreliable ones do not. (Actually we are referring to ends of a continuous scale of character reliability.) A common line of reasoning among numerical taxonomists states that if enough characters are used, the error from unreliable characters will be mostly cancelled out, and the residual error is the lessor of two evils when compared to the personal bias introduced by any attempt to evaluate characters. The reasoning is false for several reasons. With unevaluated characters, the same factor may be

Digitized by the Hunt Institute for Botanical Documentation

repeated indefinitely; and with a non random sample of characters, the errors may reinforce each other instead of tending to cancel out. Even without introducing any personal bias, it is possible to evaluate characters in order to eliminate repetition of the same factor and to weight characters according to their reliability. This is discussed in greater detail later in this paper. But the objective criteria for evaluating character reliability (for a phenetic classification) ~~are~~ can be based on evolutionary concepts. In my opinion it is the phyletic classification in actual practice, not the phenetic classification in actual practice, that more closely resembles the ideal phenetic classification. But by "phyletic classification in actual practice," I mean a predominately phenetic classifications in which evolutionary concepts have been fully utilized wherever possible. A classification is based on a sample of characters, and there is no escape from probability and statistical sampling. If our knowledge of evolutionary principles helps us to evaluate probabilities and to say that character A is probably more reliable than character B, we ought not ignore this information. The tedious part of numerical taxonomy is gathering data about characters. Once having obtained the data, we would be foolish not to make the most of it. There is another advantage for phyletic classification; it is less likely to be upset by discovery of new ~~types.~~ taxa. If we had complete knowledge of both living and extinct forms, we would have a continuum. Evaluation of characters and clusters in light of evolutionary principles can help shift our classification in favor of one consistent with the continuum, and therefor more stable with addition of new ~~types.~~ taxa.

If the above agreement is accepted, the remaining points of difference between conventional and numerical taxonomy should fall into place. The scope of taxonomy per se is limited, but this should not inhibit the taxonomist from contributing to other fields as he had done so vitally in the past. The taxonomist who considers evolution, ecology, or some other field more aesthetically pleasing may spend as much time as he wishes in the area of his interest. If we accept evolutionary criteria as useful in classification, the importance of the biological species becomes obvious. Or better yet, we should emphasize the evolutionary species as defined by Simpson (1961:153).

If my view of the merits of conventional and numerical taxonomy is correct, a next step in implementation should be the development of more sophisticated computer programs that utilize more of the information previously ignored by numerical taxonomists. With this in mind, I prepared a program that incorporates five innovations. The first is a method to eliminate correlations between characters as found within subgroups of OTU's. The second and third are alternative

Digitized by the Hunt Institute for Botanical Documentation

~~ways to evaluate characters reliability and determine objective character weights.~~

If my view of the merits of conventional and numerical taxonomy is correct, a next step in implementation should be the development of more sophisticated computer programs that utilize more of the information previously ignored by numerical taxonomists. The program of Camin and Sokal (1965) to determine cladograms with the principle of parsimony of evolutionary change is a major step in this direction. The procedure developed by Throckmorton (in press) for eliminating characters that vary between the most closely related species is another significant advance. It demonstrated the value of operationally defined character evaluation. Both of these programs were designed for two state or multistate characters. My own program is designed to accomodate continuous ~~variables~~ characters and incorporates five innovations in data manipulation. The first is a method to eliminate correlations between characters as found within subgroups of OTU's. The second and third are alternative ways to evaluate character reliability and determine objective character weights. The remaining two are variations in clustering algorithms. The innovations and a number of older numerical procedures were tested with actual data consisting of 20 measurements on 146 adult male grasshoppers belonging to 15 species. Results are evaluated according to how well individuals of the same species are recognized as more similar than individuals of different species.

MATERIALS AND METHODS

The material for this study consisted of 146 adult male grasshoppers of the species listed in Table 1. These species were selected primarily because of an excellent revision of the genus Orphulella Giglio-Tos (Gurney, 1940). The thoroughness of this revision provides assurance that the taxonomic species correspond closely to the species as found in nature even though several of the species are so similar that they had been seriously confused by competent taxonomists. Gurney was able to distinguish the species by extracting the concealed male genitalia. Thus we have a standard against which to measure the results of the present numerical study, which uses only superficial measurements as characters. Species from several genera related to Orphulella were added to provide a variety of degrees of similarity.

The characters used were measurements that could be made easily and accurately from the dorsal view using a toolmaker's measuring microscope. They are based on 22 points on the head and pronotum. The coordinates, accurate to 0.0001 inch, were recorded for each point by two persons. The coordinates were processed by computer to shift and rotate the coordinates,

Digitized by the Hunt Institute for Botanical Documentation

test the symmetry, average right and left sides, average measurements by two persons, convert to millimeters, and derive the measurements listed in Table 2. When the test of symmetry or the test of agreement between measurements by two persons exceeded established limits, the specimens were remeasured. Since the method of measurement corresponds to a two dimensional projection of points actually in three dimensions, correct rotation of the specimen was critical. Rotation around the dorso-ventral axis was mathematically corrected and of no further consequence. Rotation around the longitudinal axis would appear as a lack of symmetry. Rotation around the transverse axis could be tested only by carefully observing the specimen at the time of measurement. Forward or backward tilt of the head probably caused the only significant inaccuracy in the data. Characters HR1 through HR12 and PR1 through PR8 (see Table 2) are the ones derived directly from the measurements. These are referred to as "raw characters." It is obvious that all of these characters would be more or less correlated with size. The remaining characters are derived by manipulating the original characters to intuitively eliminate correlation between characters. These are referred to as "intuitive characters."

The following procedure was developed for systematically and objectively eliminating correlations between characters. The procedure requires that OTU's (individual specimens in this study) have been previously assigned to clusters in an acceptable ~~(theoretically justifiable)~~ manner. ~~(In this study the OTU's are the individual specimens.)~~ For the method to work effectively, there must be at least three large clusters (at least about ten OTU's per cluster). There may be additional small clusters which participate only partially in the analysis. The algorithm is explained by the flowchart in Figure 2. Block A is to compute intraclass correlation for each character using both large and small clusters. A "perfect" character will vary between clusters but not within clusters, an intraclass correlation of 1.0. A "worthless" character will vary between clusters only to the extent of sampling error within clusters, an intraclass correlation of 0.0. For a more extensive explanation, see Snedecor (1956:282). Block B is to select the "best" character, the one with the largest intraclass correlation.

Digitized by the Hunt Institute for Botanical Documentation

This "best" character is designated as a finalized character and will not be adjusted any further. If the intraclass correlation of the "best" character is very high (more than about 0.98), it cannot be effectively tested for correlation with other characters. Otherwise ~~Step C must be performed on each non finalized character. Step C is to test the character~~ all non finalized characters must be tested for correlation with the current "best" character. Characters are considered correlated if the mean of the coefficients of correlation in the large clusters differ significantly from a value of zero using Student's t-distribution for a confidence level of 0.95. If the correlation is very high, the non finalized character must be deleted from the study. Failure to test this can cause the character to be adjusted so that only rounding off error remains. If the correlation is found to be significant but not complete, it is necessary to adjust the non finalized character so that it is not correlated with the "best" character. This is ~~Step D~~ Block C in Fig. 2. The correction factor is the mean of the slopes of the regression lines of the non finalized character on the "best" character in the large clusters. All values for the non finalized characters are adjusted by substracting the correction factor times the corresponding value for the "best" character. A new value must then be computed for the intraclass correlation of the non finalized character.

Each character was standardized by substracting the minimum and dividing by the range. Resulting values vary from 0.0 to 1.0. In some analyses the characters were weighted in proportion to the intraclass correlations. Similarities between OTU's were measured using the coefficient of correlation, the mean character difference and the average taxonomic distance.

Clustering was done using the centroid method described by Proctor (1966) but applying weights computed in a new manner. When two or more elements are combined to form a new cluster, the weight of each element is the sum of the line segments connecting different levels of similarity in a dendrogram for the element divided by the total weight. This could be expressed as

$$w_e = \frac{w_p + (c_n - c_p)}{[w_p + (c_n - c_p)]}$$

where w_e = relative weight of element in cluster being formed

w_p = total weight of prior cluster

c_n = level of new cluster

Digitized by the Hunt Institute for Botanical Documentation

and c_p = level of prior cluster in which the element was created

This assumes similarity coefficients in which the lowest value represents greatest similarity. If the reverse is true, change "$c_n - c_p$" to "$c_p - c_n$." If the element is a single OTU; then w_p equals zero, and c_p equals the similarity value for identical OTU's. Fig. 3 helps to illustrate the method. When cluster ABC is formed, the weights are computed as

$$w_A = w_B = w_C = \frac{0+(1-0)}{(1+1+1)} = 1/3$$

When D joins the cluster, weights are again computed.

$$w_{ABC} = \frac{3+(2-1)}{[3+2(2-1)]+[0+(2-0)]} = 4/6 = 2/3$$

$$w_D = \frac{0+(2-0)}{6} = 2/6 = 1/3$$

And when E joins the cluster, the new weights are

$$w_{ABCD} = \frac{6+(4-2)}{[6+(4-2)]+[0+(4-0)]} = 8/12 = 2/3$$

$$w_E = \frac{0+(4-0)}{12} = 4/12 = 1/3$$

For use in FORTRAN programs, the following shortcuts may be used. An array, W, is set to zero before the start of clustering with W(I) representing a weight to be associated with OTU I. When it is determined that OTU's I and J are to be combined at clustering level CL to become a new cluster replacing OTU I, the following steps are needed:

```
WJ = W(J) + CL
W(I) = W(I) + WJ
WJ = WJ/(W(I) + CL)
WI = 1.0-WJ
```

WI and WJ are the relative weights to be used in combining OTU's I and J. Proctor (1966) has shown the advantage of the centroid method of clustering. In this study these weights were used in computing average characters for the new cluster and computed new similarity coefficients for the new OTU I with each remaining OTU. Those not wishing to use the centroid method may apply

Digitized by the Hunt Institute for Botanical Documentation

the weights directly to the similarity coefficients of the old OTU's I and J with other OTU's.

A completely different form of clustering was used in some cases. Similarity values were converted, when necessary, to a scale in which identical OTU's have a dissimilarity value of zero. The dissimilarity between cluster A and cluster B was defined as the harmonic mean of the dissimilarities of each member of cluster A with each member of cluster B. The computer programming is not so difficult as might be suspected. It is necessary to add only one vector to keep track of the number of OTU's in each cluster. (Initially each OTU is defined as a cluster of one.) When cluster A and B join, the dissimilarity between the new cluster and cluster C is

$$\frac{n_A + n_B}{n_A / d_{AC} + n_B / d_{BC}}$$

where n_A is the number of members in cluster A

and d_{AC} is the dissimilarity between clusters A and C.

In some analyses the sequence in which the OTU's form clusters was used as a starting point for further character analysis. One procedure was to obtain a total for each character of the character differences between the elements joined in each new cluster. This was done for the first half of the clustering sequence. A large total would imply that the character is unreliable. Characters could than be weighted in inverse proportion to the sums of character difference except that no character was weighted more than three times as heavily as the least weighted character. In another form of analysis, the clustering sequence was used to define clusters for a new computation of intraclass correlation and adjustment of correlated characters. The cutoff level in the sequence was selected by computer primarily to maximize the number of clusters with ten or more members and secondarily to minimize the number of clusters.

The matrix of similarity coefficients provided the basis for judging the appropriateness of the results. ~~In runs using 20 characters on 146 OTU's~~ A frequence tally was made of values of similarity coefficients between OTU's (individuals) of the same species. A separate frequency tally was made of values of similarity coefficients between OTU's belonging to different species. The two frequency tallies was analyzed by computing the means, standard deviations and standard error of the means. The difference in means was divided by the mean of the standard deviations to provide a value that I am calling the "index of ~~cluster~~ matrix reliability."

Digitized by the Hunt Institute for Botanical Documentation

All operations were programmed in FORTRAN IV and executed on an IBM 360 Model 75. The accuracy of the results was verified by spot checking with independently written programs on an IBM 1130. All computations on the IBM 360 were incorporated in a single program, which is still being improved by the addition of more options. It is able to handle missing data and to plot dendrograms. Addition options are yet to be programmmed for clustering. When this is completed the program will be made available directly or through SHARE.

RESULTS

The results are summarized in Table 4 ~~which~~ and can best be explained by giving the sequence of computer operations necessary to obtain some of the entries. The following are the major steps executed in a single run taking five minutes of computer time.

Digitized by the Hunt Institute for Botanical Documentation

1 Read in OTU identification names and character set I (20 intuitive characters for each of 146 OTU's).

2 Standardize characters, producing character set II.

3 Compute matrix of average taxonomic distances. (Index of matrix reliability at this stage would be 1.78.)

4 Obtain clustering sequence.

5 Weight character set II in inverse proportion to sums of character differences, producing character set III.

6 Compute matrix of average taxonomic distances. (Index of matrix reliability at this stage would be 1.96.)

7 Obtain clustering sequence.

8 Weight character set III in inverse proportion to sum of character differences, producing character set IV.

9 Compute matrix of average taxonomic distances. (Index of matrix reliability at this stage would be 2.02.)

10 Obtain clustering sequence.

11 Define optimum clusters from latest clustering sequence.

12 Restore character set I.

13 Adjust characters to eliminate correlations, producing character set V.

14 Standardize character set V, producing character set VI.

15 Compute matrix of average taxonomic distances.

16 Obtain clustering sequence.

17 Weight character set VI in inverse proportion to sums of character differences, producing character set VII.

18 Compute matrix of average taxonomic distances.

19 Obtain clustering sequence.

20 Weight character set VII in inverse proportion to sums of character differences, producing character set VIII.

21 Compute matrix of average taxonomic distances.

22 Obtain frequency tallies of average taxonomic distances within and between species and compute index of matrix reliability (2.00).

23 Obtain clustering sequence.

24 Define optimum clusters from latest clustering sequence.

25-37 Repeat steps 12 through 24. The index of character reliability is 2.11.

38-49 Repeat steps 12 through 23. The index of matrix reliability is 2.04.

50 Plot dendrogram. (This step is not necessary to obtain results presented in Table 3.)

Digitized by the Hunt Institute for Botanical Documentation

51 Restore character set VI.

52 Weight characters in proportion to final intraclass correlations computed in step 13.

53-62 Repeat steps 15 through 24. The index of matrix reliability is 1.96.

63-65 Repeat steps 12 through 14.

66 Weight characters in proportion to final intraclass correlations computed in step 67.

67-76 Repeat steps 15 through 24. The index of matrix reliability is 2.12.

77-89 Repeat steps 63 through 75. The index of matrix reliability is 2.07.

90 Repeat step 50.

DISCUSSION

Criterion for evaluating procedures. -- My criterion for desirability of the matrix of similarity coefficients is the "index of matrix reliability" defined ~~earlier in this paper~~ above. Essentially it is a measure of the reliability of producing a matrix that will lead to the "correct" clusters. Identification of "correct" clusters must depend on some external criterion considered more reliable than the numerical method being tested. In this case the "reliable external criterion" is the biological species as defined by conventional taxonomy. I am well aware of the sensitivities of numerical taxonomists to having their results judged in terms of faithfulness in reproducing conventional taxonomy. Nevertheless, I am convinced that the approach is justified in this case. The species were selected with this point in mind; and the established, reliable, conventional characters were not utilized in the numerical study. In this way the numerical study was deliberately made less reliable, Also, use of only twenty characters falls considerably short of recommended numerical taxonomic standards. Thus I am determining which numerical procedure can make the most out of inadequate data. An alternative value, more or less proportional to the index of cluster reliability, is the difference in means of the frequency tallies of similarity coefficients divided by the standard error of the mean. I have not emphasized this value because it is highly dependent on the number of OTU's. In this study there is no need to question whether the two means are significantly different. The values for the difference in means divided by the standard error of the mean ranged from 64 to 114.

Digitized by the Hunt Institute for Botanical Documentation

Correlated characters. -- Numerical taxonomists are gradually becoming aware of problems caused by using correlated characters. The early attitude was to ignore the matter or to dismiss it as unimportant. For example, Sokal and Sneath (1963:67) state, "When a character B depends in part upon another character A, the decision whether to employ B as well as A will depend on the nature of the factors other than A that affect B. If to the best of our knowledge, these factors reflect heritable variations, we would include B." In other words, there would be no need to evaluate what proportion of the variance in B is dependent on A. Yet there have been indications of awareness of the difficulty. Sokal (1962) and Rohlf and Sokal (1965) recommend using correlation coefficients and not distances when most of the characters involve significant size components. Criticism such as that of Mayr (1965) has undoubtedly contributed to the greater awareness of this problem in recent years. For example see Rohlf (1967) and Steyskal (1968).

Four basic approaches may be suggested for overcoming the problem of correlated characters: 1) use so many characters that the results are not overly influenced by any single factor, 2) discard correlated characters, 3) perform manipulations which reduce the weight of factors which influence multiple characters, 4) modify the characters in such a way as to eliminate correlations. Approach one is naive and fallacious; it is listed here only because it is implied in some publications. If a study uses ten characters, each of which is 50% dependent on size, the resulting classification will be 50% dependent on size. Using a hundred characters, each of which is 50% dependent on size, will still produce a classification that is 50% dependent on size. The second approach is not much better. No one would want to include two characters that measure the same thing. On the other hand, to accept only completely independent characters would rule out all but a few of the usual characters. In actual practice most taxonomists must deal with partially correlated characters and use approach three or four. The correlation coefficient is an example of the third approach, but Eades (1965) and Minkoff (1965) point out that its theoretical basis is unsound. There are numerous possibilities such as the coefficient of Mahalanobis for weighting characters according to their degree of independence. However, this is a poor use of data; the independent component of a character is partially suppressed along with the correlated component. Approach four has the advantage of attempting to correct for correlations while fully utilizing

Digitized by the Hunt Institute for Botanical Documentation

any independent component that can be located. Use of ratios instead of absolute measurements is a well established practice in conventional taxonomy; but with computers at our disposal, it should now be possible to achieve character independence in a more systematic and less subjective manner. The trend-surface analysis (Sneath, 1967) is a most interesting approach. It can remove most correlations due to basic trends or distortions in body shape. It suffers, though, from inability to deal with correlations that do not fit patterns for which tests are made (linear, quadratic and cubic trends). Also, it cannot deal with meristic characters, which are often highly size correlated.

Before discussing other methods, it is essential to emphasize differences between three basic causes for character correlations: variation in multiple characters resulting from the influence of a common environmental factor, multiple effects of a single genetic factor, and independent genetic changes with correlated occurrence due to the phylogeny. If we had perfect knowledge, most taxonomists would eliminate characters measuring the first, would prevent redundancy of characters measuring the second, and would fully utilize characters measuring the third. Any procedure that attempts to deal with character correlations is doomed to failure if it does not attempt to separate these causes. In a study of living vertebrates, we would find complete correlation between the presence of hair and the presence of three ear ossicles. Characters correlated because of independent genetic events occurring at related places in the phylogeny provide important reinforcement of good classifications. On the other hand, if we study five species of earwigs and discover that in each case the forceps of large individuals are more than proportionately large, we are not being unreasonable in suspecting a developmental relationship between body size and proportional forcep size. Then a large species with extremely large forceps should initially be considered different from the average species in only one character. Assuming that body size is used as that one character, what can we do with forcep size? The conventional taxonomist is likely to ignore it because he realizes the correlation. Early numerical taxonomists would be likely to use it in spite of the correlation. The ideal solution would be to discover the mathematical relationship between the two characters so that forcep size could be corrected to eliminate the influence of body size. In this way, if there is an independent factor influencing forcep size, that independent factor will be recognized and utilized.

Factor analysis could be a highly useful form of analysis, but it has not, to my knowledge, been applied to taxonomy in a suitable manner. The Q-type analysis (producing factor weightings for OTU's) has been used most in taxonomy (for example, see Rohlf and Sokal, 1962). This approach is unsuitable because

Digitized by the Hunt Institute for Botanical Documentation

it makes no distinction between different causes of correlation and because conclusions are based on the single most heavily weighted factor in each OTU. The R-type analysis of Rohlf (1967) utilizing projections on centroid axes is much more sophisticated and closer to a satisfactory method. One serious deficiency remains, there is still no distinction between different causes of correlation. Also, characters with low communalities (low correlation with other characters) should be among the most valuable characters; instead they tend to be lost. However, a properly done factor analysis would be better than the form of analysis presented in this paper because factor analysis is able to evaluate all characters simultaneously without imposing some preselected sequence of analysis. The factor analysis that someone should attempt requires a separate R-type analysis within each of several clusters of OTU's, using as many orthogonal factors as there are characters, using communalities of one, averaging in some manner factor rotations of the separate analyses, obtaining projections on centroid axes, weighting projections on the basis of intraclass correlation, and proceeding with a usual numerical analysis with the resulting highly modified characters. To prepare such a computer program would obviously be no small task.

The search for a practical way to eliminate correlations of the undesirable kind led me to the algorithm described in the prior section, which is a generalization of a procedure used earlier (Eades, 1964). It assumes a common mathematical relationship between characters in each species in the group. This assumption may be wrong, and for this reason we must test several species (or other predefined clusters) within the group. Also, correlation in a single species could be the accidental result of independent genetic changes, but this explanation is unlikely when similar correlations occur in each of several species tested. When a mathematical relationship exists, it will often be nonlinear; but in very few cases will the data be adequate to prove this. In this study I have tested only for linear relationships. Correction for linear correlation is far better than no correction, and a systematic search for nonlinear correlation would involve more computation than feasible or even useful in most studies. The distinction between the causes for character correlations is not perfect, but it is a significant step in that direction. Correlations due to multiple influences of an environmental factor are reduced to one character, not eliminated as we would like. Repetition of multiple effects of a single genetic factor is corrected if the factor varies within species but not if it varies only between species. Characters that are correlated because of independent genetic changes ~~in~~

Digitized by the Hunt Institute for Botanical Documentation

in related places in the phylogeny are allowed to retain their full rank as independent characters.

If we are to test for correlations within species, we must have species or other acceptable clusters that can be recognized before the main analysis. It is impossible to define useful clusters without some sort of prior taxonomic study. To satisfy the numerical taxonomic purist, we may perform a preliminary study without analysis of correlated characters in order to derive preliminary clusters for the main study. Thus in Table 3, the clusters derived after analysis as shown in line 1 were used as the preliminary clusters for lines 2 and 6. A more effective procedure in most studies would be to define the preliminary clusters as the units subjectively recognized by an experienced taxonomist. In using the procedure proposed in this paper for eliminating correlation between characters, it is important to realize that there is a tendency for adjustments to reinforce the predefined clusters. Thus if preliminary clusters are poorly chosen, an element of unfortunate bias is introduced. However, the bad effect of this bias is more than offset by the gain in the effective number of independent characters. There is no need to stop with only one cycle of adjusting correlated characters. If the taxonomist feels uneasy about the validity of his predefined clusters, he may use the results after his first analysis of correlated characters in order to define better clusters for a second analysis of correlated characters. This recycling could be repeated as often as desired. In a technical logical sense the ~~analysis~~ repetition is recycling and not circular reasoning. However, there is a self reinforcing element closely akin to circular reasoning, ~~Thus~~ and too much recycling could be bad. Table 4 is a summary of Table 3 emphasizing recycling to adjust for correlated characters. From Table 4 we can see that adjusting to eliminate correlations between characters improves the results. If the species can be recognized and utilized as predefined clusters, the results are best. But even without this non numerical input, we can see that lines 2 and 6 are better than the source of predefined clusters (the analyses for line 1). When the predefined clusters are unreliable, another cycle is useful; lines 3 and 7 are better than lines 2 and 6. However, a third cycle of adjusting correlated characters (lines 4 and 8) cannot be recommended. In many studies, indivduals would be the OTU's with species as the predefined clusters. After character adjustment, the individuals of each species could be averaged, and the analysis could proceed with species as OTU's.

Digitized by the Hunt Institute for Botanical Documentation

A simpler way to adjust characters is by intuitive or subjective selection of ratios which would not have any obvious correlations. Table 3 shows that this is clearly better than no correction at all, but it is much less desirable than the systematic adjustment described above. The next question is whether both forms of adjustment should be used together. If there is a reliable source of predefined clusters, the initial conversion to ratios is undesirable. In other cases the predefined clusters would rest on uncertain ground, and preliminary conversion to ratios would be helpful.

Character weighting.--The attitude toward character weighting is another difference between conventional and numerical taxonomy. Conventional taxonomists weight characters for numerous reasons, some of them highly subjective. Numerical taxonomists ~~are~~ have been suspicious of the subjective element and usually do not weight characters except in the sense of breaking down complex characters into unit characters. When they do apply weights to characters, it is usually to offset some deviation from their goal of equal weights. There are a number of justifiable reasons conventional taxonomists weight characters. These include: 1) Origin of a new structure as opposed to its loss (Brown, 1965). A wingless beetle would hardly create the excitement of a winged spider. 2) The ancestral condition as opposed to the modified condition (Wilson, 1964). Turtles and crocodilians share a number of ancestral reptilian traits, but this is not strong evidence for a close relationship. 3) Characters that vary greatly in lower taxa are unreliable when classifying higher taxa. 4) Characters which are unreliable in related groups are likely to be unreliable in the group under study. Numerical taxonomists tend to reject these reasons because they are often subjective, require information outside the group studied, and reflect evolutionary concepts. Subjectivity may be an objection against a procedure as used, but not against the theoretical basis behind the procedure. Many subjective procedures can be defined in an operational way and made objective. The description of a character as "unreliable" is subjective if we do not clearly understand the meaning of "unreliable." As used here, it means that the character does not sufficiently agree with most other characters. In this paper I measured character "unreliability" by adding the absolute values of character differences between elements as they were combined in the first half of the clustering sequence. Intraclass correlation provided an alternative means to objectively test ~~this factor~~ "reliability." Another problem in character weighting is the use of information from outside the group being studied. Any taxonomic

Digitized by the Hunt Institute for Botanical Documentation

study must have its limit in scope, and information from outside the scope is likely to be somewhat more subjective and less documented than information from the group under study. The conventional taxonomist who introduces 30% error and personal bias in his information from outside is less to be criticized than the numerical taxonomist who discards 70% correct information from outside. The legitimate criticism of conventional taxonomists is that they should have a better understanding of the theoretical basis and greater consciousness of the personal judgements introduced in character weighting. A third problem is the matter of phenetic or phyletic classification. If we want a phenetic classification, it would seem an "unreliable" character should be just as important as a "reliable" character in determining the final classification. This would be true if we had an infinite sample of characters. But lacking this, we should utilize information that tells us which characters are more likely to agree with the non existant infinite sample.

To return now to the data at hand, let us consider the usefulness of weighting characters according to how well they agree with other characters. To evaluate this we must consolidate the data in Table 3. This has been done by averaging first, second and third cycles of eliminating correlations within predefined clusters; eliminating analyses by coefficient of correlation; and averaging results from raw and intuitive data. Table 5 shows that unless the characters have already been weighted based on intraclass correlations, one weighting cycle significantly improves results. A second weighting cycle produces little, if any, additional improvements. Another form of weighting was based on intraclass correlation of characters in the predefined clusters after the characters had been adjusted to eliminate correlations between characters within the predefined clusters. Table 6 is a consolidation of Table 3 averaging the weighting cycles and averaging the cycles of eliminating correlations between characters. Table 6 shows that this form of weighting caused a consistent improvement in results.

Similarity coefficients.--The coefficient of correlation, the mean character difference, and the average taxonomic distance were used. Analyses involving weighting were not done with the coefficient of correlation because the weighting would have to be handled in a way far more complex than with the other similarity coefficients. Matching coefficients were not used because they are less suitable for continuous characters states. Results with the coefficient for correlation were inconclusive; they included both the lowest and the highest value for the index of ~~cluster~~ matrix reliability. One of the

Digitized by the Hunt Institute for Botanical Documentation

reasons often suggested for using the coefficient of correlation is that it helps to eliminate a common factor such as size from the data. Table 6 shows that this is better done by adjusting the characters directly. In view of serious theoretical shortcomings (Eades, 1965 and Minkoff, 1965) and mildly negative experimental data (Rohlf, 1963 and Moss, 1968), the coefficient of correlation should not be used.

A useful discussion of the relative merits of mean character difference and average taxonomic distance may be found in Sokal and Sneath (1963:146). The main advantage of the mean character difference is its simplicity and ease of computation. However, with the availability of high speed computers, this is not important. The suitability of the average taxonomic distance for partitioning into size and shape components is unimportant. The same objective can be acheived more easily by dividing all characters of each OTU by the sum of characters in that OTU. When stated this way, other difficulties are apparent. The partitioning is appropriate only if every character is positively correlated with an important common factor. The taxonomist's intuition is far better than such a crude assumption. The correct solution is a systematic analysis of correlation between all pairs of characters such as the procedure proposed in this paper. But whether by conversion to intuitively selected ratios or by systematic analysis, correlated characters should be dealt with before beginning the computation of similarity coefficients. Sokal and Sneath (1963:146) mention several advantages of the average taxonomic distance relating to the fact that it applies more than proportional weight to large character differences. The analogy to standard deviation and average deviation can be most enlightening. Standard deviation is appropriate when the sample is derived from a universe that fits a normal curve. Since large deviations are rare in a normal distribution, the large deviations actually found deserve more than proportional weight. Under these circumstances, the standard deviation of a finite sample, when suitably adjusted, gives a better estimate of average deviation in an infinite sample than does the average deviation of the finite sample. Our universe is the difference between two OTU's in all possible characters. We want an estimate of the average difference. There is good reason to suspect that differences in standardized, unweighted characters may not fit a normal curve. If we start with a normal curve of differences in "good" characters and add differences in random ("bad") characters, the result is a platykurtic distribution. Under these circumstances,

Platelike curved

I can't find this word in the dictionary. Why don't you just say flat?

Digitized by the Hunt Institute for Botanical Documentation

large differences are more likely to be found in "less reliable" characters and they should not receive more than proportional weight. If there has been an attempt to weight characters according to their reliability (their degree of concordance with other characters), then the expected distribution would be much closer to normal. Table 5 fits well with the conclusions of this theoretical reasoning. If the characters have not been weighted, the mean character difference gives better results. As weights are applied by either method, the results with average taxonomic distance improve more than those with mean character difference.

Clustering procedures.--A number of clustering methods have been summarized by Sokal and Sneath (1963). Minkoff (1965) demonstrated that the choice of clustering method can have considerable effect on the appearance of the final dendrogram. Sokal and Michener (1958) suggested the desirability of a compromise between the weighted pair group method and the unweighted pair group method. Such a compromise was described earlier in this paper. The weights were utilized in averaging character values in accordance with the centroid method described by Proctor (1966).

An area that needs more study is in clustering algorithms which permit clusters of shapes other than hyperspheroidal in character space. Taxa in conventional taxonomy are defined more by discontinuities between taxa than by the amount of variation within taxa. One way to handle this is to cluster by minimum linkage, but this is probably too extreme. Rohlf (1967) emphasized the need to provide for hyperellipsoidal clusters. Carmichael, George, and Julius (1968) offer an algorithm to permit clusters of any shape. Their proposal is not likely to be widely adopted, however, because it is too complex and contains too many arbitrary components. It provides no progressive, continuous clustering, but instead requires a separate analysis for each of a number of different levels. But most serious, the results can be different according to the cluster which is selected as the starting point. The search for a simpler method to accomplish the same ends led me to try the harmonic mean of distance values. In effect, the harmonic mean gives greater weight to smaller values. ~~The magnitude of the smaller distances between OTU's in different clusters (and therefor the size of the gap between them), is given greater importance.~~ Preliminary results with this method have been very encouraging. A more systematic discussion of the steps from the matrix of similarity values to the classification will be presented in a later publication.

Digitized by the Hunt Institute for Botanical Documentation

Additional study.--The conclusions of this paper cannot be considered well established until they have been tested on additional sets of data. To provide a good test, I prefer data with the following characteristics: OTU's are individuals, species (or other clusters) can be reliably identified by means independent of the data set, at least three species are represented by at least ten individuals each, at least 20 multistate or continuous characters are included, and missing data are rare or else one of the numbers is much larger than quoted above. I would welcome the opportunity to analyze data sets that meet these qualifications and can be provided on punched cards, punched tape, or magnetic tape. Alternatively, I will provide the program for those who wish to use it on other computer systems.

SUMMARY

1. A variety of numerical taxonomic procedures were used on 20 characters of 146 adult male grasshoppers belonging to 15 species. The procedures were evaluated according to their effectiveness in associating individuals belonging to the same biological species.
2. Character analysis is a valuable step which should be incorporated into numerical taxonomy. Conventional procedures can be operationally defined so as to avoid the criticism of subjectivity.
3. Evolutionary concepts may be useful even in an analysis aiming toward a phenetic classification. The difference between the ideal phenetic classification and the ideal evolutionary classification is insignificant when compared to the difference between the ideal phenetic classification and the phenetic classification in actual practice. An ideal phenetic classification would be based on a random sample of characters; actual samples show drastic deviations from randomness.
4. Failure to deal with the problem of correlated characters is a serious defect in a taxonomic study. It is necessary to distinguish between correlations due to phylogeny (which should not be eliminated) and those due to factors operating on individuals but affecting multiple characters (which should be eliminated).
5. A procedure described in this paper offers a useful means to eliminate correlations of the undesired type. It requires use of predefined clusters, which may be obtained from a previous numerical analysis using unadjusted characters or from some independent source. If the analysis of correlated characters is not performed, or if it is performed with predefined clusters that lack a reliable basis, it is

Digitized by the Hunt Institute for Botanical Documentation

useful to convert the initial characters to ratios that do not have any obvious intuitive correlation.

6. Results are improved by weighting characters. This may be done by using weights proportional to the intraclass correlations. Another approach is to use weights inversely proportioned to the sums of characters differences between elements being combined in the first half of the clustering sequence. In either case, the purpose is to reduce the weight of characters that are less in agreement with most other characters.
7. When characters have not been weighted, the mean character difference is the best similarity coefficient. When characters have been weighted to reflect the extent of their concordance with other characters, the average taxonomic distance is best.
8. A compromise between the weighted pair group method and the unweighted pair group method is described for use in clustering. It is suggested that the weights be applied to character values in the centroid method.
9. An alternative method of clustering is described. The dissimilarity between two clusters is defined as the harmonic mean of the dissimilarities between OTU's in the two clusters. Preliminary results suggest this is a promising procedure to permit clusters of irregular shape.
10. The procedures presented constitute a step in the direction of numerical phyletics. They are overwhelmingly phenetic but they utilize some information that is relevant because of a phyletic context. Much more work is needed before good cladograms can be produced.

ACKNOWLEDGEMENTS

This research was supported by grant number GB5981 from the National Science Foundation and a grant from the University of Illinois Research Board. The basic character data was obtained by Dennis Nyberg and others. Portions of the computer programming were done by Stephen D. Cullen.

REFERENCES

Brown, W. L., Jr. 1965. Numerical taxonomy, convergence, and evolutionary reduction. Syst. Zool. 14:101-109.

Digitized by the Hunt Institute for Botanical Documentation

Camin, J. H., and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. Evolution 19:311-326.

Carmichael, J. W.,J. A. George, and R. S. Julius. 1968. Finding natural clusters. Syst. Zool. 17:144-150.

Dupraw, E. J. 1964. Non-Linnean taxonomy. Nature. 202 (4935: 849-852.

Eades, D. C. 1964. General biology and geographic variation of Ceuthophilus guttulosus Walker (Orthoptera: Gryllacrididae: Rhaphidophorinae. Trans. Amer. Ent. Soc. 90:73-110.

1965. The inappropriateness of the correlation coefficient as a measure of taxonomic resemblance. Syst. Zool. 14:98-100.

Gurney, A. B. 1940. A revision of the genus Orphulella Giglio-Tos from America north of Mexico (Orthopters; Acrididae). Entomologica Americana XX:85-157.

Mayr, E. 1965. Numerical phenetics and taxonomic theory. Syst. Zool. 14:73-97.

Minkoff, E. C. 1965. The effects on classifications of slight alterations in numerical technique. Syst. Zool. 14:196-213.

Moss, W. W. 1968. Experiments with various techniques of numerical taxonomy. Syst. Zool. 17:31-47.

Proctor, J. R. 1966. Some processes of numerical taxonomy in terms of distance. Syst. Zool. 15:131-140.

Rohlf, F. J. 1963. Congruence of larval and adult classifications in Aedes (Diptera, Culicidae). Syst. Zool. 12:97-117.

Rohlf, F. J. 1967. Correlated characters in numerical taxonomy. Syst. Zool. 16:109-126.

Rohlf, F. J., and R. R. Sokal. 1962. The description of taxonomic relationships by factor analysis. Syst. Zool. 11:1-16.

1965. Coefficients of correlation and distance in numerical taxonomy. Univ. Kansas Sci. Bull. 43:3-27.

Sneath, P. H. A. 1967. Trend-surface analysis of transformation grids. J. Zool. (Proc. Zool. Soc. London). 151:65-122.

Snedecor, G. W. 1956. Statistical methods applied to experiments in agriculture and biology, fifth ed. Iowa State College Press, Ames, Iowa. 534 p.

Sokal, R. R. 1962. Variation and covariation of characters of alate Pemphigus populi-transversus in eastern North America. Evolution 16:227-245.

Sokal, R. R., and J. H. Camin. 1964. The two taxonomics: area of agreement and conflict. Syst. Zool. 14:176-195.

Digitized by the Hunt Institute for Botanical Documentation

Sokal, R. R., and H. A. Sneath, 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco and London. 359 p.

Steyskal, G. C. 1968. The size factor in numerical taxonomy. Syst. Zool. 17:210.

Simpson, G. G. 1961. Principles of animal taxonomy. Columbia Univ. Press, New York. 1961.

Throckmorton, L. H. 1966. Concordance and discordance of taxonomic characters in Drosophila classification. Talk presented before Society of Systematic Zoology, Washington, Dec. 29, 1966.
(in press) Concordance and discordance of taxonomic characters in Drosophila classification. Syst. Zool. 17 (4).

Wilson, E. O. 1964. A consistency test for phylogenies based on contemporaneous species. Syst. Zool. 14:214-220.

Digitized by the Hunt Institute for Botanical Documentation

Table 1. Specimens used.

Species	Number
Phlibostroma quadrimaculatum (Thomas)	6
Cordillacris occipitalis (Thomas)	5
Cordillacris crenulata (Bruner)	3
Sisantum notochloris Bruner	1
Orphula pagana (Stål)	2
Orphulella speciosa (Scudder)	30
Orphulella pelidna (Burmeister)	31
Orphulella olivacea (Morse)	9
Orphulella tenaneca (Saussure)	1
Orphulella compta Scudder	2
Orphulella concinnula (Walker)	4
Orphulella punctata (DeGeer)	12
Clinocephalus elegans Morse	8
Dichromorpha viridis (Scudder)	31
Dichromorpha robusta (Bruner)	1

Digitized by the Hunt Institute for Botanical Documentation

Table 2. Characters.

Character number	Description
Raw characters	
Head characters	
HR1	Longitudinal distance from most lateral point on fastigium to junction of fastigium and eye as seen from above.
HR2	Longitudinal distance from apex of fastigium to most anterior point an eye (zero in some specimens).
HR3	Longitudinal distance from most anterior point on eye to junction of fastigium and eye as seen from above (zero in some specimens).
HR4	Longitudinal distance from most anterior point on eye to most lateral point on eye.
HR5	Longitudinal distance from most anterior point on eye to most medial point on eye.
HR6	Longitudinal length of eye.
HR7	Transverse distance from most lateral point on fastigium to most anterior point on eye (zero in some specimens).
HR8	Distance from median sagittal plane to most anterior point on eye.
HR9	Distance from median sagittal plane to junction of fastigium and eye as seen from above.
HR10	Distance from median sagittal plane to most lateral point on eye.
HR11	Distance from median sagittal plane to most medial point on eye.
HR12	Distance from median sagittal plane to most posterior point on eye.
Pronotum characters	
PR1	Median length of prozona.
PR2	Median length of metazona.
PR3	Longitudinal distance from front margin at median carina to front margin at lateral carina (negative value in some cases).
PR4	Longitudinal distance from front margin at lateral carina to intersection of lateral carina and principal sulcus.

Digitized by the Hunt Institute for Botanical Documentation

PR5 Longitudinal distance from intersection of longitudinal carina and principal sulcus to posterior margin at lateral carina.

PR6 Distance from median sagittal plane to lateral carina at front margin.

PR7 Distance from median sagittal plane to lateral carina at principal sulcus.

PR8 Distance from median sagittal plane to lateral carina at posterior margin.

Intuitive characters

Head characters

HI1 HR1 ÷ (HR2 + HR6)

HI2 HR2 ÷ (HR2 + HR6)

HI3 HR3 ÷ (HR2 + HR6)

HI4 HR4 ÷ (HR2 + HR6)

HI5 HR5 ÷ (HR2 + HR6)

HI6 HR7 ÷ HR10

HI7 HR8 ÷ HR10

HI8 HR9 ÷ HR10

HI9 HR11 ÷ HR10

HI10 HR12 ÷ HR10

HI11 Ratio of length to width. ((HR2 + HR6) + HR10

HI12 Size. ((HR2 + HR6) X HR10) (Intermediate character not used in analysis.)

Pronotum characters

PI1 Ratio of median length of metazona to median length of prozona. (PR2 ÷ PR1)

PI2 Ratio of lateral length of metazona to lateral length of prozona. (PR5 ÷ PR4)

PI3 PR3 ÷ (PR1 + PR2)

PI4 Ratio of length of median carina to length of lateral carina. ((PR1 + PR2) ÷ (PR4 + PR5))

PI5 PR6 ÷ PR7

PI6 PR8 ÷ (PR6 + PR7 + PR8)

PI7 Ratio of length to width. ((PR1 + PR2 + PR4 + PR5) ÷ (PR6 + PR7 + PR8))

PI8 Size. ((PR1 + PR2 + PR4 + PR5) X (PR6 + PR7 + PR8)) (Intermediate character not used in analysis.)

Overall characters

SZR Ratio of size of pronotum to size of head. (PI8 ÷ HI2)

TSZ Overall size. (PI8 + HI12)

Digitized by the Hunt Institute for Botanical Documentation

Table 3

Index of matrix reliability using raw characters (roman type) and intuitive characters (italics).

	Line number	Source of predefined clusters	Similarity coefficient						
			Coefficient of correlation	Mean character difference			Average [illegible] distance		
				No weighting cycle	One weighting cycle	Two weighting cycles	No weighting cycle	[illegible] ighting cle	Two weighting cycles
Characters not adjusted to eliminate correlations	1		1.57 1.88	1.68 1.78	1.70 1.89	1.69 1.89	1.69 1.78	1. [illegible] .96	1.80 2.02
Characters adjusted to eliminate correlations within predefined clusters									
Characters not weighted by intraclass correlations									
First cycle	2	Line 1	1.96 1.97	1.83 1.92	2.04 2.12	1.98 2.18	1.53 1.87	1.85 1.97	2.06 2.00
Second cycle	3	Line 2	2.04 2.04	2.01 2.02	2.08 2.26	1.99 2.17	1.49 1.97	1.89 2.04	2.04 2.11
Third cycle	4	Line 3	2.15 2.08	2.00 2.07	2.00 2.03	1.96 2.16	1.52 1.81	1.82 2.07	1.90 2.04
Control	5	Actual species	2.52 2.05	2.19 2.08	2.32 2.16	2.32 2.17	2.19 2.03	2.41 2.21	2.41 2.25
Characters weighted by intraclass correlations									
First cycle	6	Line 1		1.90 2.01	2.04 2.15	1.98 2.18	1.66 2.03	1.88 2.01	2.05 1.97
Second cycle	7	Line 6		2.14 2.18	2.18 2.21	1.96 2.08	1.75 2.22	2.03 2.08	2.04 2.12
Third cycle	8	Line 7		2.25 1.98	2.27 2.12	1.90 2.21	1.88 2.28	2.03 2.03	2.06 2.07
Control	9	Actual species		2.35 2.31	2.35 2.24	2.32 2.18	2.46 2.44	2.44 2.30	2.41 2.21

Digitized by the Hunt Institute for Botanical Documentation

Table 4

Average index of matrix reliability by line from Table 3

	Line number	Source of predefined clusters	Average value
Characters not adjusted to eliminate correlations	1		1.79
Characters adjusted to eliminate correlations within predefined clusters			
Characters not weighted by intraclass correlations			
First cycle	2	Line 1	1.95
Second cycle	3	Line 2	2.01
Third cycle	4	Line 3	1.97
Control	5	Actual species	2.23
Characters weighted by intraclass correlations			
First cycle	6	Line 1	1.99
Second cycle	7	Line 6	2.08
Third cycle	8	Line 7	2.09
Control	9	Actual species	2.33

Digitized by the Hunt Institute for Botanical Documentation

Table 5

Condensation of Table 3 as explained in text.

	Mean character difference			Average taxonomic distance		
	No weighting cycle	One weighting cycle	Two weighting cycles	No weighting cycle	One weighting cycle	Two weighting cycles
Characters not adjusted to eliminate correlations	1.73	1.80	1.79	1.73	1.87	1.91
Characters adjusted to eliminate correlations within predefined clusters						
Characters not weighted by intraclass correlations						
Average of cycles	1.97	2.09	2.07	1.70	1.94	2.02
Control	2.14	2.24	2.25	2.11	2.31	2.33
Characters weighted by intraclass correlations						
Average of cycles	2.08	2.16	2.05	1.97	2.01	2.05
Control	2.33	2.29	2.25	2.45	2.37	2.31
Average	2.05	2.12	2.08	1.99	2.10	2.12

Digitized by the Hunt Institute for Botanical Documentation

Table 6

Consolidation of Table 3 as explained in text.

	Coefficient of correlation		Mean character difference		Average taxonomic distance	
Characters not adjusted to eliminate correlations	1.57	1.88	1.69	1.85	1.76	1.92
Characters adjusted to eliminate correlations within predefined clusters						
Characters not weighted by intraclass correlations						
Average of cycles	2.05	2.03	1.99	2.10	1.79	1.99
Control	2.52	2.05	2.28	2.14	2.33	2.16
Characters weighted by intraclass correlations						
Average of cycles			2.07	2.13	1.93	2.09
Control			2.34	2.24	2.44	2.32

Digitized by the Hunt Institute for Botanical Documentation

Legends for figures

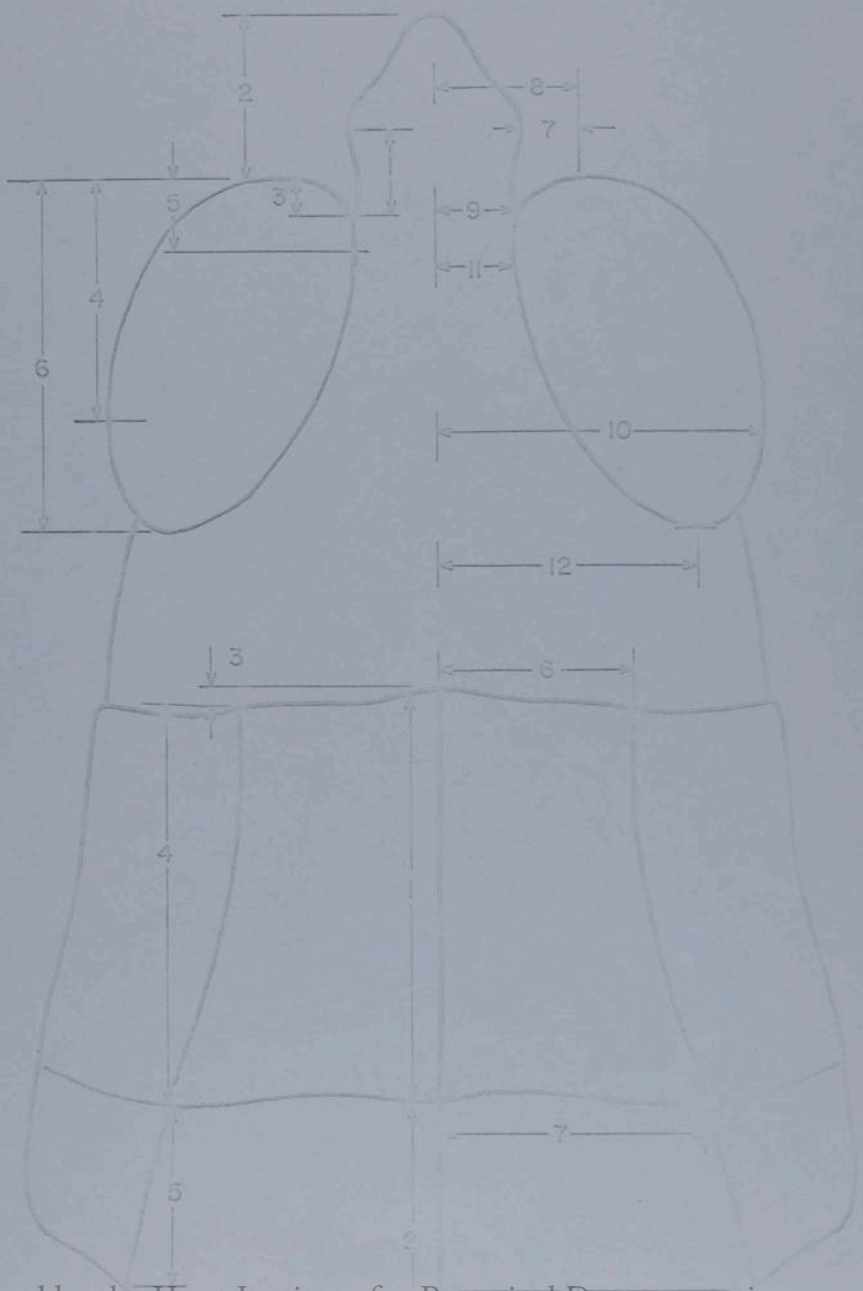
Fig. 1. Diagram of the head and pronotum of a grasshopper in dorsal view showing the characters utilized.

Fig. 2. Flowchart for adjusting characters to eliminate undesirable correlations. Lettered blocks are as follows: A. compute intraclass correlation for each character, and designate all characters as non finalized; B. select non finalized character with greatest intraclass correlation and designate it as the "best" character; C. compute mean of slopes of regressions of non finalized character on "best" character in large clusters, for each OTU subtract the mean slope times the "best" character from the non finalized character, compute new intraclass correlation for the non finalized character.

Fig. 3. Hypothetical dendrogram.

Digitized by the Hunt Institute for Botanical Documentation

Fig. 1



Digitized by the Hunt Institute for Botanical Documentation

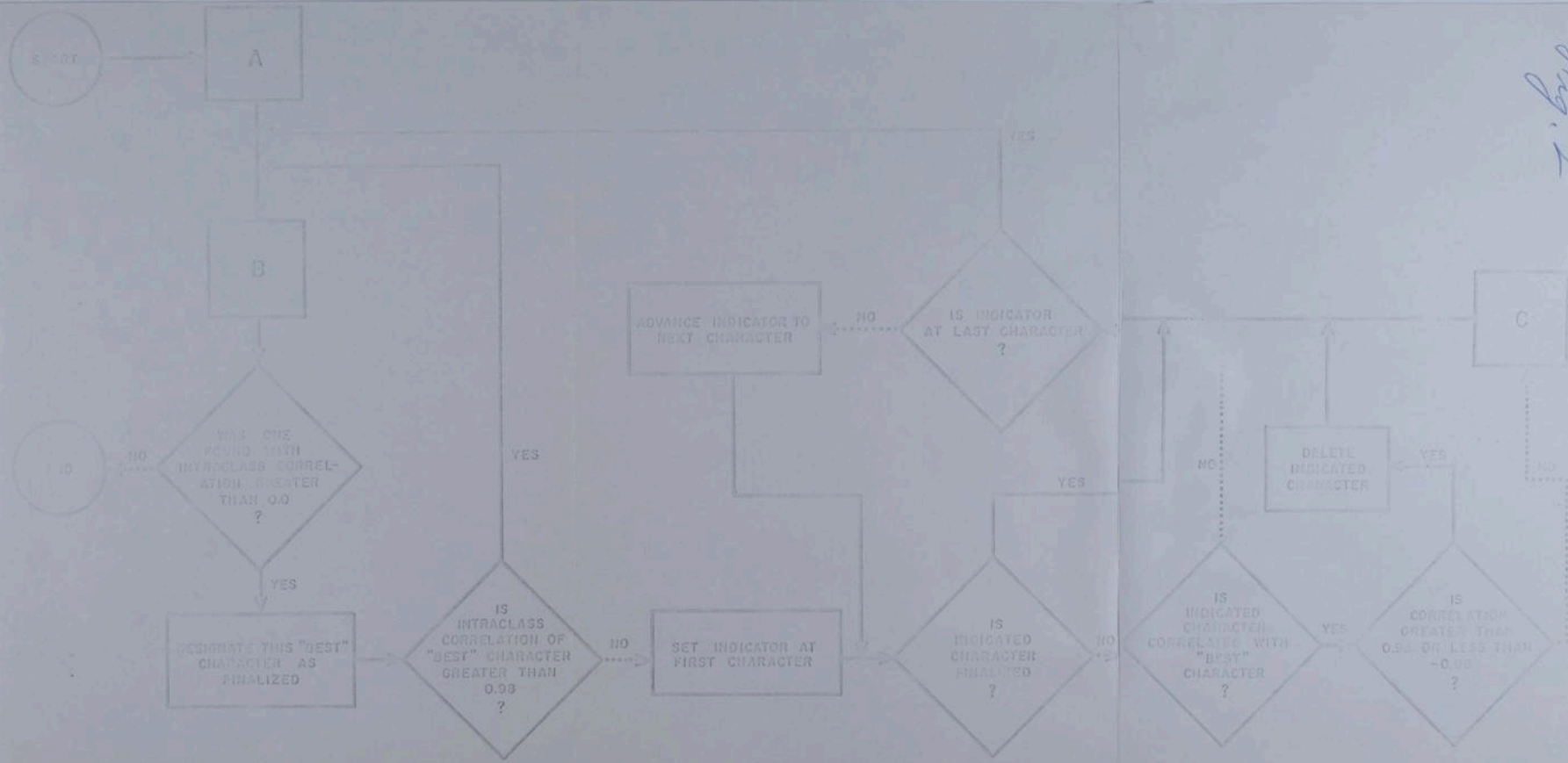


Fig. 2

Digitized by the Hunt Institute for Botanical Documentation

Fig. 3

Digitized by the Hunt Institute for Botanical Documentation

August 19, 1968

Dr. P. C. Silva
University of California
Berkeley

Dear Dr. Silva:

Several thinks have caused my delay in answering your letter of the 2nd. The most obvious is the change of address, which please note. We left New York in 1965.

Levin's paper is absolutely worthless. I have not made any suggestions for corrections, though I have seen the paper which he must have turned in to you. It is too hope-less because of a number of failures on his part. The first is that I told him some time ago that the computer program which he used, and which I designed in 1961, was no good, but another one we had developed was better. He chose to ignore my suggestion that he abandon the method which he has continued to use. Secondly, Levin is about the world's worst taxonomist, with no insight, ability, or direction. Anything he may have said about the Nemalionales would not be worth anything. We agree, therefore, that Levin's paper should be rejected.

I think that I should say a word or two in defense of my position in relation to this effort, however. I tried desperately to get a good phycological taxonomist to work with me when we started the project, but could find no other person than Levin. He was recommended from Rutgers, having received parts of his training from the Rutgers algologist, whose name escapes me at the moment. We saw Levin's short-comings almost immediately, and fired him after three months.

Further, I should also defend the procedure we used to test our computer procedures. Since we had no other criterion for the value of the program for taxonomic purposes, we needed to discover whether we could reproduce the decisions of a taxonomist, based on some published work. We are not, nor ever have, attempted to change taxonomic classifications, but wanted to produce a tool which would aid the taxonomist. This, then, was the basis for using some established classification as a criterion. As mentioned earlier, we found that the initial efforts were not useful, and have since produced several programs which do what we want them to do, and with which we are quite well satisfied. These programs have been used successfully on several new problems, none unfortunately with the algae, but with organisms in several different divisions. If you have any student who would like to have some instruction in our procedures, we'll be glad to give him an intensive course which will tell him what he must do to proceed. But he has to be a good algologist, first and foremost.

Hope these comments are useful to you.

Sincerely,

David J. Rogers
Professor, Dept. of Biology

Digitized by the Hunt Institute for Botanical Documentation

AUG 8 1

UNIVERSITY OF CALIFORNIA, BERKELEY

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

SANTA BARBARA • SANTA CRUZ

HERBARIUM
DEPARTMENT OF BOTANY

BERKELEY, CALIFORNIA 94720

INTERNATIONAL PHYCOLOGICAL SOCIETY

August 2, 1968

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

Dear Dr. Rogers:

A manuscript has been submitted to PHYCOLOGIA by M. H. Levin of the University of Manitoba entitled: The Nemalionales: a comparison of taximetric and conventional taxonomic results. In his letter of transmittal, Levin says that the manuscript was sent to Dr. Humm, Dr. Crovello, and to you, and that the major part of your criticisms have been incorporated.

I am singularly unimpressed with the paper, but seek opinions from those more competent in this field than I am. Levin states that his program was essentially that used by Rogers & Fleming with the Halimeda monograph. Is there any point to continuing to 'test' existing monographs? My feeling is that we should be using computers at the time the taxonomic conclusions are drawn, not afterward. Moreover, he doesn't seem to have worked closely with a phycologist, to judge from some painfully naive statements.

If you could make some comments on the basis of your recollection of the paper, or if you would like me to send you the manuscript, I would certainly appreciate and value your opinion.

There is no more effective way to kill a good thing than to misuse or misrepresent it, and I anticipate that the phycological community would receive this paper with horse-laughs. But I would be pleased to be convinced differently.

Sincerely,

P. C. Silva
Editor, PHYCOLOGIA

Digitized by the Hunt Institute for Botanical Documentation

UNIVERSITY OF CALIFORNIA, BERKELEY

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

SANTA BARBARA • SANTA CRUZ

HERBARIUM
DEPARTMENT OF BOTANY

BERKELEY, CALIFORNIA 94720

INTERNATIONAL PHYCOLOGICAL SOCIETY

September 6, 1968

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

Dear Dr. Rogers:

First, my apologies for not having kept up with your whereabouts. I am certainly impressed with the strength that the University of Colorado is building up in biology.

I greatly appreciate your comments on Levin and his manuscript. From another continent I have received a similar appraisal, so I have no choice but to reject the paper.

I have been pushing the need for getting into taximetrics in the algae at staff meetings for the past several years, but the difficulty is that I have absolutely no time to even think about diversifying my own efforts and no likely candidate has yet arrived on the Berkeley scene, but I keep hoping. It would certainly be desirable from our point of view to have an interested student trained in your laboratory. If I had time, I would like to bring my Codium problems to you. Perhaps one of these years.

With many thanks and good wishes,

Sincerely,

P. C. Silva
Editor, PHYCOLOGIA

Digitized by the Hunt Institute for Botanical Documentation

27 August 1968

Dr. Murray F. Buell
Box 234
Middlebush, New Jersey 08873

Dear Murray:

I'm really surprised that Dave Fairbrothers would let this paper out with his name on it. I don't think it's worth publishing, at least with its present status..

First of all, one cannot do much with the taxonomy of a species, and certainly cannot come to any classificatory rearrangements on the basis of a study which admittedly takes in only a very small portion of the total range of the species, which does not consider the relationships of this species to any other species in the genus, and which is only descriptive of plants in eleven sites. Very obviously, the authors did not examine type material, where such exists, nor compare historical usage of the names, and all they did say was that Gleason and Cronquist are probably right in their choice of the name for this variable taxon. Secondly, I can see little value in the primitive statistical techniques employed when pure old "eye-balling" the data would have shown the same things.

The paper's purpose is not given, in spite of the fact that there is a paragraph headed with that title. There is no summary at the end, which is probably a good thing because there's nothing to summarize; though not very wise in such a paper where the title leads one to think that there should be.

About the only suggestion I can make is that this paper be mimeographed, and stuck in the files of the botany department, and later picked up and added to by more studies along the lines suggested in paragraph two above.

Sincerely,

David J. Rogers
Professor of Biology

DJR:em

Digitized by the Hunt Institute for Botanical Documentation

RUTGERS · THE STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

Biological Station
Lake Itasca, Minn.
July 21, 1968

NEW BRUNSWICK, NEW JERSEY 08903

Dr. David Rogers,
Dept. of Botany
University of Colorado
Fort Collins, Colorado

Dear Dave,

The middle of Sept. will be ok so I am enclosing the LaBonde-Farbrothers ms. herewith. Thanks for taking on this job.

If you know of anyone who might have a manuscript ready which would be suitable for the Torrey Bulletin we can give rather prompt publication at this time.

Sincerely,

Murray

Murray F. Buell

Digitized by the Hunt Institute for Botanical Documentation

JUL 3 0 1968

JUL 1968

THE UNIVERSITY OF MANITOBA

FACULTY OF ARTS AND SCIENCE

DEPARTMENT OF BOTANY

WINNIPEG, CANADA

23 July 1968

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

Dear Dave:-

I have prepared a copy of the manuscript which you may keep. Minus the acknowledgements this incorporates the majority of the suggestions received to date. In this form, and with encouragement, it has been sent to Phycologia.

I have recommended to the Editor that the Similarity Ratio Matrix, Distance Function Matrix, and possibly the R, H, and T values be deposited in a data bank. My translation of the Nemalionales section is on deposit with the Translations Center of the John Crerar Library, 35 W. 33rd Street, Chicago, Illinois 60616. The publication of the characters and character states could be considered as an appendix to the manuscript.

Other criticism will be incorporated into the paper as necessary. At this point I feel that my obligation to the study is nearing completion.

Sincerely,

Mike

Encl.

M. H. Levine

Digitized by the Hunt Institute for Botanical Documentation

THE MEMALIONALES : A COMPARISON OF TAXIDIPTRIC AND CONVENTIONAL TAXONOMIC RESULTS

by M.H. Levin

Department of Botany
University of Manitoba
Winnipeg, Manitoba, Canada

Digitized by the Hunt Institute for Botanical Documentation

INTRODUCTION

The research reported in this paper is an effort to analyse the results obtained by one taxonomist by using his monographic conclusions as a basis. In the process I have been able to demonstrate which characters he used and, by using one method, I have been able to demonstrate that his conclusions, the genera listed, can for the most part be 'visualized' by a digital computer. Although any monograph could have been employed, the interest in marine biology suggested that a study of an algal group be carried out. Preference was given to a monograph of a class of algae rather than to a regional floristic or species treatment. A monograph of the Red Algae (Kylin, 1956) was chosen for analysis. The Nemalionales an order containing 453 species in 39 genera of 6 families was selected for study (Table 1).

METHODS

The Operational Taxonomic Units (OTU) employed in this study were 38 genera and the two sections of the genus Lemanea.

The first step involved the translation, from German to English, of the generic descriptions. Although taxonomic description is very precise certain idiomatic constructions defied further simplication. These were listed as they translated, and if sufficient difference in meaning was detected they were retained intact as character states.

After completing the translation of generic description these were compared in order to decide which characters could be subjected to further

---

Present Address:

Digitized by the Hunt Institute for Botanical Documentation

analysis. The list that resulted (Appendix 1) suggested the following breakdown of characters and character states (Table 2). All of the characters suggested by the monographer were used.

The character states utilized in the running of the program agree closely with those suggested by Sokal and Sneath (1963). Four different types of character states were recognized (Table 3). The first of these was a mutually exclusive one such as the example of incrustation. Here, incrustation of the thallus qualifies as being either uncalcified or calcified. The second type is one such as the character monosporangia arrangement with character states that overlap. The third type is represented by a character with a single character state as, for example, the twisting by some members and either not possessed or no information can be derived for other members. The fourth type of character state is not listed in Table 3. This is one in which a combination of exclusive and partial overlap appear under the same character.

The program used in this analysis was first described by Rogers and Tanimoto (1960) and Rogers and Fleming (1963, 1964). The program was reconstructed from their summary of procedure and adapted to the Univac 1107 computer. This program gives equal weight to each character state, calculates similarity ratio (Appendix 1), converts this to a distance function which defines semimetric space, forms a centroid determination, distance array, and finally clusters the objects. The computer is given the values of three constants; Beta, an absolute value of inhomogeneity (.00430), Delta jump (.04400) and EN threshold distance (.50000). These constants were set at the same level as in the work which successfully separated the species of the green algal genus Halimeda (Rogers and Fleming, 1964).

Digitized by the Hunt Institute for Botanical Documentation

TABLE I

SUMMARY OF THE TAXA OF THE NEMALIONALES, NUMBER OF SPECIES, AND INFINITE DISTANCES AS REPRESENTED IN THE DISTANCE FUNCTION MATRIX

	TAXON (OTU, OBJECT)	NO. OF SPECIES	INFINITE DISTANCE FROM
I.	CHANTRANSIACEAE	166	16,17,18,26,27,28,32
	1. Audouinella	5	18,26,27,28,32
	2. Kylinia	35	
	3. Acrochaetium	100	
	4. Liagorophila	1	
	5. Grania	1	16,17
	6. Rhodochorton	20	16,17
	7. Colaconema	3	
	8. Conchocelis	1	
II.	BATRACHOSPERMACEAE	50	26,27,28
	9. Batrachospermum	50	26,27,28
	10. Sirodotia	5	26,27,28
	11. Nothocladus	2	26,27,28
	12. Tuomeya	1	26,27,28
	13. Scarteladia	1	26,27,28
III.	LEMANEACEAE	15	18,19,20,21,22,23,24,25 29,30,31,32
	Lemanea	15	
	14. Section Sacheria		As enumerated above
	15. Section Lemanea		
IV.	THOREACAE	8	5,6,33,34,35,36,38,39,40
	16. Thorea	6	5,6
	17. Nemalionopsis	2	33,34,35,36,38,39,40,5,6
V.	HELMINTHOCLADIACEAE	88	1,14,15
	18. Nemalion	10	1,14,15
	19. Trichoclea	2	14,15
	20. Helminthora	7	14,15
	21. Helminthocladia	5	14,15
	22. Liagora	60	14,15
	23. Liagoropsis	1	14,15
	24. Cumopsis	1	14,15
	25. Dermonema	2	14,15
VI.	CHAETANGIACEAE	93	1,9,10,11,12,13,14,15
	26. Scinaia	10	1,9,10,11,12,13
	27. Gloiophloea	8	1,9,10,11,12,13
	28. Pseudoscinaia	2	1,9,10,11,12,13
	29. Chaetangium	1	14,15
	30. Actinotrichia	1	14,15
	31. Galaxaura	70	14,15
	32. Rhizophyllis	1	1,14,15
VII.	NACCARIACEAE	4	17
	33. Acrotylopora	1	17
	34. Neonaccaria	1	17
	35. Naccaria	2	17
VIII.	DUMONTIACEAE	20	17
	36. Dumontiacea	4	17
	37. Asparagopsis	4	17
	38. Delisea	7	17
	39. Cystoclonium	3	17
	40. Leptophyllis	1	17

Digitized by the Hunt Institute for Botanical Documentation

TABLE 2

SUMMARY OF CHARACTER GROUPINGS IN THE RED ALGAE

	Major Character Groupings	Number of Characters	Number of Character States
I.	Habitat, Incrustation, Sex Distribution	4	12
II.	Thallus	39	152
III.	Monosporangia	4	13
IV.	Spermatangia	7	36
V.	Carpogonium	9	43
VI.	Gonimoblast	11	32
VII.	Cystocarp	8	13
VIII.	Tetrasporangia	3	11
	Totals	85	312

Digitized by the Hunt Institute for Botanical Documentation

TABLE 3

TYPES OF CHARACTER STATES IN THE RED ALGAE

Category	Characters and Character States
Exclusive	Incrustation of the Thallus
	108 uncalcified
	109 calcified
Partial Overlap	Monosporangia arrangement
	334 often 2-3 together
	335 2-6 together
Independent	Ribbon shaped chromatophore
	324 spirally twisted

Digitized by the Hunt Institute for Botanical Documentation

RESULTS AND CONCLUSIONS

The initial printout was the similarity ratio matrix and the distance function matrix. No similarity, hence infinite distance, was found between certain taxa. These are listed in Table 1.

The next step in the program involved the calculation of the R, H, and T values (Table 4). The R value indicates the number of times that an object (genus or section) has at least one attribute in common with another object. An object with a high R value is more typical as it shares more character states with the whole group of objects. The H value is the sum of the rows of the distance matrix. This expresses the sum of the distance of each object to every other object. The highest ranking object would have the most attributes in common. The T value was calculated by dividing R into H: ($\frac{H}{R} = T$). The results of the T ranking demonstrated that the genus Ptilonia was first followed closely by objects 38 (Dalissea) and 37 (Asparagoopsis). Ptilonia could be considered as 'most typical' of the order. The two sections of the genus Lenanea ranked last in the T ranking and this suggested that they could be considered least typical of the order. It should be noted in the case of Lenanea that only the characteristics and attributes of the sections were used. A set which included the characters of the genus was not constructed.

The program continued through the cluster analysis and this only clustered the following objects: 38 (Dalissea), with 14 and 15 the sections Sacheria and Lenanea respectively (Table 5,7). Objects 9 and 10, the genera Sirodotia and Baryachospermum were clustered, (Table 6), and also objects 26, 27, 28 Scinaia, Gloiophloea, Pseudoscinaia (Table 7,8). All other objects were designated as separate entities forming clusters individually.

Certain objects were infinitely distant from others. Figure 1. For example, three genera of the Chaetangiaceae (objects 26, 27, 28) were distant

Digitized by the Hunt Institute for Botanical Documentation

Table 4 : R, H, AND T VALUES

R	H	T	T RANK	OBJECT
38.00000	65.67418	1.72827	1	39
38.00000	71.42172	1.87952	2	38
39.00000	81.31983	2.08512	3	37
38.00000	80.78991	2.12605	4	34
38.00000	83.97000	2.20974	5	40
37.00000	85.80705	2.31911	6	29
38.00000	89.14705	2.34597	7	33
38.00000	90.61016	2.38448	8	36
37.00000	88.22800	2.38454	9	30
38.00000	93.08295	2.44955	10	35
37.00000	94.87806	2.56427	11	19
37.00000	95.01934	2.56809	12	20
39.00000	101.01962	2.59025	13	2
37.00000	97.36828	2.63158	14	31
39.00000	103.32861	2.64945	15	8
39.00000	103.42547	2.65194	16	7
37.00000	98.31814	2.65725	17	21
39.00000	103.63955	2.65742	18	3
36.00000	95.67145	2.65754	19	13
37.00000	100.59105	2.71868	20	16
39.00000	106.09060	2.72027	21	4
36.00000	99.27339	2.75759	22	12
37.00000	102.09871	2.75942	23	22
37.00000	103.38360	2.79415	24	24
37.00000	103.61493	2.80040	25	25
36.00000	101.47866	2.81885	26	32
36.00000	101.60157	2.82227	27	9
37.00000	106.48625	2.87801	28	23
36.00000	103.81043	2.88362	29	11
36.00000	104.38917	2.89970	30	18
37.00000	109.15011	2.95000	31	6
36.00000	107.86617	2.99628	32	10
37.00000	111.27777	3.00751	33	5
34.00000	118.06731	3.47257	34	1
33.00000	116.31391	3.52466	35	26
33.00000	116.50829	3.53055	36	27
33.00000	116.50829	3.53055	37	28
30.00000	131.32489	4.37750	38	17
27.00000	127.13767	4.70880	39	14
27.00000	127.13767	4.70880	40	15

Digitized by the Hunt Institute for Botanical Documentation

from the five genera of the Batrachospermaceae. All 8 genera of the Helminthocladiaceae were distant from the two clustered sections of the genus Lemanea. The seven genera of the Chaetangiaceae were split and distant relationship between two families was shown. Finally, a single genus and its distant relationship from another one, two, or three genera was shown (Table 1).

In practice then, the program did not fully return the family groupings. To have done this it would have been necessary to show that all genera of each family were distantly related to all genera of at least one other family; all possibilities of eight families being used in the process.

The clustering of the genera Batrachospermum and Sirodotia is of interest in that Sirodotia is a genus erected by Kylin. The difference between these is based chiefly upon the organization and location of the goninoblast filaments.

The clustering of the two sections of the genus Lemanea is a function of equal distances and the same R, H, and T values though the character states are exactly opposite in each section.

Last, the clustering of Scinaia, Gloiophloea, and Pseudoscinaia, objects 26, 27, 28, was inspected (Table 7,8). Scinaia and Pseudoscinaia do not differ in their generic descriptions, however, the goninoblasts and surrounding covering are well separated in Scinaia and Gloiophloea and these are grown together in Pseudoscinaia. These attributes appear in the family key. Gloiophloea differs from Scinaia chiefly in the nature of large hyaline cells and their separation from epidermal cells by short cortical filaments. Based upon the generic descriptions, however, it is not possible to separate Scinaia and Gloiophloea. The single feature which separates them is a character used in the key to the Chaetangiaceae.

Digitized by the Hunt Institute for Botanical Documentation

It becomes clear that in dealing with the genera of the Nemalionales that the monographer turned attention to gonimoblast characters as the final source of separation when all else failed.

In some ways there are parallels between 'picture formation' and the delimitation of genera. Specimens are examined and the easily observed information is drawn from this. If the genera form a sharp separation one need list other characters primarily for the sake of completeness. If the members of a family do not form mutually exclusive categories then the search for a separating character must continue.

Most of the descriptions in this monograph enable the researcher to arrive at a clear-cut choice between genera. The gaps in the descriptions are in the no information categories, the phase of the life cycle or the structure that was not demonstrable either because it was missing from the sheet or unable to be observed adequately. The presence of this information would allow a further basis to separate or combine the various genera.

If it is assumed that the taxonomic treatment in the Nemalionales was applied consistently to the other orders of the Rhodophyceae, and by cursory inspection this appears to be true, then Kylin's monograph should become the starting point for new taxonomic studies.

Digitized by the Hunt Institute for Botanical Documentation

TABLE 5

COMPARISON OF SECTIONS OF LEMANEA

The designations in this and other tables are: + (positive or yes)
- (negative or no)
NC (no information)

Characters and Character States	Sect. Sacheria	Sect. Lemanea
1. Singly or sparingly branched	+	-
2. Often profusely branched	-	+
3. Spermatangia building separate protuberances	+	-
4. Spermatangial protuberances building a ring about the thallus	-	+
5. Central axis without down-running cell filaments	+	-
6. Central axis with down-running cell filaments	-	+
7. Carpogonial branch 2-4 celled	+	-
8. Carpogonial branch 5-7 celled	-	+

Digitized by the Hunt Institute for Botanical Documentation

TABLE 6

COMPARISON OF BATRACHOSPERMUM AND SIRODOTIA

	Characters and Character States	Batrachospermum	Sirodotia
1.	Thallus upright to 10 cm high, profusely branched on all sides, gelatinous	+	+
2.	Slippery	+	NC
3.	Central filaments much thicker than the outer filaments; each articulate cell of the central filament with a whorl of branched short shoots	+	+
4.	Moniliform habit	+	NC
5.	Carpogonial branch consisting of the lower part of the short shoot	+	+
6.	On 3 - 5 cells	+	-
7.	On 2 - 5 cells	-	+
8.	Gonimoblasts with terminal carposporas	+	+
9.	Gonimoblasts with limited organization	+	-
10.	Gonimoblasts enclosed by assimilatory filaments	+	-
11.	Gonimoblasts going out from between the assimilatory branches	-	+

Digitized by the Hunt Institute for Botanical Documentation

TABLE 7

COMPARISON OF SCINAIA, GLOIOPHLOEA, PSEUDOSCINAIA

	Characters and Character States	Scinaia	Gloiophloea	Pseudoscinaia
1.	Thallus terete or somewhat flattened	+	+	+
2.	Bifurcate branches and of equal height	+	+	+
3.	Core filaments crowded together in a rather thin central strand	+	+	+
4.	Inner cortex loosened	+	+	+
5.	Outer cortex thickly crowded together with a surface layer of compact, grown together, large hyaline cells	+	-	+
6.	Large hyaline cells becoming intermingled with short-celled cell rows, ultimately becoming crushed	-	+	-
7.	Conimoblasts sunken in the thallus, of thickly entangled thin filaments, in the shape of fruit cases, and discharging by a pore	+	+	+
8.	Carpospores arranged in rows	+	NC	NC
9.	Spermatangia scattered on the upper thallus surface	+	+	+
10.	Monosporangia occasionally present, scattered on upper thallus surface	+	NC	NC
11.	Carpospore filaments grown together with gonimoblast cover	NC	NC	+

Digitized by the Hunt Institute for Botanical Documentation

TABLE 8

COMPARISON OF DIFFERENCES BETWEEN
CLUSTERED TAXA OF NEMALIONALES

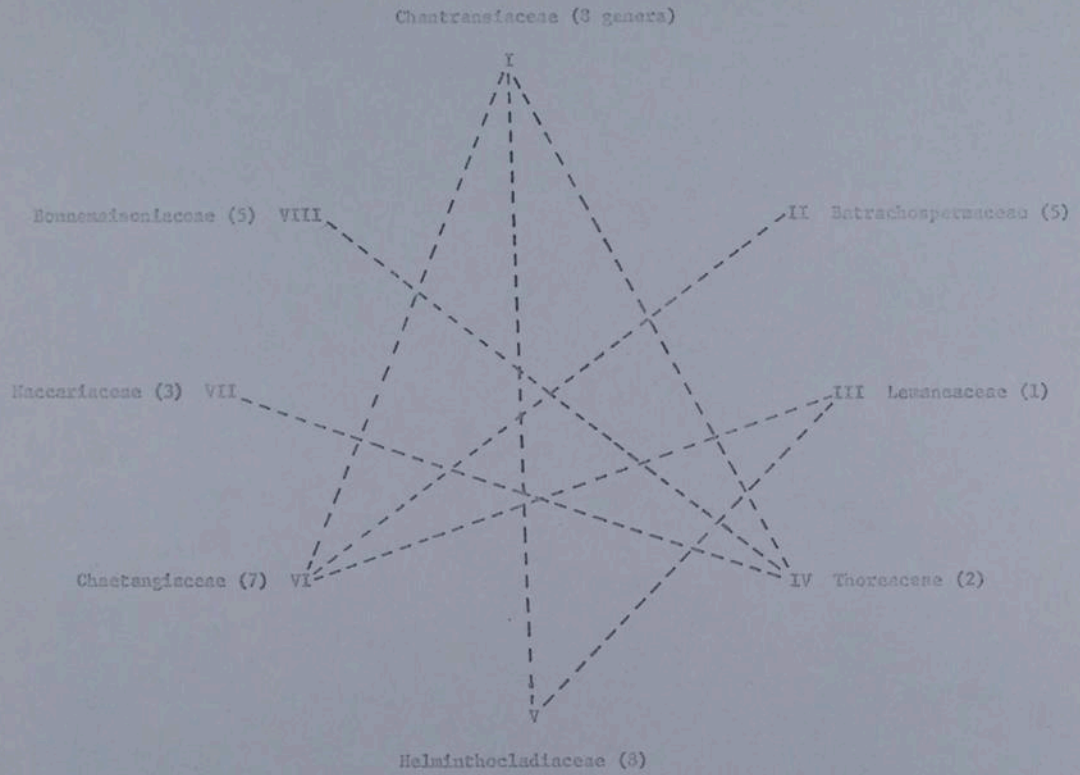
	Characters	Batrachospermum	Sirodotia
1.	Carpogonial branch	3-5 cells	2-5 cells
2.	Conimoblasts	limited organization	unlimited organization
3.	Conimoblasts	enclosed	enclosed then going out

	Characters	Scinaia	Gloiophloea	Pseudoscinaia
1.	Outer cortex	thickly crowded together	not thickly crowded together	thickly crowded together
2.	Surface layer of outer cortex	compact, grown together, large hyaline cells	not thickly crowded together	compact, grown together, large hyaline cells
3.	Large hyaline cells	not intermingled, and not becoming crushed	becoming intermingled with short-celled cell rows, ultimately becoming crushed	not intermingled, and not becoming crushed

Digitized by the Hunt Institute for Botanical Documentation

FIGURE 1

REPRESENTATION OF INFINITE DISTANCES
AS SHOWN IN THE DISTANCE FUNCTION MATRIX



Digitized by the Hunt Institute for Botanical Documentation

LITERATURE CITED

Kylin, H.(arald). 1956. Die Gattungen der Rhodophyceen. CWK Gleerups Förlag, Lund, Schweden. xv and 673 p.

Rogers, D.J. and T.T. Tanimoto. 1960. A computer program for classifying plants. Science 132(3434): 1115-1118.

Rogers, D.J. and H. Fleming. 1963. Data recording: an inexpensive and efficient method. AIBS Bulletin 13(2): 42.

Rogers, D.J. and H. Fleming. 1964. A computer program for classifying plants. II. A numerical handling of non-numerical data. BioScience 14(9): 15-28.

Sokal, R.R. and P.H.A. Sneath. 1963. Principles of numerical taxonomy. W.H. Freeman and Company, San Francisco, Calif. xvi + 359 p.

Digitized by the Hunt Institute for Botanical Documentation

CHARACTERS AND CHARACTER STATES

IN THE

NEMALIONALES

I. Habitat and Miscellaneous

1. Habitat
   - 101 fresh water
   - 102 salt water

2. Thallus location (4)
   - 103 epiphytic or exceptionally endophytic
   - 104 wholly endophytic or endozoic
   - 105 in calcareous shells
   - 106 endophytic in Bonnemaisonia or Asparagopsis.
   - 107 endophytic in Liagora orientalis

3. Incrustation of the thallus (1)
   - 108 uncalcified
   - 109 calcified

4. Sex distribution
   - 110 monoecious or dioecious
   - 111 dioecious
   - 112 monoecious

II. Thallus

5. Thallus orientation
   - 113 generally upright, occasionally prostrate and creeping
   - 114 creeping
   - 115 upright

6. Thallus height
   - 116 nearly 2 mm.
   - 117 to 5 mm.
   - 118 2-4 mm.
   - 119 2-5 mm.
   - 120 to 5 mm., exceptionally to 25 mm.
   - 121 to 5 cm.
   - 122 3-10 cm.
   - 123 to 10 cm.
   - 124 10-20 cm.
   - 125 10-40 cm.
   - 126 5-20 cm.
   - 127 15-30 cm. and up to 1.5 cm. wide
   - 128 3-7 cm.
   - 129 2-10 cm.
   - 130 about 5 cm.
   - 131 3-25 cm.
   - 132 to 15 cm.
   - 133 4-8 cm.
   - 134 10-15 cm.
   - 135 8-15 cm.
   - 136 7-20 cm.
   - 137 5-30 cm.
   - 138 20-50 cm.

Digitized by the Hunt Institute for Botanical Documentation

7. Thallus branching
- 139 scarcely
- 140 irregular
- 141 on all sides
- 142 moderately strong on all sides to more dichotomous
- 143 profusely on all sides
- 144 singly or sparingly branched
- 145 often profusely branched
- 146 irregularly on all sides
- 147 singly or dichotomously
- 148 dichotomously or on all sides
- 149 mostly profusely branched on edges
- 150 unbranched or sparingly branched
- 151 profusely dichotomously branched and rather compactly
- 152 dichotomous, equally high
- 153 mostly dichotomous, occasionally heavily proliferating on the edge and flat side, in the inner thick or somewhat loosened, or inflated and holbw tubular and bladder shaped.
- 154 dichotomous, above transversely ringed by whorls of short thickly tangled articulated deciduous hairs
- 155 dichotomous, now and then constricted articulated, loosened in the inner, occasionally nearly tubular hollow.
- 156 profusely repeatedly branched on all sides
- 157 two (bilateral) rowed alternating, individual shoots two rowed alternating fringed
- 158 profusely branched on all sides with very much ramified branches
- 159 pinnately branched

8. Thallus midrib
- 160 occasionally with the indication of a midrib
- 161 occasionally with weakly developed midrib
- 162 beneath with weakly developed midrib

9. Thallus border
- 163 alternating ciliate or denticulate
- 164 more or less saw-toothed
- 165 indistinctly saw-toothed

10. Clasping organ
- 166 single cell
- 167 cell layer

11. Thallus cross section
- 168 terete
- 169 terete or flattened
- 170 flattened, in uppermost thallus parts cylindrical
- 201 leaflike flat, irregularly lobed
- 202 tubular, terete, or slightly compressed
- 203 two-edged flattened

Digitized by the Hunt Institute for Botanical Documentation

12. Long Shoots
    204 in abundant quantity by spreading

13. Short Shoots
    205 bordered

14. Shoot constriction
    206 all moniliform constricted
    207 not moniliform constricted
    208 wavy constricted
    209 strictly, not moniliform

15. Short shoot systems
    210 grown together
    211 not grown together

16. Shoot articulate
    212 by whorls of completely freely filamentous branched short shoots
    213 symmetrical, bordered by branched assimilatory filaments embedded in jelly
    214 thallus tip with thick tangled branching assimilatory filaments which rather soon fall and are replaced with a thick, dense, many layered thallus cover.

17. Branching of filaments
    215 dichotomously in upper part of thallus

18. Thallus organization
    216 uniaxial
    217 multiaxial

19. Uniaxial construction
    218 monosiphonous
    219 central filament
    220 filaments [fitting neither category]

20. Central filament thickness
    221 many times thicker than outer cortical filaments
    222 thinner than heaviest cortical filaments, those in fully developed parts thicker than central filaments
    223 central filaments distinct, much stronger than extended cell filaments
    224 central axis in the young shoots thin, surrounded by thin along running cell filaments. Central filaments in older shoots built up of large cells
    225 central filaments thin
    226 central filament strongly developed
    227 slightly developed
    228 central axis enclosed by an inward large celled, outward universal ground like small celled cortex without distinct row arrangement of cells

Digitized by the Hunt Institute for Botanical Documentation

Central axis
229 without down running cell filaments
230 with down running cell filaments

Core (medulla filaments) bundles
231 strongly relaxed

Assimilatory filament consistency
232 embedded in jelly (mucilage)--jelly-like slippery
233 not embedded in jelly (mucilage)--slippery
234 gelatinous
235 strongly gelatinous
236 rather firm

Assimilatory filaments
237 parenchymatous
238 not parenchymatous

Assimilatory filament union
239 united with one another

Building of parenchymatous cortical tissue by (from) short shoot systems
240 present
241 absent

Fertile branches
242 build no crowded cortex cover

Internal anatomy (core, medulla)
243 inner branched thin filaments rather loosely interlaced to a central strand
244 inner filaments large celled, combined solidly to a central strand
245 inner filaments rather thickly combined to a central strand
246 core (medulla) filaments sparingly branched about 4-9 μ. thick
247 central tissue with loose longitudinal and transverse filaments
248 core (medulla) filaments crowded together to a a rather thin central strand
249 core filaments crowded together to a thick central strand, accompanying numerous rhizoids
250 central strand of the core filaments is more or less strongly loosened

Internal anatomy (cortex)
251 outer cortical layer compounded from repeatedly dichotomously branched assimilation filaments
252 cortical cells repeatedly dichotomously branched
253 cortical filaments unbranched or dichotomously branched, externally expanded often umbrella shaped

Digitized by the Hunt Institute for Botanical Documentation

Internal anatomy (cortex) (continued)

254 assimilatory filaments once dichotomous, upper with short densely crowded branches, lower about 4 μ--upper 15 μ thick

255 cortical filaments dichotomously branched

256 inner cortex of large cells. Outer cortex of 3-4 celled rather loosely crowded together assimilatory filaments

257 inner cortex being loosened, outer cortex thickly crowded together with a surface layer of compact grown together large hyaline (epidermal) cells

258 layer of large hyaline cells becoming intermingled little by little with short celled cell rows --ultimately become crushed. Large epidermal cells are separated from one another by short celled cell filaments

259 central and inner cortex finely filamentous braided through with longitudinal and transversely running rhizoids more or less profusely braided through. Outer cortex being thick, short celled, anticlinally arranged with thickly tangled thin cortical filaments

260 inner cortex profusely braided through by rhizoids. Filaments of the rather thin outer cortex repeatedly dichotomous short celled and laterally thickly crowded together

261 cortex 3 layered as in the flowering plants, in the case of the tetraspore bearing plants mostly 4 layered with the outer cell layers loosened. The upper surface cells of the cortex not rarely developed to short 1-2 celled hairs or growing out to longer articulating hairs, now and then entirely without hairs

262 Core covering of longitudinally running cell filaments which bend out toward the periphery, itself di- and trichotomously branched and whose apical cells build the one layered outer cortex of the shoots; between these longitudinal filaments run an abundant layer of thin rhizoids which build up the bulk of the shoots

30. Parietal layer (of thallus ="cortex-like")

263 larger quantity of roundish cells and a layer of shorter (smaller cells)

264 several layers; inner layer with shorter (smaller) cells

31. Cortex density

265 very crowded

266 cortex thickly grown together with central filaments

32. Terminal cells of cortex

267 not enlarged

268 strongly enlarged

269 reduced

Digitized by the Hunt Institute for Botanical Documentation

33 Apical cell
270 terminating in a single celled hair
301 hair formation absent

34 Apical growth
302 by transversely articulated apical cell and by short disk-shaped cells
303 by 2-sided alternating obliquely articulated apical cell
304 apical cell transversely articulated

35 Thallus cell thickness
305 nearly 16 u thick varying in upper, middle, lower parts of the thallus.

36 Thallus cell shape
306 irregular
307 of different forms (elongated cylindric to irregular and 2 isodiametric).

37 Pyrenoid occurrence
308 absent
309 present
310 absent or present

38 Pyrenoid location
311 central on chromatophore

39 Chromatophore position
312 parietal (marginal)
313 central or parietal

40 Chromatophore shape
314 ribbon
315 stellate
316 stellate, disc or ribbon
317 ribbon or disc
318 ribbon or stellate

41 Chromatophore quantity
319 one occasionally two
320 one to two, exceptionally three
321 more

42 Chromatophore lobing
322 not rarely irregularly
323 often irregularly

43 Ribbon shaped chromatophore
324 spirally twisted

43 Digitized by the Hunt Institute for Botanical Documentation

III. Monosporangia

44. Monosporangia occurrence
325 present
326 occasionally present

45. Monosporangia location
327 sessile on short clustered lateral
327 stalked or unstalked lateral on long shoots
329 in clusters lateral on vegetative cells
330 in special branch systems with thick short cells
331 terminal on short branches between lower parts of the assimilatory filaments
332 umbelliform position between assimilatory filaments
333 scattered on upper thallus surface

46. Monosporangia arrangement
334 often 2-3 together
335 2-6 together

47. Monospore formation
336 in terminal cells
337 terminal and intercalary cells

IV. Spermatangia

48. Spermatangia position
338 on carpogonial branches or lateral on long shoots
339 on short lateral shoots on carpogonial branch
340 on short laterals or at the end of aggregiate long shoots
341 on short laterals or sessile on carpogonial branches
342 on upper cells of the short shoots
343 on twisted branchings of the short shoots in the inner part of the shoots
344 at top of divided thallus branches
345 spermatangia bearing branch between cortical filaments, above provided with a tuft of short branches
346 on reduced apical cells of the cortical filaments
347 pedestals on the long assimilatory filaments or on the assimilatory upper surface cells or formed on the stalk cells of the assimilatory upper surface cells
348 in short branch clusters on the lowermost part of the assimilatory filaments, sometimes being numerous and thickly entangled at the fertile tips.
349 on stalked oval or catkin shaped branches
350 on stalked, catkin shaped branches
351 pedestals lateral on outer part of cortical filaments
352 pedestals originating in large part in outer cells of the assimilatory filaments

Digitized by the Hunt Institute for Botanical Documentation

Spermatangia position (continued)
353 pedestals thickly crowded on apical cells of the cortical filaments
354 pedestals thickly crowded at the tip of the assimilatory filaments
355 on the outer cells of the corticals
356 pedestals from lower cells of assimilatory filaments

49 Spermatangia organization
357 2-3 together
358 in stalked clusters or 2-3 together
359 single, terminal
360 terminal
361 arranged singly or irregularly verticillately whorled with a short over surface layer
362 building separate protuberances
363 protuberances building a ring about the thallus

50 Spermatangia shape
364 oval

51 Spermatangia pedestals
365 in shape swollen like prominent nemathecia
366 cells each with a single whorl of 3-4 spermatangia
367 twice branched
368 short

52 Spermatangia (pedestal [1 only]) location
369 in special cavities crowded together under upper surface of the thallus
370 spermatangia scattered on upper thallus surface
801 spermatangia sunken in the vegetative short shoots

53 Number of cells in spermatangia pedestal
402 2-4

54 Color of spermatangia pedestal cells
403 colorless

V. Carpogonium

55 Carpogonium location
404 laterally on long shoots
405 on short lateral short shoots
406 consisting of lower part of short shoot
407 mostly of lower basal cells of the short shoots
408 origin from lower cells of assimilatory fils.
409 between the cortical filaments (carpogonial branch with short laterals)
410 terminal on younger lateral branches of the cortex filaments
411 attaching laterally to cortical filaments
412 some one of the middle cells attaching laterally to cortical filaments

55

Digitized by the Hunt Institute for Botanical Documentation

Carpogonium location (continued)
413- terminal or subterminal on the assimilation filaments
414 ondeformed short shoots, aduncate, with a cluster of significant cells from the first and second cells of the carpogonial branch
415 on the basal cell of the secondary short shoots with a cluster of significant cells from the lowermost cell of the carpogonial branch
416 branches in uppermost part of filament

56. Carpogonium position
417 on 2-3 celled stalk
418 unstalked or on single celled stalks
419 unstalked or on 1-2 celled stalks
420 on 3-5 cells
421 on 2-5 cells
422 on 2-4 cells
423 on 5-7 cells
4 424 3 celled
425 2 celled
426 4 called
427 3 rarely 4 celled
428 2-6 celled on the middle of the cortical filaments
429 6-10 celled
430 3 rarely 4-5 celled
431 3 celled, consisting of the large cells of the inner cortex
432 on long distinct branch, consisting of more cells

57 Carpogonial branch twisting
433 somewhat twisted
434 spirally twisted

58 Carpogonial branch
435 ultimately constructing the column of the gominoblasts

59 Carpogonial venter
436 unsymmetrical, supplied with a hemispherical swelling

60 Trichogyne shape
437 cylindrical and short-stalked
438 clavate (club-shaped)

61 First division of carpogonium after fertilization
439 longitudinal
440 transverse
441 twice transversely septate
442 by an oblique wall

62 Primary pit connections
443 carpogonial branch and upper half of stalk cell and gonimoblast become involved in these

Digitized by the Hunt Institute for Botanical Documentation

63 Fertilized carpogonium
444 builds more filaments which are profusely branched
445 splits off a first gonimoblast cell which builds more filaments
446 constructs directly a quantity of sporogenous filaments which crawl around extensively in the cortex without producing organized structures

VI. Gonimoblast

64 Gonimoblast location
447 lateral on long shoots
448 between assimilatory filaments
449 between the short shoots
650 sunken in cortex
451 scattered in cortex
452 crowded together between the rather lengthy assimilatory filaments of the fertile terminal ramifications

65 Gonimoblast position
453 confined in upper part of thallus or scattered on special fertile branches
454 on both sides of upper parts surface shoots

66 Gonimoblast filaments
455 short
456 creeping and going out the carpospore bearing branch
457 thickly entangled thin filaments with some in shape of fruit cases
458 profusely branched growing along between the assimilatory filaments of the central filaments of the tip and producing external sporogenous filaments

67 Gonimoblast development
459 on upper part of constructed cell
460 on either constructed cell

68 Gonimoblast arrangement
461 single compact cluster of thick crowded together filaments
462 hemispherically rounded cluster of thick compact filaments
463 hemispherically rounded cluster of more or less thick compact filaments
[illegible] organized

69 Gonimoblast organization
464 organized
465 unorganized

Digitized by the Hunt Institute for Botanical Documentation

70 Gonimoblast and surrounding covering wall
466 separated from one another
467 grown together with each other

71 Gonimoblast discharge
468 externally by a pore

72 Protective gonimoblast filaments
469 absent
470 X present on carpogonial branch
501 crown of branched cell filaments
502 present or absent
503 without special covering branches, often intermingled with assimilatory filaments
504 enclosed by assimilatory filaments

73 Carpospore position
505 terminal in gonimoblasts
506 in rows
507 singly at joints

74 Number of carpospores in row
508 2-3 in each row

VII. Cystocarp

75 Cystocarp stalk occurrence
509 stalked
510 unstalked

76 Cystocarp stalk length
511 short stalk

77 Cystocarp shape=
512 spherical avoid

78 Cystocarp attachment
513 nearer the border
514 marginal

79 Cystocarp quantity
515 in less quantity on the flattened side of the branch

80 Cystocarp arrangement
516 single or divided
517 single
518 distinctly divided

81 Cystocarp location
519 at same places as the ♂ branches
520 on uppermost thallus branches

82 Fruiting receptacle shape
521 more or less compact

Digitized by the Hunt Institute for Botanical Documentation

VIII. Tetrasporangia

Tetrasporangia occurrence
- 522 present
- 523 rarely present

Tetrasporangia structure
- 524 pair divided
- 525 irregularly 4-parted

Tetrasporangia location
- 526 sessile on short clustered laterals
- 527 on short laterals
- 528 terminal
- 529 terminal or on short laterals
- 530 in wart-like parietal sori on the tip of the thallus branches
- 531 in small sori on the upper thallus parts along both sides of the border above the upper surface
- 532 on the articulated assimilatory filaments on the upper thallus surface

Digitized by the Hunt Institute for Botanical Documentation

SIMILARITY RATIO MATRIX

OBJECT NUMBERS	1	2	3	4	5	6	7	8	9	10
1	1.000	.462	.407	.250	.304	.333	.333	.375	.200	.176
2	.462	1.000	.571	.333	.231	.474	.429	.375	.125	.111
3	.407	.571	1.000	.400	.308	.500	.300	.300	.125	.111
4	.250	.333	.400	1.000	.429	.375	.286	.286	.091	.067
5	.304	.231	.308	.429	1.000	.429	.273	.333	.067	.059
6	.333	.474	.500	.375	.429	1.000	.375	.333	.091	.077
7	.333	.429	.300	.286	.273	.375	1.000	.556	.111	.091
8	.375	.375	.300	.286	.333	.333	.556	1.000	.111	.091
9	.200	.125	.125	.091	.067	.091	.111	.111	1.000	.600
10	.176	.111	.111	.067	.059	.077	.091	.091	.600	1.000
11	.231	.125	.111	.111	.077	.091	.111	.111	.304	.333
12	.231	.125	.111	.111	.077	.091	.111	.111	.333	.300
13	.250	.091	.077	.143	.111	.111	.111	.111	.385	.385
14	.167	.067	.077	.091	.167	.111	.111	.111	.333	.273
15	.167	.067	.077	.091	.167	.111	.111	.111	.333	.273
16	.200	.091	.077	.143	.000	.000	.091	.091	.500	.333
17	.200	.091	.091	.143	.000	.000	.111	.111	.111	.111
18	.000	.077	.125	.091	.091	.091	.111	.111	.176	.053
19	.067	.143	.111	.091	.077	.091	.111	.111	.222	.158
20	.059	.125	.176	.077	.067	.091	.111	.111	.200	.143
21	.059	.125	.111	.077	.143	.091	.111	.111	.143	.143
22	.059	.176	.158	.077	.067	.091	.111	.111	.158	.100
23	.059	.125	.125	.111	.077	.111	.143	.143	.143	.100
24	.059	.125	.125	.091	.067	.091	.111	.111	.100	.200
25	.067	.143	.143	.111	.067	.091	.111	.111	.125	.111
26	.000	.077	.077	.091	.250	.143	.091	.091	.000	.000
27	.000	.077	.077	.091	.250	.143	.091	.091	.000	.000
28	.000	.077	.077	.091	.250	.143	.091	.091	.000	.000
29	.167	.273	.231	.111	.167	.167	.111	.111	.200	.167
30	.200	.231	.250	.111	.200	.200	.111	.111	.200	.167
31	.077	.167	.143	.143	.077	.091	.111	.111	.167	.143
32	.000	.091	.091	.143	.091	.111	.143	.143	.091	.077
33	.111	.176	.176	.167	.143	.167	.200	.200	.200	.238
34	.167	.273	.231	.250	.167	.167	.200	.200	.250	.294
35	.100	.158	.158	.143	.125	.167	.200	.200	.182	.250
36	.091	.200	.167	.333	.200	.200	.250	.250	.231	.231
37	.091	.200	.167	.333	.200	.200	.250	.250	.231	.231
38	.167	.333	.231	.333	.273	.231	.333	.333	.429	.429
39	.250	.429	.333	.333	.429	.333	.333	.333	.429	.429
40	.143	.273	.200	.250	.231	.200	.250	.250	.273	.273

Digitized by the Hunt Institute for Botanical Documentation

11	12	13	14	15	16	17	18	19	20	21	22
.231	.231	.250	.167	.167	.200	.200	.000	.067	.059	.059	.059
.125	.125	.091	.067	.067	.091	.091	.077	.143	.125	.125	.176
.111	.111	.077	.077	.077	.077	.091	.125	.111	.176	.111	.158
.111	.111	.143	.091	.091	.143	.143	.091	.091	.077	.077	.077
.077	.077	.111	.167	.167	.000	.000	.091	.077	.067	.143	.067
.091	.091	.111	.111	.111	.000	.000	.091	.091	.091	.091	.091
.111	.111	.111	.111	.111	.091	.111	.111	.111	.111	.111	.111
.111	.111	.111	.111	.111	.091	.111	.111	.111	.111	.111	.111
.304	.333	.385	.333	.333	.500	.111	.176	.222	.200	.143	.158
.333	.300	.385	.273	.273	.333	.111	.053	.158	.143	.143	.100
1.000	.280	.294	.273	.273	.200	.250	.059	.176	.176	.111	.111
.280	1.000	.571	.273	.273	.429	.143	.077	.231	.231	.143	.125
.294	.571	1.000	.333	.333	.600	.200	.167	.273	.273	.167	.077
.273	.273	.333	1.000	.429	.200	.200	.000	.000	.000	.000	.000
.273	.273	.333	.429	1.000	.200	.200	.000	.000	.000	.000	.000
.200	.429	.600	.200	.200	1.000	.273	.333	.333	.333	.200	.250
.250	.143	.200	.200	.200	.273	1.000	.111	.250	.250	.111	.143
.059	.077	.167	.000	.000	.333	.111	1.000	.407	.357	.259	.214
.176	.231	.273	.000	.000	.333	.250	.407	1.000	.429	.241	.286
.176	.231	.273	.000	.000	.333	.250	.357	.429	1.000	.379	.250
.111	.143	.167	.000	.000	.200	.111	.259	.241	.379	1.000	.241
.111	.125	.077	.000	.000	.250	.143	.214	.286	.250	.241	1.000
.111	.143	.091	.000	.000	.200	.111	.273	.304	.231	.200	.250
.176	.143	.200	.000	.000	.200	.111	.182	.217	.200	.333	.273
.143	.167	.200	.000	.000	.200	.111	.222	.263	.238	.263	.200
.000	.000	.000	.111	.111	.111	.111	.333	.200	.286	.333	.286
.000	.000	.000	.111	.111	.111	.111	.333	.200	.286	.333	.286
.000	.000	.000	.111	.111	.111	.111	.333	.200	.286	.333	.286
.200	.200	.111	.000	.000	.250	.143	.385	.333	.429	.429	.429
.200	.200	.250	.000	.000	.250	.143	.250	.375	.294	.333	.294
.167	.167	.091	.000	.000	.250	.143	.222	.333	.263	.294	.412
.111	.111	.143	.000	.000	.333	.200	.400	.333	.286	.286	.286
.294	.333	.500	.167	.167	.091	.000	.176	.222	.200	.200	.158
.333	.400	.500	.200	.200	.200	.000	.200	.250	.250	.250	.200
.263	.286	.400	.167	.167	.091	.000	.158	.200	.182	.238	.143
.231	.273	.273	.333	.333	.111	.000	.167	.167	.167	.167	.167
.231	.273	.400	.333	.333	.111	.143	.273	.273	.273	.273	.167
.429	.429	.333	.500	.500	.143	.000	.250	.250	.250	.250	.250
.429	.429	.333	.500	.500	.143	.000	.250	.250	.250	.250	.429
.273	.333	.273	.333	.333	.111	.000	.200	.200	.200	.200	.200

Digitized by the Hunt Institute for Botanical Documentation

23	24	25	26	27	28	29	30	31	32	33	34
.059	.059	.067	.000	.000	.000	.167	.200	.077	.000	.111	.167
.125	.125	.143	.077	.077	.077	.273	.231	.167	.091	.176	.273
.125	.125	.143	.077	.077	.077	.231	.250	.143	.091	.176	.231
.111	.091	.111	.091	.091	.091	.111	.111	.143	.143	.167	.250
.077	.067	.067	.250	.250	.250	.167	.200	.077	.091	.143	.167
.111	.091	.091	.143	.143	.143	.167	.200	.091	.111	.167	.167
.143	.111	.111	.091	.091	.091	.111	.111	.111	.143	.200	.200
.143	.111	.111	.091	.091	.091	.111	.111	.111	.143	.200	.200
.143	.100	.125	.000	.000	.000	.200	.200	.167	.091	.200	.250
.100	.200	.111	.000	.000	.000	.167	.167	.143	.077	.238	.294
.111	.176	.143	.000	.000	.000	.200	.200	.167	.111	.294	.333
.143	.143	.167	.000	.000	.000	.200	.200	.167	.111	.333	.400
.091	.200	.200	.000	.000	.000	.111	.250	.091	.143	.500	.500
.000	.000	.000	.111	.111	.111	.000	.000	.000	.000	.167	.200
.000	.000	.000	.111	.111	.111	.000	.000	.000	.000	.167	.200
.200	.200	.200	.111	.111	.111	.250	.250	.250	.333	.091	.200
.111	.111	.111	.111	.111	.111	.143	.143	.143	.200	.000	.000
.273	.182	.222	.333	.333	.333	.385	.250	.222	.400	.176	.200
.304	.217	.263	.200	.200	.200	.333	.375	.333	.333	.222	.250
.231	.200	.238	.286	.286	.286	.429	.294	.263	.286	.200	.250
.200	.333	.263	.333	.333	.333	.429	.333	.294	.286	.200	.250
.250	.273	.200	.286	.286	.286	.429	.294	.412	.286	.158	.200
1.000	.167	.200	.143	.143	.143	.286	.333	.294	.231	.200	.286
.167	1.000	.333	.143	.143	.143	.286	.429	.222	.200	.238	.333
.200	.333	1.000	.143	.143	.143	.286	.429	.222	.200	.222	.286
.143	.143	.143	1.000	.875	.875	.429	.158	.294	.333	.091	.091
.143	.143	.143	.875	1.000	.765	.429	.158	.294	.333	.091	.091
.143	.143	.143	.875	.765	1.000	.429	.158	.294	.333	.091	.091
.286	.286	.286	.429	.429	.429	1.000	.368	.500	.286	.231	.231
.333	.429	.429	.158	.158	.158	.368	1.000	.368	.231	.333	.333
.294	.222	.222	.294	.294	.294	.500	.368	1.000	.333	.231	.231
.231	.200	.200	.333	.333	.333	.286	.231	.333	1.000	.143	.167
.200	.238	.222	.091	.091	.091	.231	.333	.231	.143	1.000	.667
.286	.333	.286	.091	.091	.091	.231	.333	.231	.167	.667	1.000
.130	.273	.200	.077	.077	.077	.231	.333	.200	.143	.407	.500
.167	.167	.167	.111	.111	.111	.167	.167	.143	.250	.333	.333
.167	.273	.273	.111	.111	.111	.167	.273	.143	.250	.455	.455
.250	.250	.250	.200	.200	.200	.333	.231	.250	.250	.500	.500
.250	.250	.250	.200	.200	.200	.333	.333	.250	.250	.500	.500
.200	.200	.200	.143	.143	.143	.273	.200	.200	.250	.333	.333

Digitized by the Hunt Institute for Botanical Documentation

35	36	37	38	39	40
.100	.091	.091	.167	.250	.143
.158	.200	.200	.333	.429	.273
.158	.167	.167	.231	.333	.200
.143	.333	.333	.333	.333	.250
.125	.200	.200	.273	.429	.231
.167	.200	.200	.231	.333	.200
.200	.250	.250	.333	.333	.250
.200	.250	.250	.333	.333	.250
.182	.231	.231	.429	.429	.273
.250	.231	.231	.429	.429	.273
.263	.231	.231	.429	.429	.273
.286	.273	.273	.429	.429	.333
.400	.273	.400	.333	.333	.273
.167	.333	.333	.500	.500	.333
.167	.333	.333	.500	.500	.333
.091	.111	.111	.143	.143	.111
.000	.000	.143	.000	.000	.000
.158	.167	.273	.250	.250	.200
.200	.167	.273	.250	.250	.200
.182	.167	.273	.250	.250	.200
.238	.167	.273	.250	.250	.200
.143	.167	.167	.250	.429	.200
.130	.167	.167	.250	.250	.200
.273	.167	.273	.250	.250	.200
.200	.167	.273	.250	.250	.200
.077	.111	.111	.200	.200	.143
.077	.111	.111	.200	.200	.143
.077	.111	.111	.200	.200	.143
.231	.167	.167	.333	.333	.273
.333	.167	.273	.231	.333	.200
.200	.143	.143	.250	.250	.200
.143	.250	.250	.250	.250	.250
.407	.333	.455	.500	.500	.333
.500	.333	.455	.500	.500	.333
.000	.286	.385	.500	.500	.455
.286	1.000	.444	.333	.556	.333
.385	.444	1.000	.333	.556	.333
.500	.333	.333	1.000	.368	.333
.500	.556	.556	.368	1.000	.412
.455	.333	.333	.333	.412	1.000

Digitized by the Hunt Institute for Botanical Documentation

DISTANCE FUNCTION MATRIX (INFINITE DISTANCE REPRESENTED BY 5.248)

OBJECT	1	2	3	4	5	6	7	8	9	10	11
1	.000	1.115	1.295	2.000	1.716	1.585	1.585	1.415	2.322	2.503	2.115
2	1.115	.000	.807	1.585	2.115	1.078	1.222	1.415	3.000	3.170	3.000
3	1.295	.807	.000	1.322	1.700	1.000	1.737	1.737	3.000	3.170	3.170
4	2.000	1.585	1.322	.000	1.222	1.415	1.807	1.807	3.459	3.907	3.170
5	1.716	2.115	1.700	1.222	.000	1.222	1.874	1.585	3.907	4.087	3.700
6	1.585	1.078	1.000	1.415	1.222	.000	1.415	1.585	3.459	3.700	3.459
7	1.585	1.222	1.737	1.807	1.874	1.415	.000	.848	3.170	3.459	3.170
8	1.415	1.415	1.737	1.807	1.585	1.585	.848	.000	3.170	3.459	3.170
9	2.322	3.000	3.000	3.459	3.907	3.459	3.170	3.170	.000	.737	1.716
10	2.503	3.170	3.170	3.907	4.087	3.700	3.459	3.459	.737	.000	1.585
11	2.115	3.000	3.170	3.170	3.700	3.459	3.170	3.170	1.716	1.585	.000
12	2.115	3.000	3.170	3.170	3.700	3.459	3.170	3.170	1.585	1.737	1.837
13	2.000	3.459	3.700	2.807	3.170	3.170	3.170	3.170	1.379	1.379	1.766
14	2.585	3.907	3.700	3.459	2.585	3.170	3.170	3.170	1.585	1.874	1.874
15	2.585	3.907	3.700	3.459	2.585	3.170	3.170	3.170	1.585	1.874	1.874
16	2.322	3.459	3.700	2.807	5.248	5.248	3.459	3.459	1.000	1.585	2.322
17	2.322	3.459	3.459	2.807	5.248	5.248	3.170	3.170	3.170	3.170	2.000
18	5.248	3.700	3.000	3.459	3.459	3.459	3.170	3.170	2.503	4.248	4.087
19	3.907	2.807	3.170	3.459	3.700	3.459	3.170	3.170	2.170	2.663	2.503
20	4.087	3.000	2.503	3.700	3.907	3.459	3.170	3.170	2.322	2.807	2.503
21	4.087	3.000	3.170	3.700	2.807	3.459	3.170	3.170	2.807	2.807	3.170
22	4.087	2.503	2.663	3.700	3.907	3.459	3.170	3.170	2.663	3.322	3.170
23	4.087	3.000	3.000	3.170	3.700	3.170	2.807	2.807	2.807	3.322	3.170
24	4.087	3.000	3.000	3.459	3.907	3.459	3.170	3.170	3.322	2.322	2.503
25	3.907	2.807	2.807	3.170	3.907	3.459	3.170	3.170	3.000	3.170	2.807
26	5.248	3.700	3.700	3.459	2.000	2.807	3.459	3.459	5.248	5.248	5.248
27	5.248	3.700	3.700	3.459	2.000	2.807	3.459	3.459	5.248	5.248	5.248
28	5.248	3.700	3.700	3.459	2.000	2.807	3.459	3.459	5.248	5.248	5.248
29	2.585	1.874	2.115	3.170	2.585	2.585	3.170	3.170	2.322	2.585	2.322
30	2.322	2.115	2.000	3.170	2.322	2.322	3.170	3.170	2.322	2.585	2.322
31	3.700	2.585	2.807	2.807	3.700	3.459	3.170	3.170	2.585	2.807	2.585
32	5.248	3.459	3.459	2.807	3.459	3.170	2.807	2.807	3.459	3.700	3.170
33	3.170	2.503	2.503	2.585	2.807	2.585	2.322	2.322	2.322	2.070	1.766
34	2.585	1.874	2.115	2.000	2.585	2.585	2.322	2.322	2.000	1.766	1.585
35	3.322	2.663	2.663	2.807	3.000	2.585	2.322	2.322	2.459	2.000	1.926
36	3.459	2.322	2.585	1.585	2.322	2.322	2.000	2.000	2.115	2.115	2.115
37	3.459	2.322	2.585	1.585	2.322	2.322	2.000	2.000	2.115	2.115	2.115
38	2.585	1.585	2.115	1.585	1.874	2.115	1.585	1.585	1.222	1.222	1.222
39	2.000	1.222	1.585	1.585	1.222	1.585	1.585	1.585	1.222	1.222	1.222
40	2.807	1.874	2.322	2.000	2.115	2.322	2.000	2.000	1.874	1.874	1.874

Digitized by the Hunt Institute for Botanical Documentation

12	13	14	15	16	17	18	19	20	21	22
2.115	2.000	2.585	2.585	2.322	2.322	5.248	3.907	4.087	4.087	4.087
3.000	3.459	3.907	3.907	3.459	3.459	3.700	2.807	3.000	3.000	2.503
3.170	3.700	3.700	3.700	3.700	3.459	3.000	3.170	2.503	3.170	2.663
3.170	2.807	3.459	3.459	2.807	2.807	3.459	3.459	3.700	3.700	3.700
3.700	3.170	2.585	2.585	5.248	5.248	3.459	3.700	3.907	2.807	3.907
3.459	3.170	3.170	3.170	5.248	5.248	3.459	3.459	3.459	3.459	3.459
3.170	3.170	3.170	3.170	3.459	3.170	3.170	3.170	3.170	3.170	3.170
3.170	3.170	3.170	3.170	3.459	3.170	3.170	3.170	3.170	3.170	3.170
1.585	1.379	3.585	1.585	1.000	3.170	2.503	2.170	2.322	2.807	2.663
1.737	1.379	1.874	1.874	1.585	3.170	4.248	2.663	2.807	2.807	3.322
1.837	1.766	1.874	1.874	2.322	2.000	4.087	2.503	2.503	3.170	3.170
.000	.807	1.874	1.874	1.222	2.807	3.700	2.115	2.115	2.807	3.000
.807	.000	1.585	1.585	.737	2.322	2.585	1.874	1.874	2.585	3.700
1.874	1.585	.000	1.222	2.322	2.322	5.248	5.248	5.248	5.248	5.248
1.874	1.585	1.222	.000	2.322	2.322	5.248	5.248	5.248	5.248	5.248
1.222	.737	2.322	2.322	.000	1.874	1.585	1.585	1.585	2.322	2.000
2.807	2.322	2.322	2.322	1.874	.000	3.170	2.000	2.000	3.170	2.807
3.700	2.585	5.248	5.248	1.585	3.170	.000	1.295	1.485	1.948	2.222
2.115	1.874	5.248	5.248	1.585	2.000	1.295	.000	1.222	2.051	1.807
2.115	1.874	5.248	5.248	1.585	2.000	1.485	1.222	.000	1.399	2.000
2.807	2.585	5.248	5.248	2.322	3.170	1.948	2.051	1.399	.000	2.051
3.000	3.700	5.248	5.248	2.000	2.807	2.222	1.807	2.000	2.051	.000
2.807	3.459	5.248	5.248	2.322	3.170	1.874	1.716	2.115	2.322	2.000
2.807	2.322	5.248	5.248	2.322	3.170	2.459	2.202	2.322	1.585	1.874
2.585	2.322	5.248	5.248	2.322	3.170	2.170	1.926	2.070	1.926	2.322
5.248	5.248	3.170	3.170	3.170	3.170	1.585	2.322	1.807	1.585	1.807
5.248	5.248	3.170	3.170	3.170	3.170	1.585	2.322	1.807	1.585	1.807
5.248	5.248	3.170	3.170	3.170	3.170	1.585	2.322	1.807	1.585	1.807
2.322	3.170	5.248	5.248	2.000	2.807	1.379	1.585	1.222	1.222	1.222
2.322	2.000	5.248	5.248	2.000	2.807	2.000	1.415	1.766	1.585	1.766
2.585	3.459	5.248	5.248	2.000	2.807	2.170	1.585	1.926	1.766	1.280
3.170	2.807	5.248	5.248	1.585	2.322	1.322	1.585	1.807	1.807	1.807
1.585	1.000	2.585	2.585	3.459	5.248	2.503	2.170	2.322	2.322	2.663
1.322	1.000	2.322	2.322	2.322.	5.248	2.322	2.000	2.000	2.000	2.322
1.807	1.322	2.585	2.585	3.459	5.248	2.663	2.322	2.459	2.070	2.807
1.874	1.874	1.585	1.585	3.170	5.248	2.585	2.585	2.585	2.585	2.585
1.874	1.322	1.585	1.585	3.170	2.807	1.874	1.874	1.874	1.874	2.585
1.222	1.585	1.000	1.000	2.807	5.248	2.000	2.000	2.000	2.000	2.000
1.222	1.585	1.000	1.000	2.807	5.248	2.000	2.000	2.000	2.000	1.222
1.585	1.874	1.585	1.585	3.170	5.248	2.322	2.322	2.322	2.322	2.322

Digitized by the Hunt Institute for Botanical Documentation

23	24	25	26	27	28	29	30	31	32
4.087	4.087	3.907	5.248	5.248	5.248	2.585	2.322	3.700	5.248
3.000	3.000	2.807	3.700	3.700	3.700	1.874	2.115	2.585	3.459
3.000	3.000	2.807	3.700	3.700	3.700	2.115	2.000	2.807	3.459
3.170	3.459	3.170	3.459	3.459	3.459	3.170	3.170	2.807	2.807
3.700	3.907	3.907	2.000	2.000	2.000	2.585	2.322	3.700	3.459
3.170	3.459	3.459	2.807	2.807	2.807	2.585	2.322	3.459	3.170
2.807	3.170	3.170	3.459	3.459	3.459	3.170	3.170	3.170	2.807
2.807	3.170	3.170	3.459	3.459	3.459	3.170	3.170	3.170	2.807
2.807	3.322	3.000	5.248	5.248	5.248	2.322	2.322	2.585	3.459
3.322	2.322	3.170	5.248	5.248	5.248	2.585	2.585	2.807	3.700
3.170	2.503	2.807	5.248	5.248	5.248	2.322	2.322	2.585	3.170
2.807	2.807	2.585	5.248	5.248	5.248	2.322	2.322	2.585	3.170
3.459	2.322	2.322	5.248	5.248	5.248	3.170	2.000	3.459	2.807
5.248	5.248	5.248	3.170	3.170	3.170	5.248	5.248	5.248	5.248
5.248	5.248	5.248	3.170	3.170	3.170	5.248	5.248	5.248	5.248
2.322	2.322	2.322	3.170	3.170	3.170	2.000	2.000	2.000	1.585
3.170	3.170	3.170	3.170	3.170	3.170	2.807	2.807	2.807	2.322
1.874	2.459	2.170	1.585	1.585	1.585	1.379	2.000	2.170	1.322
1.716	2.202	1.926	2.322	2.322	2.322	1.585	1.415	1.585	1.585
2.115	2.322	2.070	1.807	1.807	1.807	1.222	1.766	1.926	1.807
2.322	1.585	1.926	1.585	1.585	1.585	1.222	1.585	1.766	1.807
2.000	1.874	2.322	1.807	1.807	1.807	1.222	1.766	1.280	1.807
.000	2.585	2.322	2.807	2.807	2.807	1.807	1.585	1.766	2.115
2.585	.000	1.585	2.807	2.807	2.807	1.807	1.222	2.170	2.322
2.322	1.585	.000	2.807	2.807	2.807	1.807	1.222	2.170	2.322
2.807	2.807	2.807	.000	.193	.193	1.222	2.663	1.766	1.585
2.807	2.807	2.807	.193	.000	.387	1.222	2.663	1.766	1.585
2.807	2.807	2.807	.193	.387	.000	1.222	2.663	1.766	1.585
1.807	1.807	1.807	1.222	1.222	1.222	.000	1.441	1.000	1.807
1.585	1.222	1.222	2.663	2.663	2.663	1.441	.000	1.441	2.115
1.766	2.170	2.170	1.766	1.766	1.766	1.000	1.441	.000	1.585
2.115	2.322	2.322	1.585	1.585	1.585	1.807	2.115	1.585	.000
2.322	2.070	2.170	3.459	3.459	3.459	2.115	1.585	2.115	2.807
1.807	1.585	1.807	3.459	3.459	3.459	2.115	1.585	2.115	2.585
2.939	1.874	2.322	3.700	3.700	3.700	2.115	1.585	2.322	2.807
2.585	2.585	2.585	3.170	3.170	3.170	2.585	2.585	2.807	2.000
2.585	1.874	1.874	3.170	3.170	3.170	2.585	1.874	2.807	2.000
2.000	2.000	2.000	2.322	2.322	2.322	1.585	2.115	2.000	2.000
2.000	2.000	2.000	2.322	2.322	2.322	1.585	1.585	2.000	2.000
2.322	2.322	2.322	2.807	2.807	2.807	1.874	2.322	2.322	2.000

Digitized by the Hunt Institute for Botanical Documentation

33	34	35	36	37	38	39	40
3.170	2.585	3.322	3.459	3.459	2.585	2.000	2.807
2.503	1.874	2.663	2.322	2.322	1.585	1.222	1.874
2.503	2.115	2.663	2.585	2.585	2.115	1.585	2.322
2.585	2.000	2.807	1.585	1.585	1.585	1.585	2.000
2.807	2.585	3.000	2.322	2.322	1.874	1.222	2.115
2.585	2.585	2.585	2.322	2.322	2.115	1.585	2.322
2.322	2.322	2.322	2.000	2.000	1.585	1.585	2.000
2.322	2.322	2.322	2.000	2.000	1.585	1.585	2.000
2.322	2.000	2.459	2.115	2.115	1.222	1.222	1.874
2.070	1.766	2.000	2.115	2.115	1.222	1.222	1.874
1.766	1.585	1.926	2.115	2.115	1.222	1.222	1.874
1.585	1.322	1.807	1.874	1.874	1.222	1.222	1.585
1.000	1.000	1.322	1.874	1.322	1.585	1.585	1.874
2.585	2.322	2.585	1.585	1.585	1.000	1.000	1.585
2.585	2.322	2.585	1.585	1.585	1.000	1.000	1.585
3.459	2.322	3.459	3.170	3.170	2.807	2.807	3.170
5.248	5.248	5.248	5.248	2.807	5.248	5.248	5.248
2.503	2.322	2.663	2.585	1.874	2.000	2.000	2.322
2.170	2.000	2.322	2.585	1.874	2.000	2.000	2.322
2.322	2.000	2.459	2.585	1.874	2.000	2.000	2.322
2.322	2.000	2.070	2.585	1.874	2.000	2.000	2.322
2.663	2.322	2.807	2.585	2.585	2.000	1.222	2.322
2.322	1.807	2.939	2.585	2.585	2.000	2.000	2.322
2.070	1.585	1.874	2.585	1.874	2.000	2.000	2.322
2.170	1.807	2.322	2.585	1.874	2.000	2.000	2.322
3.459	3.459	3.700	3.170	3.170	2.322	2.322	2.807
3.459	3.459	3.700	3.170	3.170	2.322	2.322	2.807
3.459	3.459	3.700	3.170	3.170	2.322	2.322	2.807
2.115	2.115	2.115	2.585	2.585	1.585	1.585	1.874
1.585	1.585	1.585	2.585	1.874	2.115	1.585	2.322
2.115	2.115	2.322	2.807	2.807	2.000	2.000	2.322
2.807	2.585	2.807	2.000	2.000	2.000	2.000	2.000
.000	.585	1.295	1.585	1.138	1.000	1.000	1.585
.585	.000	1.000	1.585	1.138	1.000	1.000	1.585
1.295	1.000	.000	1.807	1.379	1.000	1.000	1.138
1.585	1.585	1.807	.000	1.170	1.585	.848	1.585
1.138	1.138	1.379	1.170	.000	1.585	.848	1.585
1.000	1.000	1.000	1.585	1.585	.000	1.441	1.585
1.000	1.000	1.000	.848	.848	1.441	.000	1.280
1.585	1.585	1.138	1.585	1.585	1.585	1.280	.000

Digitized by the Hunt Institute for Botanical Documentation

JUN 14 1968

THE UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712

The Department of Botany

June 12, 1968

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

Dear David,

Many thanks for your very helpful letter of June 4th and especially for your willingness to pass along the information given on your NSF rating sheet. I discussed the letter with Dr. Flake and we both agree that there is much to be said for the views covered, and we will take these into consideration as our collaborative efforts progress.

Hopefully we can put our heads together sometime this summer, perhaps best at the AIBS meetings in Ohio although, unless our grant develops, Dr. Flake might have some difficulty in attending these.

Again, many thanks for all of your assistance.

With warmest regards.

Most sincerely yours,

Bill

B. L. Turner

BLT:ja

Digitized by the Hunt Institute for Botanical Documentation

4 June 1968

Dr. Billie Turner, Chairman
Department of Botany
University of Texas
Austin, Texas 78712

Dear Billie:

We have just reviewed yours and Flake's proposal to NSF, and a copy of the "confidential" review is enclosed. I am prompted to write this letter for two reasons: (1) We want to see you funded for this work, and (2) we feel that some aspects of your proposed research might prove infeasible. George Estabrook (the mathematician who has been working on these problems with me since 1964) and I have spend considerable time wondering how we might provide you with a useful back-up for your work. In the spirit of trying to be helpful, the following comments about the proposal are made.

1. Research in the areas of cluster analysis has been active for a number of years now. Much meaningful progress has been achieved (in Kansas and Colorado). We should begin to pass out of the phase of empirical development and "after the fact" evaluation of the several extant methods (on the basis of how well the various computer programs fit some a priori opinions of the taxonomic specialists whose groups have been analyzed), and into the phase of the development and recognition of biological principles involved.

2. The areas of research which are in desperate need of investigation now are the points of interface between biology and the mathematical methodologies employed to link biology to computing machines. Efforts whould be made to abstract principles from biology which may be used as objective yardsticks to evaluate (and improve) mathematical methods. An ideal objective would be to abstract sound enough principles from systematic biology such that these along might constitute the axiomatic base upon which the appropriate mathematical methodologies might be built.

3. Perhaps the most elusive area of interface is the recognition, establishment, evaluation, modification, etc. of characters. Here is where biological information is being carried to the mathematical methodologies. There are presently very few principles governing this procedure and yet we must be confident with this aspect of the interface if we are to ultimately adopt these new methodologies as ones which deserve our confidence.

Digitized by the Hunt Institute for Botanical Documentation

4. I would suggest that the statement of objectives under the heading "Characterization" appearing on pp. 3 and 4 of the Grant Proposal be expanded and made more precise. This will give you a clearer picture of what your objectives are. The statement (p. 4) including "... may not require much investigation to obtain a satisfactory approach..." certainly indicates an optimism on your part - I hope it does not reflect your failure to recognize the importance of this phase of your work. Although a little progress has been made, this merely provides you with a starting place.

5. In my opinion Mayr's "prediction criterion" is a misrepresentation of the problem; it stands too close to the current issues. One needs to stand further from the problem, disentangle oneself from the conventional points of view, view the problem in more generality and bring more imagination to bear. I think that the results (still outstanding) on the first part of the research plan (Characterization) are needed before meaningful work can be done on the second (Prediction). I would strongly urge that the substance (90%) of your effort go into a collaborative effort between yourself (and your students) and your methodologist to shed some light on the very knotty problems of characterization which underlie all considerations of methods further along in the taxonomic process. If you can contribute to the solution of these fundamental problems in characterization you shall have made a profound contribution to the art. Fiddling around with ad hoc classification methods is easy work for methodologists; shedding light on characterization is hard work for biologists. We need the experience, professionalism, and willingness of men such as yourself to address this difficult area.

These points, said out in the cold, may not seem entirely relevant, but they indicate some very fundamental points of view. Perhaps the only way that their relevance can be made meaningful is by face-to-face talks with you and Flake, and perhaps with your graduate students. We do not always come across with our intentions intact until such time as we can be certain that our terminology and yours coincide. Even with a complete rapport on the words used, we need to give some more explanation of our attitudes about the complex interrelations between taxonomist, mathematician and computer programmer. Some of our ideas appear in print, but since they do not all occur in one paper, it is difficult to gain from our published papers what is intended here in this letter.

With this background, let me suggest that we put our heads together some time this coming summer, with sufficient time to gain an insight into our respective problems, and some means of attacking them. We are entering a very significant phase in what I prefer to call taximetrics, having been through a few years of empirical efforts and some more fundamental problem solving techniques.

I trust this suggestion is meaningful to you, for we all need as much intensive effort on the work as we can get.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION

PROPOSAL RATING SHEET

Reviewer David J. Rogers

Proposal No.: B8 2359R
Investigator: TURNER & FLAKE
Institution: U Texas
Please return to: Syst Biol.
if possible by: May 27, 1968

Comments *(Continue on additional sheet if necessary)*

This proposal addresses itself to an important and useful area of plant taxonomy. The work to be performed is of such magnitude that it definitely calls for applications of computers to aid in analysis of the data. The objectives of the proposal indicate that the investigators intend to study the problems involved in establishment of characters of significance (weighting problems) to classification, an important area for taxonomy, and predictive clustering schemes as well. I criticized an earlier proposal (B8-0963R) from the same school, with Flake as the principal investigator, because there was insufficient relation between him (a programming type with engineering training) and the botanists for whom the work was to be done. This new proposal has indicated a much closer tie between the botanist and programmer, and thus has removed this criticism.

The authors indicate that they have good knowledge of activities in the field in their cited references, and will use available work from others in the same type of investigation. Both investigators are well qualified for this work.

Budgetary items seem well in line with the proposed activities and the incorporation of graduate students in the investigation is meritorious.

OVERALL RATING
- ☐ EXCELLENT
- ☐ VERY GOOD
- ☐ GOOD
- ☐ FAIR
- ☐ POOR

Signature of Reviewer:

Date: June 3, 1968

Other suggested reviewers *(optional)*:

NSF Form 9-52 July 1967

"Reviewers' comments are solicited and received in confidence by the National Science Foundation, an agency of the U.S. Government, and will be given maximum protection from disclosure permitted under applicable laws, including 5 U.S.C. 552 (The Freedom of Information Act)."

Digitized by the Hunt Institute for Botanical Documentation

B8 2359 R

Proposal for Support of Research Entitled

NUMERICAL CLASSIFICATION STUDIES FOR SYSTEMATIC BIOLOGY

Submitted to

National Science Foundation

Co-Principal Investigators

R. H. Flake

Department of Electrical Engineering

and

B. L. Turner

Department of Botany

The University of Texas at Austin

Austin, Texas

April 15, 1968

Digitized by the Hunt Institute for Botanical Documentation

Proposal to the

National Science Foundation

from The University of Texas at Austin

for Support of Research Entitled:

Numerical Classification Studies for Systematic Biology

Abstract

A new approach for numerical classification has been developed to aid in biological classification problems, particularly at the infraspecific level or lower. The main goals of the proposed research are to develop numerical methods for assessing the predictive quality of classifications produced by this new approach, and to apply these new techniques to selected systematic problems.

Previous Investigations

Numerous recent attempts (Sneath, 1957; Rogers & Tanimoto, 1960; Sokal & Sneath, 1963; Wirth, Estabrook & Rogers, 1966; Estabrook, 1966; Rubin, 1967; Needham, 1961; Parker-Rhodes, 1961) have been made to utilize the capabilities of the modern high-speed computer to aid in the classification problems associated with the areas of biological taxonomy and information retrieval.

Digitized by the Hunt Institute for Botanical Documentation

The co-principal investigator of this proposal, Prof. R. H. Flake, first became interested in computer-aided classification methods in 1962 while at Washington University (St. Louis) where as a research associate in the computation center he was involved in assisting a graduate student in Botany who was doing research for his doctoral dissertation on a systematic study of genera of the subtribe Oncidiinae of the Orchidaceae. The method of numerical taxonomy proposed by Rogers and Tanimoto (Rogers & Tanimoto, 1960) was employed in that study. While the study of Oncidiinae was being pursued, a research psychiatrist on the staff of the Washington University Medical School proposed a study of psychiatric patient data utilizing the Rogers and Tanimoto approach. The results of both numerical studies were somewhat unsatisfactory. However, the exercise did enhance the investigators' insight into the nature of numerical classification problems. The graduate student completed the requirements for his Ph.D. and joined Rogers' group, then at the New York Botanical Gardens.

The co-principal investigator, Prof. B. L. Turner, became interested in a numerical approach to systematics following the acquisition of large quantities of chemical data relating to populational problems in the genera Juniperus and Baptisia. Because of the complexity and the difficulty encountered in evaluating these data by conventional taxonomic methods, computer assistance seemed essential. During the past two years the co-principal investigators of this

Digitized by the Hunt Institute for Botanical Documentation

proposal in collaboration have developed a new approach for numerical classification that we believe is consistent with the classification methods employed by most taxonomists. A computer program has been developed embodying this numerical classification technique.

This new procedure is designed to assist the biological worker concerned with systematic problems. It is particularly useful in infraspecific studies, especially where the characters differentiating populations are subtle and the multiplicity of character variations makes comparative morphological studies, as well as chemical, without the aid of a computer, most difficult, if not impossible.

A detailed account of this new approach is contained in the preprint in the Appendix which has recently been accepted for publication in the Journal of Theoretical Biology.

Proposed Research

The main goals of the proposed research are to develop methods for assessing the predictive quality of classifications produced by the recently developed approach for numerical classification, and to apply these new techniques to selected systematic problems.

Characterization

The predictive quality of a classification depends on how well the data reflects or characterizes the pertinent biological information. The approach

Digitized by the Hunt Institute for Botanical Documentation

envisioned here for investigating the relative value of data characterization will involve the calculation of the relative "weight" of the coded characters in a set of taxonomic data in order to better inform the biological worker about the character weighting implied by any of a variety of methods of coding. This will make it possible to easily implement a coding process that reflects the biological worker's estimate of the proper a priori character weighting. This is a relatively well-defined aspect of the characterization problem which may not require much investigation to obtain a satisfactory approach, particularly in view of the recent work of Estabrook (Estabrook, 1967), which appears to be relevant here.

Another aspect of the characterization problem involves the identification of the "key characters", or the characters most responsible for a hierarchical structure obtained from application of the numerical classification method to a set of taxonomic data. This information will be valuable to the biological worker in assigning a posteria weight to the taxonomic characters, for re-processing the data if it appears desirable.

The approach envisioned in this portion of the characterization problem involves the dual classification, that is, the numerical classification of the characters themselves, rather than the taxa. Some preliminary investigation with this approach has been carried out, and the results are encouraging.

Prediction

Prediction with reference to a classification refers to two sets of

Digitized by the Hunt Institute for Botanical Documentation

phenomena (Mayr, 1965), previously unused characters and previously unknown taxa. A taxonomist generally works with a small fraction of the total number of characters of an organism. If an entirely new set of characters is tested against a classification and this does not lead to any major revision, then the existing classification is a good one and it is said to have a "high predictive value" in relation to the previously unused characters. Similarly, if newly discovered taxa fit easily into the existing classification, then the classification has a high predictive value.

The best classification is the classification with the most predictive value, and results from using properly characterized data. It is proposed to develop numerical methods for estimating the predictive value of a classification.

In addition, the conceptual investigation of the predictive quality of the new numerical classification scheme will be carried out in parallel with the application of these new computer aided techniques to a number of selected systematic problems.

Systematic Applications

Juniperus. Professor B. L. Turner and Dr. Ernst von Rudloff, Organic Chemist at the Prairie Regional Laboratory, Saskatoon, Saskatchewan, Canada, in collaboration, have compiled an enormous amount of terpenoid data, both qualitative and quantitative, for the two species, J. ashei and J. virginiana. The taxa are readily distinguished by their chemistry, there being approximately

Digitized by the Hunt Institute for Botanical Documentation

60-70 characters which distinguish between the species. Since the species reportedly hybridize over a wide range (in fact this is acclaimed by many workers to be the best documented case for allopatric introgression) it ought to be possible to detect gene flow between the species, if such is occurring. Treatment of the data by preliminary computer programs indicates that little if any hybridization is occuring; (von Rudloff, Irving and Turner, 1967) however, the data do suggest that there is a clinal structuring in the populations from Northeast to Southwest across the eastern United States. An intensive sampling program to test this possiblity has been undertaken during the winter of 1968. These data were collected in the form of samples from specific populational sites (i.e., 10 plants collected from each population, the latter taken at approximately 100 mile intervals in a transect from Austin, Texas to Washington, D.C.). A computer aided analysis of these data should make it possible to detect any existing clinal structuring in the populations.

Thermopsis. This genus is closely related to Baptisia and is considered by Turner (1967) to be the more primitive of the two taxa. It is also more difficult taxonomically in the sense that most of the species are poorly defined and occur as aggregations of isolated populations mostly in the higher mountainous ranges of the western United States and eastern Asia. It differs from Baptisia in that present-day hybridization among the species does not seem to be a factor in the variation of the group; instead the taxa appear to be maintained by vegetative means usually in the form of small colonies. Thermopsis is known to have some

Digitized by the Hunt Institute for Botanical Documentation

degree of polyploidy (unlike Baptisia) and this, combined with the subtle variation among the isolated populations, makes the genus an attractive one to study from a numerical standpoint. The species of Thermopsis are currently being studied in detail for their flavonoid chemistry and, for the purposes of this program, are being collected for morphological study as well. These data will be obtained from populational samples in sets of 10 from any one site and the chemical data will be taken from the same plants from which the morphological data are assembled. (This is a bit of vision that was not employed in our Baptisia study, since at the inception of our work on this later genus it was not realized how much chemical data could be assembled and how difficult it might be to evaluate these data by computer methods.) For each populational sample a statistic will be available which can be put into a computer program for classificatory purposes. It is possible that such a treatment might yield ambiguous results since it is already clear that the pod characters, if heavily weighted in a classificatory scheme, serve to delineate the two genera as presently understood. It should prove interesting to treat these data using a strictly phenetic approach as opposed to a phyletic approach. The latter has been discussed by Sokal and Camen (1965) and the program here envisioned for Thermopsis and Baptisia should make possible an assessment of the two numerical methods. In fact the results of such a study might suggest relationships unsuggested by current publications; for example using all the data, the species of the two genera might sort themselves into meaningful systematic groupings which cut across the currently accepted generic lines.

Baptisia. The program(s) developed will also be applied to a number of

Digitized by the Hunt Institute for Botanical Documentation

systematic problems involving the genus Baptisia. This taxon has been studied intensively by comparative phytochemists at the Biochemical Systematic Laboratories in the Department of Botany and there exists a plethora of chemical data both at the specific and infraspecific level. For example, all 16 species of the genus have been carefully surveyed for flavonoids and approximately 80 compounds have been structurally determined. Two of the species, B. nuttalliana (Horne, 1965) and B. leucophaea (Brehm, 1962) have been studied populationally so that infraspecific variability from over a broad geographical region might be studied. Preliminary treatment of some of the B. nuttalliana data (almost 40 flavonoid characters varying in 20 states among 640 plants collected by random sampling from throughout the range of the species; Horne, 1965) has revealed interesting correlations (referred to in the first appendix). In addition, it is hoped that a program can be developed to treat what is perhaps the most intensively studied natural hybrid swarm to date, over 1000 plants of B. leucophaea and B. sphaerocarpa and their hybrid derivatives in a 10-acre tract, each plant tagged and studied for its flavonoid components (Baetcke, 1966). By a computer treatment of the approximately 50 variables in this population it should be possible to detect the extent of hybridization and backcrossing, since the species concerned do not share common flavonoids.

Other problems. In addition to the treatment of the chemical data from Juniperus, Thermopsis and Baptisia, the programs will be applied to morphological data both separately and in combination with the chemical. The computer program

Digitized by the Hunt Institute for Botanical Documentation

will also be available to other workers who might wish to evaluate their own data. For example, Dr. Donald E. Stone of Duke University has essentially the same kind of quantitative data for the fatty acids of the genus Carya and he has been collaborating with Dr. Robert H. Flake for the treatment of these data. Needless to say, these numerical techniques will also be available to a number of graduate students working with a wide range of organisms and characters, be they morphological or chemical, and it is anticipated that the program, if supported, will be a continuing one in which students interested in systematics can obtain experience with this approach. For example, Mr. Robert Adams is currently working on a doctoral thesis under Professor B.L. Turner on a numerical study of Juniperus populations in western Texas. Mr. Adams entered the field of systematics with a major in mathematics and computer science. He has since become interested in applying this approach to biological systematics and in this connection is accumulating terpenoid data from gas chromatographic studies for a systematic evaluation. In addition he is strongly motivated in the direction of conceptual approaches to populational problems similar to those discussed by Womble (1951). In fact, Mr. Adams intends to use "trend analysis" in his own work, an approach in which the total variables are related to geography, climate, etc. through appropriate contour mapping.

To emphasize the academic aspects of the program at the graduate level we have appended the doctoral program of Mr. Adams so that the reviewers might evaluate the credentials of those students likely to become deeply involved in

Digitized by the Hunt Institute for Botanical Documentation

the program. It is envisioned that candidates can enter this field from either Engineering and/or Biology; additional course work will be prescribed to provide for adequate background knowledge in whichever field the candidate is most deficient.

Digitized by the Hunt Institute for Botanical Documentation

PROPOSED PROGRAM OF WORK FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Robert Phillip Adams — 1720-N Barton Springs Rd., Austin, Texas

B.A., June 1962, University of Texas — Major: Mathematics — Minor: Physics

Foreign Language: French, ETS Examination, Summer 1966.

MAJOR: BIOLOGICAL SCIENCES

Course	Course Description	Professor	Institution	Year	Grade
Bot 321	General Botany II	Bold	U. of Texas	1966	A
Bot 362	Plant Systematics	Johnston	U. of Texas	1966	A
Bot 385M	Palynology and Paleobotany	Larson	U. of Texas	1966	A
Bot 394.4	The Cell in Growth and Development	Whaley	U. of Texas	1966	Cr
Bot 392K	Advanced Topics in Plant Sci.	Turner	U. of Texas	1967	A
Bot 382	Advanced Study and Research	Turner	U. of Texas	1967	A
Bot 393	Plant Ecology	McMillan	U. of Texas	In progress	
Bot 372K	Secondary Compounds in Biological Processes	Mabry	U. of Texas	To be taken	
Zoo 425	Introduction to Genetics	Lee	U. of Texas	1967	B
Zoo 382L.10	Population Genetics	Kojima	U. of Texas	In progress	
Zoo 382L.7	Evolution	Stone	U. of Texas	To be taken	
Bot 699	Dissertation	Turner	U. of Texas	To be taken	

Approximate Title: Chemosystematic and Numerical Study of Juniperus.

SUPPORTING WORK

Math 114	Linear Algebra and Matrix Theory	Karlson	Stanford U.	1962	B
Stat 116	Theory of Probability and Statistics	Parson	Stanford U.	1962	B
Bot 364K	General Physiology	Myers	U. of Texas	1966	A
Chem 810a	Organic Chemistry	Bailey	U. of Texas	1966	A
Chem 810b	Organic Chemistry	Henze	U. of Texas	1966	B
Chem 369	General Biochemistry	Eakin	U. of Texas	1967	A
Chem 370	General Biochemistry	Eakin	U. of Texas	1967	A
Zoo 381K	Biometrics	Kojima	U. of Texas	1967	A

Digitized by the Hunt Institute for Botanical Documentation

REFERENCES

BAETCKE, K. P. (1966). "The Structure of an Interspecific Hybrid Population of Baptisia leucophaea and B. sphaerocarpa Based on the Use of Chemical Markers." Doctoral Dissertation, The University of Texas, Austin, Texas.

BELLMAN, R., KALABA, R. and ZADEH, L. (1966). "Abstraction and Pattern Classification." J.Math. Anal. & Appl. 13, 1.

BREHM, K. P. (1962). "The Distribution of Alkaloids, Free Amino Acids, Flavonoids, and Certain Other Phenolic Compounds in Baptisia leucophaea Nutt. var. laevicaulis Gray, and their Taxonomic Implication." Doctoral Dissertation, The University of Texas, Austin, Texas.

CHARLES, D. R., and GOODWIN, R. H. (1943). "An Estimate of the Minimal Number of Genes Differentiating Two Species of Goldenrod with Respect to Other Morphological Characters." Am. Natural 77, 53.

CLAUSIN, J. & HIESEY, W. W. (1958). "Experimental Studies on the Nature of Species." (IV) Genetic Structure of Ecological Races. Carnegie Inst. Wash. Publ. 615.

ESTABROOK, G. F. (1966). "A Mathematical Model in Graph Theory for Biological Classification." J. Theoret. Biol. 12, 297.

ESTABROOK, G. F. (1967). "An Information Theory Model for Character Analysis." Taxon 16, 86.

FARRIS, J. S. (1966). "Estimation of Conservatism of Characters by Constancy within Biological Populations." Evolution. 20, 587.

FLAKE, R. H., et al. (in press). "Populational Analysis by Methods of Numerical Systematics."

HORNE, D. B. (1965). "The Distribution of Flavonoids in Baptisia nuttaliana and B. lanceolata and their Taxonomic Implications." Doctoral Dissertation. The University of Texas, Austin, Texas.

LANCE, G. H. and WILLIAMS, W. T. (1966). "A Generalized Sorting Strategy For Computer Classifications." Nature. 212, 218.

MAYR, ERNST. (1965). "Numerical Phenetics and Taxonomic Theory." Syst. Zool. 14, 73.

Digitized by the Hunt Institute for Botanical Documentation

NEEDHAM, R. M. (1961). " A Method for Using Computers in Information Classification," Report M. L. 152, Issued by Cambridge Language Research Unit, Cambridge, England.

PARKER-RHODES, A. F. (1861). "Contributions to the Theory of Clumps," Report M. L. 138, Issued by Cambridge Language Research Unit, Cambridge, England.

ROGERS, D. J. and TANIMOTO, T. T. (1960). "A Computer Program For Classifying Plants." Science 132, 1115.

RUBIN, J. (1967). "Optimal Classification into Groups: An Approach For Solving the Taxonomy Problem." J. Theoret. Biol. 15, 103.

SNEATH, P. H. A. (1957). "The Application of Computers to Taxonomy." J. Gen. Microbiol. 17, 201.

SOKAL, R. R., and CAMDEN, J. H. (1965). "Principles of Numerical Taxonomy," San Francisco: W. H. Freeman.

TURNER, B. L. (1967). "Plant Chemosystematics and Phylogeny." J. Pure & Appl. Chem. 14, 189.

VON RUDLOFF, IRVING and TURNER, (1967). "Re-Evaluation of Allopatric Introgression Between Juniperus ashei and J. virginiana using Gas Chromatography," Am. J. Bot. 54, 600 (Abstract).

WIRTH, M., ESTABROOK, G. and ROGERS, D. J. (1966). "A Graph Theory Model for Systematic Biology, with an Example for the Oncidiinae (Orchidaceae)." Syst. Zool. 15, 89.

WOMBLE, W. H. (1951). "Differential Systematics." Science 144, 315-322.

Digitized by the Hunt Institute for Botanical Documentation

Equipment

The only significant item of equipment necessary for the proposed study is the computer. A Control Data Corporation 6600 Computer located in the University of Texas Computation Center is available for this research.

The CDC 6600 system consists of the following major items: one central processor, ten peripheral and control processors, large central memory, display console and keyboard, magnetic tapes, card equipment, printers, and card punch. The 6600 computer executes on the average over 3 million instructions per second. The memory consists of 131,072 words of rapid access, magnetic core storage located in 32 independent banks. The central processor has ten arithmetic and logic units to add, multiply, divide, long add, shift, branch, Boolean, and increment; 24 operating registers; and has floating-point and fixed-point arithmetic. The peripheral and control processors have 4095 words of magnetic core storage; 12 input/output channels; and the following instructions: logical, branch, input/output, central processor access, and central memory access.

Initial software for the CDC 6600 computer system is scientifically oriented, with FORTRAN as the basic language.

Period of Performance

It is recommended that the proposed grant, if awarded, be made effective for a period of two years, beginning September 1, 1968, or some other mutually

Digitized by the Hunt Institute for Botanical Documentation

agreeable date.

Current Support and Pending Applications

The co-principal investigators of this proposal do not have any current grant support for research and do not have any other application pending. This proposal is not being submitted to any other agency.

Nonsolicitation Statement

Offeror represents that it has not employed or retained any company or person (other than a full-time bona fide employee working solely for the offeror) to solicit or secure this grant and that it has not paid or agreed to pay any company or person (other than a full-time bona fide employee working solely for the offeror) any fee, commission, percentage, or brokerage fee, contingent upon or resulting from the award of this grant, and agrees to furnish information relating thereto as requested by the contracting officer.

Digitized by the Hunt Institute for Botanical Documentation

Personnel

Co-principal Investigator:	Robert H. Flake
Born:	Little Rock, Arkansas, March 31, 1934
Education:	D.Sc., Washington University, St. Louis, Missouri, June, 1962 Major: Electrical Engineering (Automatic Control), Minor: Physics A.M., Physics, Washington University, January 1960 B. S., Engineering Physics, Washington University, June 1956
Experience:	Assistant Professor of Engineering Science at Washington University (St. Louis) from June 1962 to September 1964; Assistant Professor of Electrical Engineering at The University of Texas since January 1966. In these capacities taught undergraduate and graduate level courses in automatic control, conducted research in automatic control theory and in the applications of computers to various taxonomic problems in medicine and biology. Senior Development Engineer and subsequently Consultant for the Aero-Division of Honeywell, Inc., St. Petersburg, Florida, from September 1964 to April 1967. In this capacity assisted in developing calibration and alignment techniques and navigation computational algorithms for inertial navigational systems.
Professional & Scientific Societies:	Institute of Electrical and Electronics Engineers, American Institute of Aeronautics and Astronautics, International Association of Plant Taxonomists, Sigma Xi.
Current Professional Interests:	Guidance and control theory, numerical classification

Digitized by the Hunt Institute for Botanical Documentation

Publications: "Volterra Series Representation of Time-Varying Nonlinear Systems," Proceedings of the Second International Federation of Automatic Control Congress, Basel, Switzerland, 1963.

"Volterra Series Representation of Nonlinear Systems," Institute of Electrical and Electronics Engineers Transactions, Part II, 1963, pp. 330-335.

(with J. Zaborsky) "On Fitting with Functional Polynomials," Transactions of the American Society of Mechanical Engineers Journal of Basic Engineering, Volume 89, Series D, Number 2, June 1967, pp. 399-405.

(with J. G. Hartwell) "A Reference Orbit Useful in Integrating Between Activity Spheres." American Institute of Aeronautics and Astronautics Journal, Volume 5, Number 3, March 1967, pp. 599-600.

(with J. R. Cox) "Filter Design for the Average Response Computer," 19th Annual Southwestern Institute of Electrical and Electronics Engineers Conference Record, April 1967.

(with J. H. Satterfield) "A Computer Classification of Psychiatric Patients," 6th Annual Southeastern Regional Meeting of the Association for Computing Machinery and National Meeting of Society for Biomedical Computing, Volume I (Preprints), June 1967.

(with B. L. Turner) "Numerical Classification for Taxonomic Problems" Journal of Theoretical Biology (Accepted for Publication March 15, 1968).

Digitized by the Hunt Institute for Botanical Documentation

Co-principal Investigator:	B. L. Turner
Born:	Yoakum, Texas, 1925
Service:	4 years, Navigator, Air Force (1st Lt.) 1943-1947
Education:	B. S., Sul Ross State (Alpine, Texas) 1947-1949 M.S., Southern Methodist University, 1949-1950 Ph.D., Washington State University 1950-1953
Experience:	Associate Investigator on an Ecological Study of Africa, sponsored by the Office of Naval Research and the University of Arizona 1956-1957 (on leave from The University of Texas) Instructor to Professor and Director of Herbarium, The University of Texas 1953-1967; Chairman, Department of Botany, 1967-- National Science Foundation Senior Postdoctoral Fellow and Visiting Professor, University of Liverpool, England, 1965-66.
Honors, etc.:	Phi Beta Kappa; Sigma Xi; Secretary, Botanical Society of America 1959-1964; President, Southwestern Association of Naturalists 1967; Membership in 10 or more American and International Societies, serving on various committees in several; New York Botanical Garden Award for Outstanding Contributions to American Botany, 1965; Council, American Society of Plant Taxonomists, 1967.
Committees:	Invited (and continuing) member of an International ad hoc Committee on Chemotaxonomy, sponsored by the International Union of Pure and Applied Chemistry. Meetings held at Kyoto, Japan, 1964; Stockholm, Sweden, 1966. Also formal member of the Committee on Chemotaxonomy of the International Association of Plant Taxonomists. Chairman, Program Committee on Systematic Botany, XI International Botanical Congress, Seattle, 1969.

Digitized by the Hunt Institute for Botanical Documentation

Also, various university committees, including the Committee to form a Division of Biological Sciences at The University of Texas and The Molecular Sciences Committee under the NSF sponsored Developmental Program.

Current Professional Interests:

Biochemical Systematics
Numerical Systematics
Chromosome Numbers in Compositae and Leguminosae, particularly as Related to Phyletic Arrangements.
Cytogenetical Studies in the Helenieae
Vegetation of the World on a Climatic Basis
Desert Vegetation of the World

Publications:

1950

Vegetative key to Texas Desmanthus (Leguminosae) and similar genera. Field & Lab. 18: 51-54.

Texas species of Desmanthus (Leguminosae). Field & Lab. 18: 54-65.

Mexican species of Desmanthus (Leguminosae). Field & Lab. 18: 119-130.

1951

Revision of the United States Species of Neptunia (Leguminosae). Field & Lab. 46: 82-92.

1953

Chromosome complements in Desmanthus (Leguminosae). Field & Lab. 21: 47-50.

A taxonomic study of the genus Chamaechaenactis Rydb. (Compositae). Madrono 12: 97-103. (with S. J. Preece)

1955

Chromosome numbers in the genus Sesbania (Leguminosae): Evidence for a conservative treatment. Rhodora 57: 213-218.

Digitized by the Hunt Institute for Botanical Documentation

The Cassia fasciculata complex (Leguminosae) in Texas. Field & Lab. 23: 87-91.

1956

Chromosome numbers in the Leguminosae. I. Am. J. Bot. 43: 577-581.

A cytotaxonomic study of the genus Hymenopappus (Compositae). Rhodora 58: 163-308. (Doctoral thesis)

Chromosome numbers and geographic distribution of Lindheimera, Engelmannia and Berlandiera (Compositae) Southwestern Naturalist 3: 125-132. (with M. C. Johnston)

1957

The chromosomal and distributional relationships of Lupinus texensis and L. subcarnosus (Leguminosae). Madrono 14: 13-16.

1958

Chromosome numbers in the genus Krameria: Evidence for familial status. Rhodora 60: 101-106.

1959

Chromosome numbers in the Leguminosae. II. African species, including phyletic interpretations. Am. J. Bot. 46: 49-57. (with O.S. Fearing)

Segregation and recombination of chemical constituents in a hybrid swarm of Baptisia laevicaulis x B. viridis and their taxonomic implications. Am. J. Bot. 46: 678-686. (with R. E. Alston)

Application of paper chromatography to systematics: Recombination of parental biochemical components in a Baptisia hybrid population. Nature 184: 285-286. (with R. E. Alston)

1960

Chromosomal relationships and taxonomic considerations in the genus Cassia. Am. J. Bot. 47: 309-318. (with H. S. Irwin)

Digitized by the Hunt Institute for Botanical Documentation

Distribution of canavanine in the family Leguminosae as related to phyletic groupings. Can J. Bot. 38: 499-505. (with B. A. Birdsong and R. E. Alston)

The basic chromosome number of the genus Neptunia. Madrono 15: 184-187. (with O. S. Fearing)

Chromosome numbers in the Leguminosae. III. Species of the southwestern United States and Mexico. Am. J. Bot. 47: 603-608. (with O. S. Fearing)

Phytogeographic reconnaissance: the western segment of the Michoacan coast. In Coalcoman and Motines del Oro (ed. D. D. Brand) Martinus Niihoff, The Hague, pp. 272-286.

1961

Chromosome numbers of some Brazilian Leguminosae. Rhodora 63: 16-19. (with H. S. Irwin)

Chromosome numbers in the Compositae. IV. North American species, with phyletic interpretations. Am. J. Bot. 48: 216-223. (with W. L. Ellison and R. M. King)

Chromosome numbers in the Compositae. V. Mexican and Guatemalan species. Rhodora 63: 121-129. (with J. H. Beaman and H.F.L. Rock)

"Operational niches" and "community-interaction values" as determined from in vitro studies of some soil algae. Evolution 15: 228-238. (with Bruce C. Parker)

A comparison of the ninhydrin positive and phenolic substances of parotoid gland secretions of certain Bufo species and their hybrids. Evolution 15: 352-359. (with Don Hunsaker, R. E. Alston and W. F. Blair)

Chromosome numbers in the Compositae. III. Certain Mexican species. Brittonia 13: 64-69. (with M. C. Johnston)

1962

Taxonomy of Hymenothrix (Helenieae, Compositae). Brittonia 14: 101-120.

New techniques in analysis of complex natural hybridization. Proc. Natl. Acad. Sci. 48: 130-137. (with R. E. Alston)

Digitized by the Hunt Institute for Botanical Documentation

Chromosome numbers in Mexican and Guatemalan Compositae. Rhodora 64: 271-276. (with J. Beaman)

A cytotaxonomic survey of Melampodium (Compositae-Heliantheae). Am. J. Bot. 49: 263-269. (with R. M. King)

Chromatographic validation of two morphologically similar hybrids of different origin. Science 137: 1048-1950. (with R. E. Alston, R. N. Lester and D. Horne)

Methods of presentation of crude biochemical data for systematic purposes, with particular reference to the genus Bahia (Compositae). Am. J. Bot. 49: 599-604. (with W. L. Ellison and R. E. Alston)

Chromosome numbers of Dyssodia (Compositae) and phyletic interpretations. Rhodora 64: 2-15. (with M. C. Johnston)

A biosystematic study of the Phlox cuspidata-Phlox drummondii complex. Am. Midl. Nat. 67: 257-281. (with L. Erbe)

1963

Natural hybridization among four species of Baptisia (Leguminosae). Am. J. Bot. 50: 159-173. (with R. E. Alston)

Taxonomy of Florestina (Compositae). Brittonia 15: 27-46.

Perspectives in chemotaxonomy. Science 142: 545-552. (with R. E. Alston and T. J. Mabry)

Cytophyletic analysis of Hymenoxys odorata: a recapitulation. Madrono 17: 77-79.

The betacyanins and their distribution. Phytochem. 2: 61-64. (with T. J. Mabry and A. Taylor)

1964

Chemical investigations of the Batidaceae. Taxon 13: 197-200. (with T. J. Mabry)

Partition chromatography as applied to taxonomic problems in the Asteraceae. Taxon 13: 11-14. (with T. J. Mabry)

Digitized by the Hunt Institute for Botanical Documentation

Chromosome numbers in the Compositae. VIII. Mexican and Central American species. Southw. Nat. 9: 27-39.

Chromosome numbers in the Compositae. VII. Additional species from the southwestern United States and Mexico. Madrono, 17: 128-140. (with A. M. Powell)

Taxonomy of Machaeranthera sect. Psilactis (Compositae--Astereae). Brittonia 15: 316-331. (with D. Horne)

A taxonomic study of the genus Amphicarpaea (Leguminosae). The Soutw. Nat. 9: 207-218. (with O. S. Fearing)

1965

Serological studies in Baptisia and certain other genera of the Leguminosae. Am. J. Bot. 52(2): 165-172. (with R. E. Alston and R. N. Lester)

Chromosome numbers in the Compositae. IX. African species. Jour. South African Bot. 31(3): 207-217. (with W. H. Lewis)

Taxonomy of Stenocarpha (Compositae-Heliantheae-Galinsoginae). The Southw. Nat. 10(4): 238-240.

1966

Chromosomal numbers in the Compositae. X. North American species. Am. J. Bot. 53: 24-33. (with D. L. Flyr)

The use of disc electrophoresis of plant proteins in systematics. Taxon 15: 135-143. (with D. Boulter and D. Thurman)

Pollen wall ultrastructure and its bearing on the systematic position of Blennosperma and Crocidium (Compositae). Am. J. Bot. 53: 555-563.

The taxonomic status of Stemmatella (Compositae-Helinatheae). Madrono 18: 203-206.

Taxonomy of Eutetras (Compositae-Heliantheae). The Southw. Nat. 11(1): 118-122.

The biochemical basis of taxonomy. The Science Teacher 32(9). (with T. J. Mabry and R. E. Alston)

Digitized by the Hunt Institute for Botanical Documentation

Chromosome numbers in Stackhousia (Stackhousiaceae). Australian Jour. Bot. 14: 165-166.

Taxonomy of Oxypappus (Compositae). SIDA 2(6): 431-433.

Waitzia conica (Compositae-Inuleae), a new species from Western Australia. SIDA 2(6): 428-430.

1967

Systematic implications from electron microscopic studies of Compositae pollen--a review. Ann. Mo. Bot. Gard. 53: 220-256. (with J. Skvarla)

Plants: Chemistry and Systematics. Science 155(3758): 63. (book review)

Electrophoretic mobilities of formic and glutamic dehydrogenases in the Fabaceae: a systematic survey. New Phytol. 66: 37-45. (with D. A. Thurman, D. Boulter and E. Derbyshire)

Distribution of canavanine in the plant kingdom. Phytochem. 6: 863-866. (with J. B. Harborne)

Plant chemosystematics and phylogeny. Pure and Applied Chem. 14: 189-213.

Systematic significance of the lupine alkaloids with particular reference to Baptisia (Leguminosae). Evolution 21(3): 508-517.

Chromosome numbers in Compositae. XI. Peruvian species. Ann. Mo. Bot. Gard. 54(2): 172-177. (with A. M. Powell and J. Cuatrecacas)

Chemosystematics: present and future applications. Bull. Nat. Inst. Sci. of India, No. 34: 190-211.

Chemotaxonomy. Chapter in the Encyclopedia of Biochemistry (eds) Roger J. Williams and Edwin M. Lansford, Jr. Reinhold Publ. Corp. 226-228. (with T. J. Mabry and R. E. Alston)

Chromosome survey of Podolepis (Compositae-Inuleae). Aust. J. Bot. 15: 445-449.

Digitized by the Hunt Institute for Botanical Documentation

Estimated Budget for First Twelve Months
(Sept. 1, 1968 through Aug. 31, 1969)

			NSF Grant	Grantee Share
A.	Salaries and Wages			
	1. Senior Personnel			
	a. 2 Co-Principal Investigators (co-principal investigator, R. H. Flake, two summer months @ 9 mo. academic rate of $13,000)	3 1/2 Summer Months $2,889	$ 6,639	
	(co-principal investigator, B. L. Turner, 1 1/2 summer months @ 9 mo. academic rate of $22,500)	3,750		
	2. Other Personnel			
	d. 1 - Graduate Student (Res. Sci. Asst. III, 1/2 time-9 mos., full-time 3 summer mos. @ full-time rate of $565 per month)	4 1/2 Academic Mos. 3 Summer Mos.	4,242	
	f. 1 - Pre-Baccalaureate Student (Approximately 650 hours @ 1.70/hr. for data coding and key punching)	3 Calendar Mos.	1,100	
		Total Salaries and Wages	$11,981	
B.	Fringe Benefits		286	
		Total of A & B	$12,267	
D.	Expendable Supplies (Computer supplies, magnetic tapes, tab cards and paper)		250	
E.	Travel-Domestic (See Budget Notes)		800	
F.	Publication Costs		800	
G.	Other Costs			
	1. CDC-6600-Computer Charges (three hours @ $200 per hour)		600	
H.	Total Direct Costs		$14,717	
I.	Indirect Costs			
	1. On Campus 40.5% of salaries and wages		4,852	
J.	Total Costs		$19,569	
	Less Grantee contribution for indirect costs (excess over 20% of TDC exlcusive of Computer charges-$600)		-2,029	$2,029
Estimated Budget for First Twelve Months			$17,540	$2,029

Digitized by the Hunt Institute for Botanical Documentation

Estimated Budget for Second Twelve Months
(Sept. 1, 1969 through Aug. 31, 1970)

			NSF Grant	Grantee Share
A.	Salaries and Wages			
	1. Senior Personnel			
	a. 2 Co-Principal Investigators (co-principal investigator, R. H. Flake, two summer mos. @ 9 mo. academic rate of $14,000)	3 1/2 Summer Months $3,111	$ 7,111	
	(co-principal investigator, B.L. Turner, 1 1/2 summer months @ 9 mo. academic rate of $24,000)	4,000		
	2. Other Personnel			
	d. 2 - Graduate Students	9 Academic Mos. 6 Summer Mos.	8,682	
	2 - Res. Sci. Asst. III(1-1/2 time-9 mos., full-time 3 summer mos. @ full-time rate of $592/mo.) (1-1/2 time-9 mos., full-time 3 summer mos. @ full-time rate of $565/mo.)			
	f. 1 - Pre-Baccalaureate Student (Approximately 650 hrs. @ 1.70/hr. for data coding and key punching)	3 Calendar Mos.	1,100	
	Total Salaries and Wages		$16,893	
B.	Fringe Benefits		490	
	Total of A & B		$17,383	
D.	Expendable Supplies (Computer supplies, magnetic tapes tab cards and paper)		250	
E.	Travel - Domestic (See Budget Notes)		1,000	
F.	Publication Costs		1,000	
G.	Other Costs			
	1. CDC-6600-Computer Charges (three hours @ $200 per hour)		600	
H.	Total Direct Costs		$20,233	
I.	Indirect Costs			
	1. On Campus 40.5% of salaries and wages		6,846	
J.	Total Costs		$27,079	
	Less Grantee contribution for indirect costs (excess over 20% of TDC exclusive of computer charges - $600)		-2,919	$2,919
Estimated Budget for Second Twelve Months			$24,160	$2,919
Estimated Budget for Twenty-four Months			$41,700	$4,948

Digitized by the Hunt Institute for Botanical Documentation

ESTIMATED BUDGET-NOTES:

A. Salaries and Wages

Salaries paid from grant funds at The University of Texas at Austin conform to the rates approved by the Board of Regents for Salaries paid from regular University funds.

B. Fringe Benefits

Federal Social Security (4.4% of 1st $7800 of each salary; 4.8% after January 1, 1969); Workmen's Compensation Insurance (0.50% of applicable salaries) coverage, as may be required by University regulations under the University's self-insuring plan. Under University accounting procedures these items are direct costs.

E. 1. Domestic Travel

For attendance of one technical meeting and one trip to Washington, D. C., to discuss progress of research with NSF. Transportation expense including transportation by plane or rail and Pullman, or by private car at an allowance of 10¢ per mile in lieu of actual expense, plus per diem at not to exceed $17.00 per day on the quarter-day basis.

Digitized by the Hunt Institute for Botanical Documentation

Administrative Information

Further information relating to the technical portions of this proposal may be obtained from Robert H. Flake, Department of Electrical Engineering, The University of Texas at Austin, telephone: Area Code 512, GR 1 - 1014 or B. L. Turner, Department of Botany, The University of Texas at Austin, Area Code 512, GR 1 - 5262. Contractual matters should be referred to the University's Office of Sponsored Projects, Post Office Box 7726, Austin, Texas 78712, telephone Area Code 512, GReenwood 1 - 1353.

Submitted by:

ORIGINAL SIGNED BY

Robert H. Flake
Assistant Professor
Department of Electrical Engineering

and

ORIGINAL SIGNED BY

B. L. Turner
Chairman
Department of Botany

APPROVED:

ORIGINAL SIGNED BY
JENS M. JACOBSEN

Jens Jacobsen
Executive Director
Office of Sponsored Projects

ORIGINAL SIGNED BY

A. W. Straiton
Chairman
Department of Electrical Engineering

I certify that the distribution of costs between the direct and indirect categories as shown in the proposal conforms to the usual accounting practices of the institution for all Federally supported or sponsored research, and to the distribution used by the cognizant Federal audit agency.

ORIGINAL SIGNED BY

James H. Colvin
Vice-President for Business Affairs

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX

NUMERICAL CLASSIFICATION FOR TAXONOMIC PROBLEMS

R. H. Flake

Department of Electrical Engineering

and

B. L. Turner

Department of Botany

The University of Texas, Austin, Texas, U.S.A.

A new procedure for numerical classification has been developed to aid in biological classification problems, particularly at the infraspecific level or lower. The numerical classification concepts that form the basis of this procedure are identified with biological taxonomic concepts. This numerical procedure appears to be consistent with the classification methods employed by most taxonomists.

1. Introduction

Numerous recent attempts (Sneath, 1957; Rogers & Tanimoto, 1960; Sokal & Sneath, 1963; Wirth, Estabrook & Rogers, 1966; Estabrook, 1966; Rubin, 1967; Needham, 1961; Parker-Rhodes, 1961) have been made to utilize the capabilities of the modern high-speed computer to aid in the classification problems associated with the areas of biological taxonomy and information retrieval.

The central problem in these computer classification procedures is the determination of subgroups, usually referred to as clumps or clusters,

Digitized by the Hunt Institute for Botanical Documentation

from a given set of objects that mutually share a majority of their characteristics, while not sharing many characteristics with the other objects not contained in the subgroups. In some classification procedures, particularly those associated with biological taxonomy, a hierarchy of the subgroups is also obtained.

Most numerical classification procedures rest on algorithms that are classified as either sorting algorithms or clump-searching algorithms. The majority of the procedures developed in the area of numerical taxonomy belong to the sorting algorithm category, whereas the clump-searching algorithms were developed largely for information retrieval classifications.

In the sorting algorithm, the calculation begins by computing a table of similarity measures between all distinct pairs of objects. The hierarchy of subgroups is then generated by a sorting strategy (Lance & Williams, 1966) applied to these measures. These subgroups are thereby defined only in the operational sense that the subgroups obtained are the result of applying the particular sorting strategy on the calculated similarity measures. However, in the clump searching algorithm, a clump is usually defined as a set of objects with associated properties expressed in terms of the object similarity measures, and the clump searching algorithm is developed in order to search efficiently for the clumps of objects represented in the data.

A new numerical classification procedure utilizing a clump-searching algorithm and tailored particularly for biological classification

Digitized by the Hunt Institute for Botanical Documentation

problems is presented in this paper.

A discussion of the principles of biological classification is outlined in the next section. The third section introduces the new numerical classification concepts and discusses their relationship to the biological classification concepts presented in the second section. A computer program has been developed for the proposed numerical classification procedure, and the basic approach underlying the programmed search algorithm is outlined in Appendix II.

2. Principles of Biological Classification

The morphologically oriented taxonomist is often asked to explain his method for arriving at a given decision or group of decisions involved in classification of taxa and their designation as to rank. Most taxonomists answer that organisms are assigned to a category according to the taxonomist's ability to discriminate between them. They are ranked subjectively, based on his best judgment as to what a species or some other category might be. In general, rank is assigned to groups according to the number of features which distinguish between the assemblages. When two or more assemblages are distinguished with difficulty one from the other, they are called varieties or subspecies of a single species; those which are relatively easily distinguished, but clearly related, are called species of the same genus; when two groups of assemblages can be readily recognized and each possesses a few peculiar kinds of characters, these are called genera, and so forth up the scale to where very

Digitized by the Hunt Institute for Botanical Documentation

unusual or wierd assemblages with a combination of many unique features might be called different families, orders, or even divisions.

This may sound non-scientific and rather arbitrary to those who have never attempted the exercise, particularly where the taxa in question are closely related and are sorted in the form of pressed plants or as preserved specimens in jars.

But under natural conditions the problem is not so ill-posed. Features that mark the organisms are easily recognized, variation can be readily evaluated and this related to enviromental features and, if the populations are large, literally thousands of individuals can be examined, albeit superficially, to test the validity of the classified assemblages.

Morphological characters, the basis for most taxononic classifications, are known to be under genetic control. The taxonomist classifies a given population or group of populations as a specific or infraspecific taxon whenever he can detect an assemblage of individuals which seem to share a common gene pool (i.e., individuals which exchange genes among themselves preferentially). The number of genes that differ between two assemblages or gene pools determines the rank that will be assigned those particular taxa.

The determination of the number of genes associated with the character states distinguishing populational assemblages is a laborious and time-consuming task, but this can be approached experimentally and such calculations have been attempted for at least a few taxa

Digitized by the Hunt Institute for Botanical Documentation

(Charles and Goodwin, 1943; Clausen and Hiesey, 1958). Because of the problems inherent in the experimental approach, most taxonomists are content to work on the assumption that genetic similarities or dissimilarities are reflected by morphological similarities or dissimilarities: the more character states shared by two assemblages, the more closely related; the more distant their relationship, the fewer the character states shared.

If the character differences which distinguish between assemblages are indeed a reflection of genetic differences (and most experimental work indicates that this is so), then it should be possible to assign biologically meaningful numerical values to the attributes which characterize these assemblages. These data can then be used in computer programs designed to cluster those individuals or populations (assemblages) which share the largest number of characters (Sokal & Sneath, 1963). Further, as will be indicated below, numerical treatments are useful in that some objective value can be assigned to the distances or gaps between assemblages and, as a result, classifications of a more nearly uniform nature can be achieved.

No new insights into the problem of assignment of rank to an assemblage (or taxon) will be offered in the discussion that follows; for whether two very distinct species are to be recognized as belonging to the same or different genera (or families!) is largely a matter of convenience or personal preference of the taxonomist. Numerical methods cannot resolve such a problem since the assignment of ranks is largely subjective. However, only rarely is the taxonomist faced with this kind

Digitized by the Hunt Institute for Botanical Documentation

of question. More often the problem resolves itself to relating one or more assemblages to other assemblages so that some spatial ordering among the assemblages (or taxa) is obtained. In short, using all the characters available to him, the taxonomist attempts, usually with rapid scanning by eye and mental assimilation of the visual pattern data, to form "natural groups" and relate them one to the other by making multiple comparisons between and among all the individuals or assemblages under consideration. "Natural groups" are biological aggregations of populations or individuals which, when compared on all their characters, are more closely related one to the other than they are to any other group or population of individuals.

Biological aggregations, once recognized as "natural groups", are usually assembled into formal categories on purely phenetic grounds (i.e., without giving weight to the characters selected for study) or on phyletic grounds (where certain characters, usually selected a posteriori, are given more weight than others in arriving at the arrangement). The philosophy behind these two kinds of taxonomies has been tediously explored (Sokal and Camen, 1965) and it need only be noted here that most taxonomists accept phyletic treatments as preferable to those which are purportedly phenetic only (Turner, 1967). Therefore, careful consideration should be given to the matter of assigning character weight.

Among plant taxa in general it is well known that certain characters are genetically more variable than others and that, given a multitude of

Digitized by the Hunt Institute for Botanical Documentation

different climatic regimes, selection can result in a rapid divergence of such characters in selected populations, the taxonomy of which may depend upon these characters. For example, leaf size, shape and texture are sufficiently variable in most taxa so that plant taxonomists tend to weight these characters as less significant to phyletic studies than the characters of the flower. Vegetative features are more plastic than floral features and consequently there is a greater likelihood that rapid divergence or perhaps convergence might lead to mistaken relationships, particularly where different phyletic offshoots contribute taxa to similar environments and superficially similar habits are achieved. In any case, differing spatial arrangements of aggregations can be expected whenever the characters concerned are weighted differently. Thus, a worker who does not weight floral characters as more important than those of the leaf may produce a classification that is different from that worker who weights vegetative characters at the expense of the floral.

The taxonomist, if he chooses to weight characters, must do so according to his experience and judgment. The numerical analysis of such data utilizing computers can produce classification models with greater precision, objectivity and a better evaluation of the recorded data than classifications produced by taxonomists without computer assistance. What should be emphasized here, however, is that the selection of the characters and the selection of the numerical coding procedures (i.e., the assignment of states) for the characters inherently involves assigning a weight to those characters; consequently it is a problem

Digitized by the Hunt Institute for Botanical Documentation

that plagues both phyletic and phenetic workers alike. The computer does not relieve the taxonomist of his responsibility as a taxonomist he must select and evaluate the states he assigns to the characters concerned. In fact it can be said that a numerical classification method will be only as successful as the ability of the worker concerned to intelligently use this new tool; computers will enable good taxonomists to produce better classifications, but poor taxonomists will not profit from this approach.

It is possible that, in the future, numerical procedures will be developed such that the assignment of character weight in a biologically reasonable fashion can be performed by the computer program itself. This is especially likely where statistical data from natural or experimentally grown populations are obtained. For example, if the mean and variance is determined from the measurements of characters from numerous individuals of a given species throughout its range, it should be possible to develop an algorithm which will weight characters in a biologically reasonable fashion according to their variance. This is essentially what most taxonomists do; they subjectively give more weight to characters which have less variation and de-emphasize characters which have a large (seemingly random) variation.

Farris (1966) and others have suggested that the numerical approach, especially where statistical data are available for populations, offers much potential for the resolution of taxonomic problems. This

Digitized by the Hunt Institute for Botanical Documentation

is particularly true at the infraspecific level or lower where the character differences which distinguish populations are subtle and the multiplicity of character variations make the standard morphological approach most difficult if not impossible.

The problem of classification at the specific or higher levels is not so difficult since, as indicated above, the taxonomist can order these taxa with considerable efficiency; what numerical taxonomy adds here is greater precision as to the density (or cohesiveness) of an aggregate and the calculation of a "gap" or "distance" between itself and a related taxon.

We believe that the numerical classification program developed from the concepts presented in the next section is consistent with the taxonomic approach employed by most workers. Nevertheless, the application of this approach will yield meaningful results only if the taxonomist has sufficient knowledge of his data to achieve a system of weighting which reflects the phyletic history of the characters studied.

3. Numerical Classification Concepts

The term numerical classification refers here to a procedure that identifies clusters of similar objects from a given object-attribute (character state) data matrix, and orders these clusters into a hierarchy. In order for this task to be performed with the aid of a computer, the concepts of a cluster and a hierarchy of clusters must first be defined in

Digitized by the Hunt Institute for Botanical Documentation

quantitative terms, and the definitions chosen should be consistent with the goals of the classification.

Quantification of these concepts is generally based on a numerical representation or characterization of the objects to be classified. As previously noted, some form or other of character weighting is inherent in defining a character and in assigning a numerical representation to a character state. Thus, careful consideration regarding the purposes of the classification should be given to the implications of which characters should be utilized in the numerical representation of the objects, and of the relative weight assigned to the characters that are employed. Then, with the numerical characterization of the objects accomplished, a quantitative measure of the number of attributes any object shares with any other object is defined, since this is the basis of the general concept of a cluster of similar objects. Several suitable numerical measures of similarity have been described in the literature (Sokal & Sneath, 1963).

The quantification of a clump-searching classification procedure also requires a definition of a clump or cluster that is associated with the notion of a "group of highly similar objects". The definition chosen should lend itself to the construction of a hierarchical arrangement of such clusters, if the method is to be applied to biological classification problems. One definition of a cluster that leads to the development of a hierarchical arrangement is most naturally described in terms of the notion of a

Digitized by the Hunt Institute for Botanical Documentation

"fuzzy" set (Bellman, Kalaba, & Zadeh, 1966).

In the usual mathematical language, a collection of all elements which satisfy some stated property is called a set. Thus, a given element is either a member of a set or it is not a member according to whether or not it satisfies the property defining the set. This notion of a set has been the basis for defining clusters in previous clump searching numerical classification approaches. However, there are many significant "groups of highly similar objects" in biological classification problems that do not uniformly possess any (known) property or attribute that could be used to define a set. A concept that appears to be more relevant for decribing such "groups" is called a fuzzy set in which a continuous range of levels of membership is permitted.

A fuzzy set A in an arbitrary space of objects X is characterized by a membership function f_A which assigns to each element $x \in X$ some number $f_A(x)$ in the interval $[0,1]$. The value of f_A at x represents the "level of membership" of x in A, which, in general, is intermediate between non-membership and full membership; that is, the nearer the value of $f_A(x)$ to unity, the higher the level of membership of x in A; and, conversely, the nearer the value of $f_A(x)$ to zero, the more tenuous the membership of x in A.

As an extension of this, the notion of a fuzzy class, or fuzzy set whose elements are sets, is a natural way to describe the concept of a

Digitized by the Hunt Institute for Botanical Documentation

"class" of clusters.

Let $\underset{\sim}{O}$ be some class of subsets of elements of X. A fuzzy class, $\underset{\sim}{C}$, in $\underset{\sim}{O}$ is represented by a membership function $f_{\underset{\sim}{C}}$ which assigns a real number $f_{\underset{\sim}{C}}(o)$ in [0,1] to each of the sets o in $\underset{\sim}{O}$ according to the level of membership of that set in $\underset{\sim}{C}$.

In particular, let $\underset{\sim}{O}$ be that class of subsets o of elements of X such that each set o contains two or more elements of X. Let $\underset{\sim}{C}$ be the fuzzy class of sets o in $\underset{\sim}{O}$ defined by the membership function $f_{\underset{\sim}{C}}$, where $f_{\underset{\sim}{C}}(o)$ is the average of the measures of similarity of all distinct pairs of elements in o. Each set o is called a cluster and the numerical value $f_{\underset{\sim}{C}}(o)$ is a measure of cluster similarity of the cluster o. $\underset{\sim}{C}$ is therefore a fuzzy class of clusters where clusters are sets of "highly similar objects."

The clusters of $\underset{\sim}{O}$ may be ordered according to their cluster similarity, and a hierarchy of partitions can be constructed by forming particular collections of the ordered clusters called "aggregations", which will now be defined.

Let $\underset{\sim}{O}_s = \{o_i : o_i \in \underset{\sim}{O},\ i \in I_s,\ f_{\underset{\sim}{C}}(o_i) \geq s\}$ be the subclass $\underset{\sim}{O}_s$ of $\underset{\sim}{O}$ indexed by a set I_s, such that a cluster o_i is a member of $\underset{\sim}{O}_s$ if the cluster similarity $f_{\underset{\sim}{C}}(o_i)$ is at least as high as the level s, where s is some real number in [0,1]. Now let $\underset{\sim}{I}_{\underset{\sim}{A}_s} \subseteq I_s$ be a subset of the indexing set for $\underset{\sim}{O}_s$ which identifies the sets contained in a particular subclass $\underset{\sim}{A}_s$ of $\underset{\sim}{O}_s$ having the following restrictive properties:

Digitized by the Hunt Institute for Botanical Documentation

(1)* $(\cup_{i \in I_{\underset{\sim}{A}_s}} o_i) \cap (\cup_{i \in I_s \sim I_{\underset{\sim}{A}_s}} o_i) = \emptyset$.

(2)** For every nonempty proper sub-indexing set $J \subset I_{\underset{\sim}{A}_s}$,

$$(\cup_{i \in J} o_i) \cap (\cup_{i \in I_{\underset{\sim}{A}_s} \sim J} o_i) \neq \emptyset .$$

Then the class $\underset{\sim}{A}_s = \{o_i : i \in I_{\underset{\sim}{A}_s}\}$ is called an aggregation in $\underset{\sim}{O}_s$ at the level s. Note that the collection of all aggregations at a level s forms a partition of $\underset{\sim}{O}_s$. A hierarchy of partitions is constructed by ordering aggregations according to their level s. Appendix I contains a discussion illustrating the hierarchy of aggregations for an artificial (non-biological) example.

The hierarchy of aggregations has the property that any object (cluster element) appearing in an aggregation belongs to a cluster in the aggregation that has a cluster similarity at least as large as that of the aggregation level, and its similarity to any object not in the aggregation is less than this level. In this sense, each object in an aggregation at a given level is a member of a "natural group"; that is, "groups" of similar objects, or clusters, that are separated from or

---

*i.e., the union of the sets indexed by $I_{\underset{\sim}{A}_s}$ is disjoint from all the other sets in $\underset{\sim}{O}_s$.

**i.e., the union of any nonempty proper subcollection of the sets indexed by $I_{\underset{\sim}{A}_s}$ shares at least one element with at least one of the other sets indexed by $I_{\underset{\sim}{A}_s}$. Thus, $\underset{\sim}{A}_s$ is a type of maximal class.

Digitized by the Hunt Institute for Botanical Documentation

dissimilar to the other objects not contained in the aggregation. The "gap" or "distance" between two aggregations appearing at a certain level may be defined as the difference between this level and the greatest lower level at which the two aggregations intersect.

Another important classification concept mentioned in the previous section was the density of an aggregation. The density of an aggregation is defined as the ratio of the number of distinct pairs of objects in a cluster containing the largest number of objects from among the clusters in the aggregation, divided by the total number of distinct pairs of objects in the aggregation.

Having chosen these classification concepts, a reasonably efficient search algorithm was programmed to yield a hierarchy of "natural groups" of objects given an object-attribute data matrix. The basic ideas underlying the programmed algorithm search strategy are discussed in Appendix II.

The numerical aid to classification presented here has been applied to the analysis of chemical data from natural populations of the plant species Baptisia nuttalliana. As a result, a geographical structuring of the populations was discovered, which was not apparent to a taxonomist who had analyzed the same data by the conventional taxonomic approach (Horne, 1965; Flake, et al., in press).

This investigation was partially supported by Research Grants GM-11111-04A1 from the National Institutes of Health, National

Digitized by the Hunt Institute for Botanical Documentation

Science Foundation Grant Number NSF-GB-5448X and the Department of Defense's Joint Services Electronics Program (U.S. Army, U.S. Navy, and U.S. Air Force) under Grant AF-AFOSR-766-67.

Digitized by the Hunt Institute for Botanical Documentation

REFERENCES

BELLMAN, R., KALABA, R. & ZADEH, L. (1966). J Math. Anal. & Appl. 13, 1.

CHARLES, D. R. & GOODWIN, R. H. (1943). Am. Natural. 77, 53.

CLAUSEN, J. & HIESEY, W. W. (1958) "Experimental Studies on the Nature of Species." (IV) Genetic Structure of Ecological Races. Carnegie Inst. Wash. Publ. 615.

ESTABROOK, G. F. (1966). J. Theoret. Biol. 12, 297.

FARRIS, J. S. (1966). Evolution. 20, 587.

FLAKE, R. H., et al. (in press).

HORNE, D. B., (1965). Doctoral Dissertation. The University of Texas, Austin, Texas.

LANCE, G. N. & WILLIAMS, W. T. (1966). Nature. 212, 218.

NEEDHAM, R. M. (1961). "A Method for Using Computers in Information Classification", Report M. L. 152, Issued by Cambridge Language Research Unit, Cambridge, England.

PARKER-RHODES, A. F. (1961). "Contributions to the Theory of Clumps", Report M. L. 138, Issued by Cambridge Language Research Unit, Cambridge, England.

ROGERS, D. J. & TANIMOTO, T. T. (1960). Science, 132, 1115.

RUBIN, J. (1967). J. Theoret. Biol. 15, 103.

SNEATH, P. H. A. (1957). J. Gen. Microbiol. 17, 201.

SOKAL, R. R., & CAMEN, J. H. (1965). Syst. Zool. 14, 176.

SOKAL, R.R., & SNEATH, P. H. A. (1963). "Principles of Numerical Taxonomy", San Francisco: W. H. Freeman.

TURNER, B. L. (1967). J. Pure & Appl. Chem. 14, 189.

WIRTH, M., ESTABROOK, G. & ROGERS, D. J. (1966). Syst. Zool. 15, 59.

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX I: Hierarchy of Aggregations in an Artificial (Non-Biological) Example

Consider the space of objects $X = \{x_1, x_2, x_3, x_4, x_5\}$ with a set of (symmetric) similarity measures defined for each pair of objects in X shown in the following table:*

	x_2	x_3	x_4	x_5
x_1	0.90	0.75	0.00	0.00
x_2		0.45	0.00	0.00
x_3			0.00	0.10
x_4				0.80

A partial list of the ordered clusters with associated cluster similarities and the hierarchy of aggregations with associated aggregation levels and densities is shown below.

i	$f_{C}(o_i)$	o_i	$O_s = \{A_s\}$	s	$\{\rho_{A_s}\}$	$\{\{U_{i \in I_{A_s}} o_i\}\}$
1	0.90	$\{x_1, x_2\}$	$\{\{o_1\}\}$	0.90	{1.0}	$\{\{x_1, x_2\}\}$
2	0.80	$\{x_4, x_5\}$	$\{\{o_1\}, \{o_2\}\}$	0.80	{1.0,1.0}	$\{\{x_1, x_2\}, \{x_4, x_5\}\}$
3	0.75	$\{x_1, x_3\}$	$\{\{o_1, o_3\}, \{o_2\}\}$	0.75	{1/3,1.0}	$\{\{x_1, x_2, x_3\}, \{x_4, x_5\}\}$
4	0.70	$\{x_1, x_2, x_3\}$	$\{\{o_1, o_3, o_4\}, \{o_2\}\}$	0.70	{1.0,1.0}	$\{\{x_1, x_2, x_3\}, \{x_4, x_5\}\}$
5	0.45	$\{x_2, x_3\}$	$\{\{o_1, o_3, o_4, o_5\}, \{o_2\}\}$	0.45	{1.0,1.0}	$\{\{x_1, x_2, x_3\}, \{x_4, x_5\}\}$
6	0.37	$\{x_1, x_2, x_3, x_5\}$	$\{\{o_1, o_2, o_3, o_4, o_5, o_6\}\}$	0.37	{0.6}	$\{\{x_1, x_2, x_3, x_4, x_5\}\}$
.	.	.	.	.		.
.	.	.	.	.		.
.	.	.	.	.		.

* e.g., the measure of similarity between x_2 and x_3 is 0.45.

Digitized by the Hunt Institute for Botanical Documentation

The last column on the right identifies the distinct objects contained in each aggregation. This information is useful in assessing the gross nature of the hierarchy of aggregations. Finer details of the hierarchy such as, for example, question of what combination of objects (cluster or clusters) a particular object may associate itself with most closely upon joining an aggregation, require detailed knowledge of the clusters in the aggregations, as well as the enumeration of the objects in the clusters.

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX II: The Search Algorithm Strategy

The definition of a cluster states that any subset (containing two or more objects) of the set of objects being examined is a cluster. Thus, if there are N objects (plants or animals, taxon, etc.) being considered in the numerical data, then there are $2^N - (N + 1)$ clusters of objects present in the data.

The experience of the authors indicates that the number of objects considered in the numerical treatment of biological data is of the order of one hundred. If $N = 100$, then $2^N - (N + 1) \simeq 10^{30}$, which is much too large a number of clusters to enumerate, even for highspeed computers. Thus, if only the identification of the "best" clusters (those with highest cluster similarity) are needed for the purposes of the numerical classification, then it appears imperative to develop a more efficient way of locating the "best" clusters than simply computing the cluster similarity of all $2^N - (N + 1)$, and sorting out the "best" clusters from this (usually) enormous list.

The following property of a cluster is the basis of a relatively efficient search procedure for locating the "best" clusters.

If a cluster o_ℓ containing $n(\geq 3)$ elements, has a cluster similarity $f_{\underset{\sim}{C}}(o_\ell) = \alpha$, $\alpha \in [0,1]$ then there exists a cluster o_k with $n-1$ elements such that $o_k \subset o_\ell$, and $f_{\underset{\sim}{C}}(o_k) \geq \alpha$. This assertion is

Digitized by the Hunt Institute for Botanical Documentation

easily verified by the following simple argument: Let $\{S_{ij}\}$ be the set of symmetric ($S_{ij}=S_{ji}$) similarity measures of each pair of objects under consideration. Then the cluster similarity of a cluster o_ℓ containing n objects is

$$f_{\underset{\sim}{C}}(o_\ell) = \frac{2}{n(n-1)} \sum_{i \in I(o_\ell)} \sum_{j \in I_i(o_\ell)} S_{ij}$$

where $I(o_\ell) = \{i: i \text{ is an index of an object in } o_\ell\}$ and $I_i(o_\ell) = \{j: j \in I(o_\ell) \text{ and } j > i,\ i \in I(o_\ell)\}$.

$$\text{Let } S_\alpha = \sum_{\substack{j \in I(o_\ell) \\ j \neq \alpha}} S_{\alpha j} = \inf_i \sum_{\substack{j \in I(o_\ell) \\ j \neq 1}} S_{ij}$$

$$\text{Then } f_{\underset{\sim}{C}}(o_\ell) = \frac{2}{n(n-1)} \left[S_\alpha + \sum_{i \in I(o_k)} \sum_{j \in I(o_k)} S_{ij} \right]$$

where o_k is the cluster composed of the n-1 objects remaining after object α is removed from o_ℓ. We wish to show that $f_{\underset{\sim}{C}}(o_\ell) \leq f_{\underset{\sim}{C}}(o_k)$. In order to accomplish this, we set

$$f_{\underset{\sim}{C}}(o_\ell) + \mathcal{E} = f_{\underset{\sim}{C}}(o_k)$$

and calculate $\mathcal{E}$. A straightforward calculation yields

$$\mathcal{E} = \frac{4}{n(n-1)(n-2)} \left[\sum_{i \in I(o_\ell)} \sum_{j \in I_i(o_\ell)} S_{ij} - \frac{n}{2} S_\alpha \right]$$

Now, since S_α is the minimum of the sum of the off-diagonal elements of any column of the matrix of similarity measures for the objects o_ℓ, then

$$\mathcal{E} \geq \frac{4}{n(n-1)(n-2)} \left[\frac{n}{2} S_\alpha - \frac{n}{2} S_\alpha \right] = 0$$

Which proves the assertion.

Digitized by the Hunt Institute for Botanical Documentation

The search strategy for locating the "best" clusters is simply to calculate an ordered list of some predetermined length of the "best" clusters containing n objects (beginning with n = 2), and generate a list of "best" clusters containing n + 1 objects by adjoining single objects to the clusters in the ordered list, accepting any new cluster of size n + 1 that has a cluster similarity at least as large as that of the cluster with the lowest cluster similarity in the previous list. The new clusters of size n + 1 are then ordered according to the cluster similarity and the process is repeated until all clusters that have a cluster similarity above the minimum cluster similarity in the first list (n = 2) are found. All of the clusters are then ordered in one list according to their cluster similarity. The generation of the hierarchy of aggregations from the list of ordered clusters is accomplished by a routine enumeration process.

Digitized by the Hunt Institute for Botanical Documentation

UNIVERSITY OF COLORADO

BOULDER, COLORADO 80302

TAXIMETRICS LABORATORY
DEPARTMENT OF BIOLOGY
ARMORY 101

PHONE: 303-443-2211
Ext. 6712

17 May 1968

Dr. John Reeder
Osborn Memorial Laboratories
Yale University
New Haven, Conn. 06520

Dear John:

You will recall that I reviewed a manuscript by Ted Crovello, University of Notre Dame, and turned it down. This was last fall, I think. Since then, I've seen the reworked manuscript, and from my standpoint, it is in good shape now. He is in the process of re-typing it, and either has, or will shortly, resubmit it to you. It has my official blessing, for what that's worth. Since there are no taxonomic decisions (revisions, new taxa, etc.) but rather, deals with methodology, I should think that you could let it run.

Do I hear that you're moving to Wyoming? It would be nice to have you for neighbors, and I hope we'll see you often.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm
cc. Ted Crovello

Digitized by the Hunt Institute for Botanical Documentation

University of Notre Dame
College of Science
Notre Dame, Indiana 46556

Department of Biology

May 13, 1968

Doctor David Rogers
Taximetrics Laboratory
Department of Biology
University of Colorado
Boulder, Colorado 80302

Dear Dave:

Thanks for your letter of 25 April.

The article is being typed in final draft now, so please write to John Reeder. Thanks.

Turning to the Greene Herbarium, whether it gets put on tape gets more doubtful everytime that NSF takes a budget cut. But I hope to have the application in before 15 June which is when I leave here for Kansas. I will get back September 1.

I probably will see you at AIBS.

The Chicago position is interesting, but I told them that I would not even consider coming until September 1969. So they are keeping it open for a year.

Have a good summer. Thanks for writing Reeder.

Sincerely,

Ted

Theodore J. Crovello
Assistant Professor and
Curator of the Herbaria

TJC:ma

Digitized by the Hunt Institute for Botanical Documentation

Night Letter

Dr Theodore Crovello
Dept Biology
Notre Dame Univ
" " Ind

Please call Tuesday afternoon
re: Buttons manuscript
Phone no 303- 443- ~~2211~~ X 6712
Dave Rogers

Digitized by the Hunt Institute for Botanical Documentation

APR 4 1968

University of Notre Dame
College of Science
Notre Dame, Indiana 46556

Department of Biology

March 25, 1968

Doctor David Rogers
Taximetrics Lab
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dave:

Enclosed is the revised manuscript for Brittonia. I have tried to incorporate all of your suggestions. I hope I succeeded with most of them. Could you give it your further study and shoot it back to me either for final copy or another revision?

Certain actions require explanation. It almost is impossible to give a "clear, brief explanation of the methods in an appendix." Tryon's method is too damn complex. As early as 1962 Sokal and Rohlf were able to state this fact simply in an article on factor analysis in Systematic Zoology. They simply referred readers interested in details to standard texts. In my case, I am publishing details of the method in an article coming out in Taxon sometime this year. So, I can refer the reader to a familiar journal. Ok?

I also explained why I used two different clustering procedures, namely to allow readers to understand better the effect of change of OTU's and to emphasize that there is no one, absolute, taximetric method.

I explained that because I cluster sites and taxospecies together does not mean I equate the two. Also, although it may confuse readers at first, if they really are serious about studying the results of the article, it is better to keep taxospeciss identified as numbers in the SPAN diagrams and sites as letters. It provides them with more information.

Finally, I tried to construct a summarizing chart of the effect of change of OTU's, but I found that this was complex if not impossible, because each analysis adds OTU's. Instead, I expanded my discussion and I think I presented the findings in a simpler fashion.

Turning to another subject, are you considering me as a collaborator of the second type as presented on page 18 in the preliminary description of your IR project that you sent me some time ago? I am generating a proposal to computerize the 100,000 sheet E. L. Greene Herbarium, which has possibly over 3,000 type specimens and all sheets are pre 1915. In my proposal I mention that I might utilize ~~your~~

Digitized by the Hunt Institute for Botanical Documentation

your software. Not only am I interested in instruction about the system, but I may be able to generate new uses for you, as well as provide a valuable data bank.

Best regards. I hope to hear from you soon.

Sincerely,

Ted

Theodore J. Crovello
Assistant Professor and
Curator of the Herbaria

TJC:ma
Enclosure

P.S. I am not including the original tables with the MS, as I changed nothing in them & you (or George) have a copy with the second copy of the MS I sent you last time but did not receive back from you. I'm trying to avoid having them retyped.

T.

Digitized by the Hunt Institute for Botanical Documentation

25 April 1968

Dr. Theodore J. Crovello
Department of Biology
University of Notre Dame
Notre Dame, Indiana 46556

Dear Ted:

Herewith, at last, your manuscript. It is so well done, and I am ashamed to say that it was just today that I got the thing started. I read it again; it reads very well, and I will say (when you are ready) to the editor of Brittonia that I am satisfied that you have done a fine job on it.

There are a couple of places in the manuscript that you can still do a little bit of editing on. I would very strongly suggest that you take yourself out of the third person wherever you have mentioned your own citations. I forget the page numbers on which it occurs, but I do recall one statement in there which says "this led Crovello to .." do so and so. Take it out and say "I did so and so" and then cite the place where you did it. The only other thing I would suggest is that since your paper with Ornduff on the Limnanthaceae is now out that under the discussion you can cite that as a more specific reference.

Turning from the subject of your paper to the IR project, we will be very, very pleased to provide you with our software. Most will be ready some time this summer. Since the input of the demonstration which we hope to have by the first of September has a great similarity to the information which you probably want to record about the Greene Herbarium it might be well if you can try to find some "scratch" to get out and take a look at it. If this can't be arranged let me know and we'll see if we can find some way to get together with you. We are most anxious to have data banks and give good tests to our procedure, and if you are willing to give us a data bank we will be happy to have it.

I just remember that you said that you were going to Chicago. Did that come through, or not? If so, I suppose that the E. L. Greene Herbarium will rest in peace without any good IR system for some time to come.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

University of Notre Dame
College of Science
Notre Dame, Indiana 46556

Department of Biology

December 14, 1967

Doctor David Rogers
Department of Biology
Taximetrics Lab
Armory 101
University of Colorado
Boulder, Colorado

Dear Dave:

As you suggested in our telephone conversation yesterday I am sending two copies of the MS for you and George to review. As we discussed you will review it thoroughly and return it to me for correction. I will then return the corrected MS to you for final analysis.

I have just reread the article and have made some minor changes on the ribbon copy. I consider the analyses of the type carried out in the paper essential for future use of numerical taxonomy by taxonomists. It is not the most exciting thing I've ever written, but seems necessary for the confident use of the methods.

With regard to the first two tables I offer the following:

Table 1. I could save space by deleting the table and listing the characters in paragraph form in the text, as: Last year's twig characters - twig color, twig length,...Last year's bud characters-bud color, bud length,....

Table 2. If you want to reduce the content of the table, indicate in pencil what should be deleted.

Dave, I appreciate the interest and effort you have been expending on my part.

Sincerely,

Ted

Theodore J. Crovello
Assistant Professor and
Curator of the Herbarium

TJC:lm

Digitized by the Hunt Institute for Botanical Documentation

14 February 1968

Dr. Theodore J. Crovello
Department of Biology
University of Notre Dame
Notre Dame, Indiana 46556

Dear Ted:

At long last, here are our comments on your paper, "The Effect of Change of Number of OTU's in a Numerical Taxonomic Study."

There are some general points which may be helpful. George pointed out recently that there are two apparent approaches to the development of models for computing machines to be used in taxonomy. (There are others, no doubt.) One is to take the empirical approach, try it, modify it in terms of desirable or undesirable results until a satisfactory procedure is derived. The other way is to decide in advance what is wanted as a result, develop rules which convey the thought processes of the biologist, discover what mathematical model is necessary to develop the desired results, and then proceed to the programming. Clearly, there is no real boundary between the approaches, but I think that the differences are sufficient such that when a person favoring one over the other reads a paper from the other approach, he tends to criticize it in his own terms. That happened with your paper when I read it. That colored my remarks about it. I will point this out to John Reeder when you are ready to return your paper. It will help him understand what happened t the first time, and hopefully, let him see that there is no one royal road to the application of computing machines to taxonomy. I would clearly indicate in the paper that yours is the empirical approach.

Another general point which may be useful. Not so much what is actually said in the introductory statements as what can be read between the lines has an effect on the reader. There is a sort of feeling which comes off that "all numerical taxonomy is better than any orthodox approach." Since you clearly do not mean this to be the case, why put your work in a light where you have to fight straw men? I think that if you look back over the past literature on classification, you will find (though not necessarily explicitly stated) that any classification is affected by the number of objects used to make the classification. So a reader will have the first reaction - "we know this to be the case, so what's new?" What's new is that you can much m more precisely demonstrate how the number of OTU's do modify the concepts. If this is indeed your major thesis in the paper, then start out with the statement of the problem itself, and make no bows to taximetrics or numerical taxonomy, or any other blanket concept. Clearly, you aren't attacking any theoretical aspect of numerical taxonomy here, so why bother beating a dead horse, which is what you do in the first few sentences.

Digitized by the Hunt Institute for Botanical Documentation

Another problem should be faced, if you want the results to get a clear reading. Any clustering method can be shown to produce different results dependent on how you structure your input data, how you include or exclude OTU's, or how you interpret the results of the clustering. Therefore, I'd suggest that you have confused the reader by using two clustering methods, neither of which can be understood unless he reads outside your paper. Many in the audience you are addressing do not know what Jim Rohlf's model is supposed to do, and certainly the explanation isn't sufficient to give a "feeling" for Tyron's. And the audience of Brittonia doesn't (as yet) really give a damn about the differences. For these reasons, I'd suggest you give a clear, brief, explanation of the methods in an appendix.

We have found that botanical readers get impatient with a large number of acronyms. You have supplied the editor with a much easier situation concerning space, but he isn't aware that you have done him a favor by using BDM, BSM, or SPAN diagram, particularly when you also designate your populations by other capital letter combinations. Too many such designations confuse rather than clarify. Why don't you redesignate your OTU's with sequential Arabic numbers?

Nearly related to this last point is the fact that you have used sites as the basis for OTU's in the same matrix with taxospecies. This is a less important point, but it is sure to cause some people to wonder how you make equivalent taxa out of the two. The fact that you use both "site" and "taxospecies" in the same matrix indicates (or may do so) that you consider the taxa equivalent in heirarchical level.

It would seem very profitable if you could build a summarizing chart to indicate which taxa are thrown together depending on the differing numbers of OTU's incorporated in the study, and according to the different clustering techniques.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

THE EFFECT OF CHANGE OF NUMBER OF OTU'S IN A NUMERICAL TAXONOMIC STUDY[1]

THEODORE J. CROVELLO

Crovello, T. J. (Dept. of Biology, Univ. of Notre Dame, Notre Dame, Indiana). The effect of change of number of OTU's in a numerical taxonomic study. Brittonia : . 196 . The effect of change of number of OTU's (operational taxonomic units) in a numerical taxonomic study is investigated using 28 taxospecies of the genus Salix in California and 22 sites of section Sitchenses of the genus. A coefficient of Euclidean distance is used to estimate relationships among the OTU's. Two methods of clustering are employed, Tryon's key communality cluster analyses and the unweighted pair group method. The latter results in a phenogram. It was found that the number of OTU's employed does make a difference. But the difference is small and is centered mostly at the levels of clustering that involve somewhat unique OTU's.

INTRODUCTION

Use of numerical taxonomy in botany is expanding. Gilmartin (1967) cites many references in the recent literature. Taximetric techniques are very impressive and their results possess an air of objective finality about them since they ordinarily are produced by an electronic computer. This is especially so when the results are presented as a single phenogram. The purpose of this and companion studies is to investigate the stability of results when certain modifications are introduced. The current analysis examines the effect of change of number of operational taxonomic units, or OTU's (Sokal and Sneath, 1963), on the results of a numerical taxonomic study.

[1]Costs of manuscript preparation were defrayed by NIH Corollary Grant GM 11935 held by Robert R. Sokal.

Digitized by the Hunt Institute for Botanical Documentation

MATERIALS AND METHODS

Crovello (1966) carried out a numerical taxonomic study of the section Sitchenses of the genus Salix. As now recognized by him, this includes two taxospecies, Salix sitchensis Sanson and Salix Jepsonii Schneider. The former ranges from the southern coast of Alaska south to Santa Barbara, California, and as far east as Idaho. The latter inhabits the Sierra Nevada of California and possibly a few adjacent sites in Nevada and Oregon. Representatives of the section occur from sea level along the Pacific Ocean up to 8000' in the mountains of California, Oregon and Washington.

A useful by-product of his study was a table, or matrix, containing information on over 500 plants for 161 characters. More than one measurement was taken for each character in each plant. As employed here the value in the table is the mean value. Table 1 lists the 99 morphological characters used in the analyses presented here. The complete list of character states for each character may be obtained from the author. The remaining 72 62 characters are unit pubescence characters which in the current study have been replaced by 12 general pubescence characters which indicate the presence or absence of pubescence on an organ. These data are in a form suitable for use in future studies. A punched card deck of this basic data matrix (BDM) is available from the author, or from the Herbarium, University of California, Berkeley. The BDM contains a sample from each of the 30 taxospecies of Salix recognized by Munz (1959) as growing in California, as well as a sample of 22 sites of the section Sitchenses from throughout its natural range. Munz (1959) recognized

Digitized by the Hunt Institute for Botanical Documentation

Salix Counteri Andersson as a distinct species while Crovello (1966) concluded that it is only a synonym of Salix sitchensis.

The above BDM is ideal for a study of the effect of change of OTU's on the results of a numerical taxonomic study because it provides a sample of taxa that are recognized at two different levels of the taxonomic hierarchy. For the present study, the individual plants were grouped into 50 OTU's, the 22 sites of the section Sitchenses (Table 2) and the 28 taxospecies recognized by Munz (1959) but not included in the section Sitchenses (Table 3).

Each character in the 50 OTU by 99 character BDM was condensed to relieve it of the arbitrary weighting arising from the use of different units of measurement in the different characters. Condensation is a term given by Crovello (1966) to a modification of a nameless method of Cain and Harrison (1958) which transforms the recorded range of observed values of a character among the OTU's to a range of zero to one. Condensation has the advantage over the more common technique of standardization (see Sokal and Sneath, 1963) since, with a small sample of OTU's, the transformed range is always the same for each character..

After condensation, the similarity among each pair of OTU's was calculated using the square root of Sokal's distance coefficient (Sokal and Sneath, 1963). This provides a value between zero and one, zero when two OTU's are exactly similar in the context of the 99 characters employed, and one when the two OTU's are completely dissimilar. For convenience and for ease of understanding, the above value was subtracted from one. This new value was used in subsequent clustering. When multiplied by 100 the

Digitized by the Hunt Institute for Botanical Documentation

percent similarity between two OTU's is obtained. These values made up the basic similarity matrix (BSM), an OTU by OTU matrix.

Two methods of clustering were employed. The first used the NT-SYS Computing System designed by F. J. Rohlf, J. Kishpaugh and R. Bartcher to construct phenograms using the unweighted pair-group method. This phase was carried out at the Computation Center, The University of Kansas, Lawrence. The reader is referred to Sokal and Sneath (1963) for an explanation of the method.

The second method of cluster analysis employed was Robert C. Tryon's Key Communality Cluster Analysis. This is a technique for reducing the number of uncorrelated dimensions necessary for depicting the salient trends of relationships among the OTU's from n, the number of OTU's, to k, some number less than n. Call this k space reduced object space. For details of the method, the reader is referred to Tryon (1958a, 1958b, 1959) or to Crovello (1966). This analysis was carried out at the Computation Center, University of California, Berkeley, using the BCTRY system of computer programs (Tryon and Bailey, 1966). The computing time donated by the respective Computation Centers is gratefully acknowledged.

An important component of the BCTRY system of cluster analysis is the geometric representation of combinations of independent dimensions in reduced object space taken three at a time called a SPAN diagram. Figure 3 is an adaptation of a computer printout of a SPAN diagram, attempting to represent three dimensions in two. The three boxes with numbers in them are points on the surface of a sphere. The numbers indicate the dimensions of reduced object space used in the figure. When lines are drawn from each point to the center of the sphere, right angles are formed. An OTU can lie anyplace on or within the sphere. If

Digitized by the Hunt Institute for Botanical Documentation

it lies on the sphere's surface, this means that all of its general variation or communality (Cattell, 1965) is explained in these three dimensions. As an OTU is situated closer to the center of the sphere, less of its variation is accounted for in these dimensions and more of it rests in other dimensions. The result is that except for OTU's lying on the surface of the sphere, an OTU's position is more or less distorted. Because of this, only OTU's that exhibit more than 80% of their communality in the three dimensions of a figure are plotted on that figure. Hence, all OTU's appearing in a figure are close (within one fifth of the radius) to the surface of the sphere. Knowing this, one can estimate the position of each OTU in the three dimensions. For example, OTU 1 lies either above or below the plane of paper close to the sphere's surface. It is not close to the origin. An OTU is below the plane of the paper if it is underlined. Thus in Figure 3 OTU's 1 through 4 lie below the surface while OTU's are almost directly opposite them, 180 degrees away. For further details on the derivation of the SPAN diagram, consult the papers by Tryon referred to above.

Given the overall 50 by 50 BSM, four groups of OTU's were analyzed: (1) the 22 sites of the section Sitchenses; (2) the 28 taxospecies in California outside of the section Sitchenses plus three representative sites of the section; (3) the 28 taxospecies plus all nine sites of the section in California; (4) the 28 taxospecies plus all 22 sites of the section.

RESULTS

The results of the analyses using four sets of OTU's and two clustering methods are presented in Tables 4 and 5 and in Figures 1 to 12. Explanation

Digitized by the Hunt Institute for Botanical Documentation

of the symbols used in the figures is found in Tables 2 and 3. As a general rule, it should be kept in mind that because phenograms represent relationships in only one significant dimension, the phenogram will be less accurate in presenting the overall patterns of variation manifested in the BSM. However, it is often more accurate near the tips, where relationships among highly similar OTU's are portrayed. This is because the clustering procedure leading to a phenogram begins with the OTU's that possess the maximum mutual similarity. After the initial clustering level, some form of average similarity among the cluster formed and the remaining OTU's is computed, which leads to distortion. But the results of the first level of clustering are accurate. On the other hand, with Tryon's technique three dimensions are considered simultaneously. The relationships among the OTU's in these three dimensions are accurate, but the effect of remaining dimensions is not considered. Thus Tryon's method provides an accurate picture of the general position of OTU's or groups of OTU's, with respect to each other, but for most accuracy for "nearest neighbor" OTU's, the phenogram should be consulted.

When studying a phenogram, the vertical distance between OTU's is arbitrary. The only important distance is measured along the horizontal axis. The level of similarity at which two OTU's or clusters join can be read directly on this axis. Each cluster of OTU's is free to rotate on its stem, much like a three dimensional mobile. For instance, in Figure 2 the stem supporting JEPSR1 and JEPSR2 may actually be in any position around SALCRK. Also, TRNSMT is probably close to JEPSR1 and JEPSR2, but it could actually be 180° away from them, on the other side of SALCRK.

Digitized by the Hunt Institute for Botanical Documentation

22 sites of section Sitchenses--Figure 1 presents the results of Tryon's technique of cluster analysis. The two sites from the Sierra Nevada in California (H and I) occupy an extreme position in relation to the others. Remember that they are reflected, their actual position is 180° from where they appear. Passing from this in a clockwise direction around the sphere the next sites (e.g., K, L, G, M, N) are from high altitudes. To the left of IDAHO (R) in the figure the remaining sites spread themselves out into two loosely constructed groups which suggest a northern and south delimitation. Sites A, B, C and E are coastal or near coastal California sites, while sites O, U, T and V are some of the more northerly sites. Viewing the figure as a whole, the strongest trend appears to be from H and I down and to the left in expanding arcs. Such a path is highly correlated with change from high altitude to low altitude.

Turning to Figure 2, note the extreme separation of the two sites from the Sierra (H, I), from the other sites. This strongly influenced Crovello (1966) to maintain Salix Jepsonii as a separate taxospecies. Furthermore, a cluster of high altitude sites is apparent. Below this, except for the presence of SALCRK as an extreme site, no clear clusters present themselves. In general these are the conclusions from Figure 1. But many specific differences occur. For example, the UPGMA method considered HMPTUL and REVSTK most similar, but Tryon's method separates these (P, S). Conversely, whereas Figure 1 depicts HMPTUL (P) closest to MACAMA (D) and ORICK (F), these are separated in Figure 2 (however, compare their relationship in Figure 10).

28 taxospecies plus 3 sites--For this set of OTU's five dimensions in reduced object space were required to portray accurately the relationships

Digitized by the Hunt Institute for Botanical Documentation

among OTU's. For convenience, we shall compare our results with conventional concepts of relationships among the species (Table 3). Commonly two subgenera are recognized in Salix, the Diandrae (stamens one or two) and the Pleiandrae (stamens greater than two). The four representatives of the Pleiandrae (1, 2, 3, 4) formed a cluster in the SPAN diagrams that was well separated from other OTU's. In Figure 3, four representatives of section Cordatae (9, 10, 11, 14) form a cluster, while the other two members of this section (according to Schneider, 1921) are not present. The rest of Figure 3 is in disagreement with conventional ideas as to the content of sections. For example, the three sites from the section Sitchenses (C, L, I) bisect the two members of section Brewerianae (27, 28).

The four members of section Longifoliae (5, 6, 7, 8) cluster together in Figure 4, as do the four members of the subgenus Pleiandrae. The three sites of section Sitchenses are now spread out from each other along the surface of the sphere. Whereas in Figure 3 the distance between any two sites was not much greater than that between some taxospecies, in Figure 4 the distance between sites is much more so.

Figure 5 contains two dimensions of Figure 3, but the third is different. Another member of section Cordatae, Salix lasiolepis (13) is situated midway between other members of the section at the right and two members of section Chrysanthae (18, 19) that inhabit the California coast. The intermediate position of Salix lasiolepis has been recognized by willow taxonomists before. Interestingly, in the present context of 99 characters this position is maintained.

Digitized by the Hunt Institute for Botanical Documentation

Figure 6 reveals several distinct clusters similar to those in the SPAN diagrams. The four species at the top are the representatives of subgenus Pleiandrae, the next four represent section Longifoliae, and the next four species are the members of section Cordatae which appeared together in Figures 3 and 5. The next five taxospecies appear together as a cluster in an order somewhat different than in Figure 3. The three sites of section Sitchenses appear next, but accompanied by Salix Scouleriana (SCOULR). The two species of section Brewerianae (BREWER, DELNRT) form the next cluster. Salix lasiolepis and the two taxospecies of section Chrysanthae also constitute a cluster, as was suggested in Figure 5. The remaining species do not appear to represent exactly the same relationships as indicated in the SPAN diagrams. This is information in itself. For example, OTU's ANGLOR and PLANIF cluster early in Figure 6 but in Figure 3 they (20 and 22) are quite far apart and other OTU's appear between them. This indicates that although OTU's 20 and 22 are relatively well separated in the first 3 dimensions of reduced object space, they remain close over subsequent dimensions, while the OTU's between their in dimensions 1 through 3 become separated from them.

28 taxospecies plus 9 California sites--Figure 7 maintains the four members of the Pleiandrae as a distinct cluster, as well as the four representatives of the Longifoliae. Eight of the nine sites in California appear along an axis that separates two groups of taxospecies. As they are arranged here, the sites suggest a cline from sea level up to the high mountain sites. The two members of section Brewerianae appear below the site axis. Above the site axis are members of various sections as recognized by Schneider.

Digitized by the Hunt Institute for Botanical Documentation

Figure 8 depicts four OTU's of section Cordatae (9, 10, 11, 14) in a loose cluster, with a fifth member (13) lying between the previous ones and the sites of section Sitchenses. The members of the Pleiandrae remain distinct. Above the cluster of sites are the two members of section Brewerianae (27, 28) followed by various "supposedly unrelated" taxospecies.

Figure 9 presents the phenetic relationships as a phenogram. As in the SPAN diagrams, the subgenus Pleiandrae forms a definite cluster, but the relationships within the cluster are different. Whereas in the SPAN diagrams OTU's 1 and 2 are not the closest OTU's to each other, in the phenogram, they are grouped first with each other, as are 3 and 4. This strongly suggests that in the remaining dimensions of reduced object space OTU's 1 and 2 remain as they are while the others diverge. Actually, even in the five space depicted in Figures 7 and 8 the real position of each OTU is not as obvious as it seems. This is because each OTU may not be actually on the surface of the sphere. Recall that an OTU may lie anywhere from the surface of the sphere to down into the sphere to a distance 80 percent of the sphere's radius. The next four OTU's in the phenogram are the representatives of section Longifoliae. They come together in the phenogram just as is suggested by Figure 7. Likewise, the relationships of the next four OTU's (9, 10, 11, 14) reflect those implied in Figure 8.

Next, LASLEP coalesces with HOOKER at a high level of similarity and these unite with PIPERI. Finally, two Sierran taxospecies merge and the entire phenogram then links to the distinct alpine willow, Salix nivalis (NIVALI).

28 taxospecies plus 22 sites--The six OTU's (9, 10, 11, 12, 13, 14) of section Cordatae all appear in the middle right of Figure 10 as a rather

Digitized by the Hunt Institute for Botanical Documentation

loose cluster. Moving down and to the left, the members of section Chrysanthae (18, 19) emerge. Finally, the 22 sites of section Sitchenses occupy the first dimension in reduced object space. As in the previous analyses the sites are arranged so as to suggest an altitudinal cline. Moving up Figure 10, the four members of the Pleiandrae appear as reflected OTU's, while the upper surface contains the majority of the remaining OTU's of the Diandrae. Figure 11 contains only one new dimension, that being defined by representatives of section Longifoliae (5, 6, 7, 8).

As in the SPAN diagrams, the subgenus Pleiandrae and section Longifoliae are clearly maintained at the top of Figure 12. The four members of section Cordatae that are closest in the SPAN diagrams form the next cluster in Figure 12. ANGLOR and PLANIF initiate the subsequent cluster. Note that although they are relatively close in Figure 10, in Figure 12 NIVALI is much closer to PLANIF than is ANGLOR. The next six taxospecies appear in the same general region as they do in Figure 10, but do not agree in detail. One is forced to conclude that for the members of subgenus Diandrae appearing above the sites in Figure 10, the depicted relationships are inaccurate.

The next 22 OTU's in the phenogram are the sites of section Sitchenses. The outline of an altitudinal cline is implied in the clusters. Note, however, exceptions such as the close relationship between MTRANR and MACAMA. The next two clusters of two and three OTU's respectively, reinforce the relationships shown in the SPAN diagrams. Finally GEYERI and NIVALI join the rest of the Diandrae as unique taxospecies, which is not the obvious relationship obtained from Figures 10 and 11.

Digitized by the Hunt Institute for Botanical Documentation

Comparisons among the four analyses--In the three analyses involving taxospecies, both in SPAN diagrams and in phenograms, the recognition of the four OTU's representing the subgenus Pleiandrae as one group was unanimous. The same may be said for the four taxospecies conventionally recognized as belonging to section Longifoliae. Both of these groups are well separated from the rest of the Diandrae but are not independent of each other. This is apparent from the appropriate SPAN diagrams and from Figure 6. It is not apparent from Figures 9 and 12, since the stems of each of the two groups in these two phenograms do not connect with each other before connecting with the remaining OTU's. Although the small distance between point of attachment of the two (0.58 to 0.54) suggests the possibility of a relationship, it cannot be confirmed in these two figures.

Four members (9, 10, 11, 14) of section Cordatae form a loose but definite cluster in each of the last three analyses, both in SPAN diagrams and in phenograms. In Figures 6, 9 and 12, ANGLOR and PLANIF initiate a new cluster. The only difference in the three phenograms is that in Figure 6 the cluster remains separated from that formed by the above four OTU's from section Cordatae until most other OTU's have been grouped. But the short length of the stem separating ANGLOR and PLANIF from the Cordatae cluster indicate that a slight change in relationship would allow them to join earlier with the four OTU's of section Cordatae. In the appropriate SPAN diagrams (Figs. 3, 7, 8) OTU's 20 and 22 maintain the same distance from each other, but the figures differ in the presence or absence of OTU's between them.

Digitized by the Hunt Institute for Botanical Documentation

OTU's 12, 15, 16, 17, 29 and 30 appear more or less as a group throughout all of the relevant analyses, but their internal relationships change. The sites of section Sitchenses exhibit an altitudinal, clinal pattern in all analyses. The two taxospecies of section Brewerianae (27, 28) always appear as a distinct cluster, closely related to section Sitchenses. OTU's 21, 23 and 31 appear to be rather unique OTU's. In the phenograms they always join other OTU's late in the clustering process, whereas in the SPAN diagrams they appear in outlying positions, between clusters, or close to other OTU's. Given the phenogram and other information, we conclude that when they appear close to another OTU in a SPAN diagram, they do not maintain such closeness in other dimensions of reduced object space.

When the first through last analyses are studied in sequence, certain generalizations can be made. Among the phenograms, the cophenetic values increase (Table 4). This is a function of the pattern of variation among each of the four sets of OTU's. The 28 taxospecies appear in the last three analyses and can thus be regarded as a constant in terms of cluster structure. From analyses two through four the number of sites is increased, and this increases the number of small clusters present. With an increase in the number of small clusters, the cophenetic correlation coefficient increases. If the above is true, then the first analysis, concerned only with the 22 sites, should have the highest cophenetic correlation. But it has the lowest. This is because the analyses used data that was condensed over the section only, not over the more variable genus. Thus the first analysis depicts more intrasectional variation than do subsequent analyses.

Digitized by the Hunt Institute for Botanical Documentation

The similarity level at which sites join with sites and taxospecies join with each other is summarized in Table 5. The last column is the most unreliable since it is based on conventional ideas of sectional limits. When Figures 2, 6, 9 and 12 are examined the OTU's that come together at the highest level in one phenogram are not always those that do so in the others.

DISCUSSION

Except for an unpublished numerical analysis of the Limnanthaceae by Crovello and Ornduff, no study of the effect of change of number of OTU's on the resulting clusters is known to the author. From the results of the present paper, we may conclude that, in general, the higher groupings remain unchanged when the only factor that is varied is number of OTU's. However, within each small cluster of taxospecies or sites, the interrelationships of the OTU's change with changes in the number of OTU's used in the analysis.

The two clustering methods used differ in the precision to which relationships among and within clusters are given. Phenograms are most accurate at the initial level of clustering, where two OTU's join to form a cluster. After that their precision decreases sharply. This is to be expected since at the first clustering level the relationships throughout the entire Euclidean object space are considered. But once clusters have been initiated the mean position of each cluster is used in comparisons with other OTU's and clusters. This can produce serious distortion. On the other hand, a method such as Tryon's does not average OTU's, since no actual clustering procedure occurs. As used by Crovello, without oblique rotation (see Cattell, 1965) the method maintains the original relationships

Digitized by the Hunt Institute for Botanical Documentation

among the OTU's but finds those dimensions in object space that exhibit the most density. But Tryon's method of cluster analysis, and factor analysis as well, must tolerate a paradox. Because the results are more accurate it uses more dimensions than do phenograms. When the number of dimensions is greater than three, the relationships are harder to comprehend. Stemming from this, the relationships in any three dimensions may be misleading. The real distance between any two OTU's in the total object space is never less than that depicted in a SPAN diagram. But it may be much more when the other dimensions are included. Thus phenograms and SPAN diagrams complement each other. To obtain the most accurate estimate of relationships, both should be computed.

If one studies the discrepancies among the various results of clustering, it becomes obvious that they center around taxospecies that appear to be remote from other OTU's. This can be determined from the basic OTU by OTU similarity matrix, but it is more readily seen from the correlation matrix derived from it and from the vector of mean similarity values of each OTU with every other OTU. The most remote OTU's have the lowest coefficient of correlation and also have the lowest average similarity value.

In conclusion, the number of OTU's used does make a difference. But the difference is small, and centered mostly at the levels of clustering that involve somewhat unique OTU's.

Digitized by the Hunt Institute for Botanical Documentation

LITERATURE CITED

Cain, A. J. and G. A. Harrison. 1958. An analysis of the taxonomist's judgement of affinity. Proc. Zool. Soc. (London) 131(1):85-98.

Cattell, R. B. 1965. Factor analysis: An introduction to essentials. I. The purposes and underlying models. Biometrics 21:190-210.

Crovello, T. J. 1966. Quantitative taxonomic studies in the genus Salix. Ph.D. Thesis. Univ. of California, Berkeley.

Gilmartin, A. J. 1967. Numerical taxonomy--an eclectic viewpoint. Taxon 16(1):8-12.

Munz, P. A. 1959. A California flora. Univ. of California Press, Berkeley.

Sokal, R. R. and P. H. A. Sneath. 1963. Principles of numerical taxonomy. W. H. Freeman and Co., San Francisco, 359 pp.

Tryon, R. C. 1958a. Cumulative communality cluster analysis. Educ. and Psych. Measmt. 18:3-35.

Tryon, R. C. 1958b. General dimensions of individual differences: Cluster analysis vs.-factor analysis. Educ. and Psych. Measmt. 18(3):477-495.

Tryon, R. C. 1959. Domain sampling formulation of cluster and factor analysis. Psychometrika 24(2):113-135.

Tryon, R. C. and D. E. Bailey. 1966. The BCTRY computer system of cluster and factor analysis. Multivariate Behav. Res. 1:95-111.

Schneider, C. 1921. Systematic enumeration of the sections, species, varieties and forms of American willows. J. Ann. Arb. 3:61-125.

Digitized by the Hunt Institute for Botanical Documentation

Table 1. The 99 Characters Used in the Present Study.

Last Year's Twig Characters

1. Twig color
2. Twig length
3. Twig width
4. Flakiness of twig
5. Twig pruinose

Last Year's Bud Characters

6. Bud color
7. Bud length (dehisced flower buds)
8. Bud appressed to twig
9. Bud scale open or fused
10. Bud apex shape

~~This Year's Bud Characters~~

Leaf Characters, Except Pubescence

11. Stipules present or absent
12. Stipule length
13. Stipule margin
14. Stipule margin glandular
15. Petiole length
16. Petiole glandular
17. Blade length
18. Blade shape
19. Blade margin
20. Blade margin glandular
21. Blade margin revolute
22. Blade base angle

Digitized by the Hunt Institute for Botanical Documentation

Table 1. (continued)

23. Blade apex angle
24. Veins prominent below
25. Color of veins below
26. Abaxial side glaucous
27. Adaxial side of blade lustre
28. Number of leaves from the tip
29. Leaf width (measured at widest point)
30. Veins protruding below abaxial surface
31. Stipule width
32. Stipule shape

Habit Characters

33. Habit
34. Height of plant

~~Unit Leaf Pubescence Characters of Leaf Blade~~

This Year's Twig Characters

35. Length of twig
36. Number of leaves on twig
37. Twig diameter
38. Flakiness
39. Pruinose
40. Color of twig

Common Reproductive Characters

41. Number of aments per lateral branch
42. Ament length
43. Ament width
44. Peduncle length from ament to first leaf
45. Overall peduncle length
46. Peduncle leaf number

Digitized by the Hunt Institute for Botanical Documentation

Table 1. (continued)

47. Peduncle leaf length
48. Peduncle leaf width
49. Peduncle leaf margin
50. Direction of flowering within the ament
51. Ament dense or lax
52. Rachis with bumps like spruce twig
53. Floral scale length
54. Scale shape
55. Scale color
56. Scale margin
57. Scales persistant or deciduous
58. Number of veins on distal end of scale
59. Adaxial nectary number
60. Adaxial nectary shape
61. Adaxial nectary length
62. Adaxial nectary width
63. Abaxial nectary number
64. Abaxial nectary shape
65. Abaxial nectary length
66. Abaxial nectary width
67. Scale pubescent on entire adaxial surface
68. Scale pubescent on entire abaxial surface

Female Characters

69. Stigma lobe length
70. Stigma lobe number
71. Style length
72. Capsule length

Digitized by the Hunt Institute for Botanical Documentation

Table 1. (continued)

73. Capsule width

74. Capsule pedicel length

75. Stigmas revolute

76. Seed length (embryo length)

77. Sead coat length

78. Seed width (embryo width)

79. Cotyledon length

80. Seed hair length

Male Characters

81. Stamen number

82. Anther length

83. Anther color

84. Filament length

85. Filaments divided

86. Pollen length

87. Percent pollen stained in lactophenol

General Pubescence Characters

88. Pubescence on stamen filaments

89. Pubescence on last year's twig

90. Pubescence on flower bud

91. Pubescence on this year's buds

92. Pubescence on abaxial leaf surface

93. Pubescence on adaxial leaf surface

94. Pubescence on this year's twig

95. Pubescence on the peduncle leaf

96. Pubescence on the rachis

97. Pubescence on adaxial side of floral scale

98. Pubescence on abaxial side of floral scale

99. Pubescence on capsule surface

Digitized by the Hunt Institute for Botanical Documentation

Table 2. The 22 Sites of the Section Stichenses Used in the Present Study

Code Letter	Code Name	Site Description
California		
A	SALCRK	Monterey County. Along stream emptying into the Pacific Ocean. From southern boundry of Monterey County north to Lucia. Main collection site was along Salmon Creek, within one mile of ocean.
B	HLYCTY	Santa Clara and Santa Cruz Counties. From 4 to 11 miles inland from ocean. Main collection site was in the vicinity of Holy City, approximately 5 miles south of Los Gatos. In cool area on east slope of Santa Cruz mountains, among Douglas fir.
C	DILLON	Northern Marin and southern Sonoma Counties. From Dillon Beach north to mouth of Russian River. Mostly along immediate coast, but one collection from Monte Rio, 7 miles east of ocean along Russian River. Main collection site was at Dillon Beach. Mostly in exposed habitats.
D	MACAMA	Sonoma County. Along Maacama Creek at junction with California Highway 128. In creekbed in deep shade of tall trees.
E	RKPORT	Mendocino County. Within 6 miles north and south of Rockport, along California Highway 1. Either exposed to salt spray, or away from ocean among liveoak trees in steep terrain.

Digitized by the Hunt Institute for Botanical Documentation

Table 2. (continued)

F	ORICK	Humboldt County. Along U.S. 101, from 5 miles north of Orick south to Eureka.
G	TRNSMT	Trinity and Siskiyou Counties. Most plants from elevations of from 5000' to 6000'.
H	JEPSR1	14 specimens from Tulare County north to Shasta County.
I	JEPSR2	14 specimens from Tulare County north to Tehama County.
Oregon		
J	COOSBY	Curry County to Lane County. From Sixes north to Florence. From ocean inland up to 3 miles. Main collection 0.6 miles east of Sixes. Most specimens from exposed banks of Sixes Creek.
K	CRLAKE	Klamath, Jackson and Lane Counties. Most specimens from west slope of Cascade Mountains between 3000' and 4000'.
L	STEEN	Harney County. Steen Mountains, from elevations of 6000' to 8000'.
M	MTHOOD	Hood River County. Slopes of Mount Hood. At elevations from 3800' to 6000'.
N	WALLOW	Wallowa County. Wallowa Mountains, at elevations from 4500' to 6000'.
Washington		
O	BINGEN	Klickitat County. Along Columbia River near Bingen.
P	HMPTUL	Grays Harbor County. Most specimens collected between Hoquiam and Humptulips.

Digitized by the Hunt Institute for Botanical Documentation

Table 2. (continued)

Q	MTRANR	Pierce County. Mount Ranier National Park. Mostly at elevations from 4500' to 5000'.
Idaho		
R	IDAHO	Bonner, Clearwater and Kooenai Counties, Idaho, and Glacier National Park, Montana.
Canada		
S	REVSTK	Between Revelstoke and Quesnel, British Columbia.
T	BCALTA	Odd locations from inland areas of Alberta, British Columbia and Alaska. Treated as one site to increase sample size.
Alaska		
U	SITKA	Most specimens from immediate area around Sitka.
V	JUNEAU	Specimens either from immediate area around Juneau, or from the Kenai Peninsula. Treated as one site to increase sample size.

Digitized by the Hunt Institute for Botanical Documentation

TABLE 3. The Taxospecies of California Arranged Into Sections According to Schneider (1921).[a]

Code Number	Code Name	Taxospecies
section Pentandrae Dumortier		
1	LASAND	S. lasiandra Benth.
2	CAUDAT	S. caudata (Nutt.) Heller
3	LAEVIG	S. laevigata Bebb
section Nigrae Loudon		
4	GOODIG	S. Gooddingii Ball
section Longifoliae Andersson		
5	HINDSI	S. Hindsiana Benth.
6	EXIGUA	S. exigua Nutt.
7	MELANO	S. melanopsis Nutt.
8[b]	PARKSI	S. Parksiana Ball
section Cordatae Barratt		
9	LUTEA	S. lutea Nutt.
10	LIGULI	S. ligulifolia (Ball) Ball
11	MACKEN	S. Mackenziana (Hook.) Barr.
12	PSCORD	S. pseudocordata And.
13	LASLEP	S. lasiolepis Benth.
14	TRACYI	S. Tracyi Ball
section Adenophyllae Schneider		
15	COMUTA	S. comutata Bebb
16	EASTWD	S. Eastwoodiae Ckll.
17	ORESTR	S. orestera Sch.
section Chrysanthae Koch		
18	PIPERI	S. Piperi Bebb
19	HOOKER	S. Hookeriana Barr.

Digitized by the Hunt Institute for Botanical Documentation

TABLE 3. (continued)

section Ovalifoliae Rydberg

20	ANGLOR	S. anglorum Cham. var. antiplasta Sch.

section Reticulatae Fries

21	NIVALI	S. nivalis Hook.

section Phylicifoliae Dumortier

22	PLANIF	S. planifolia Pursh. var. monica (Bebb) Sch.
23	DRUMSB	S. Drummondiana var. subcoerulea (Piper) Ball

section Sitchenses Bebb

24	SITCHS	S. sitchensis Sans.
25	COULTI	S. Coulteri And.
26	JEPSON	S. Jepsonii Sch.

section Brewerianae Schneider

27	BREWER	S. Breweri Bebb
28	DELNRT	S. delnortensis Sch.

section Discolores Barratt

29	SCOULR	S. Scouleriana Barr.

section Fulvae Barratt

30	LEMMON	S. Lemmonii Bebb
31	GEYERI	S. Geyeriana And.

[a] the only exception to Schneider's assignments is S. Jepsonii. He placed it in section Phylicifoliae.

[b] described after 1921, placed in this section by Ball.

Digitized by the Hunt Institute for Botanical Documentation

Table 4. Summary Statistics on the Cluster Analyses.

Analysis number	OTU's in each analysis	Number of dimensions	Column 4[a]	Column 5[b]	Cophenetic correlation coefficient
1	22 sites	3	.9313	.9243	.803
2	28 taxospecies plus 3 sites	5	.9621	.9820	.808
3	28 taxospecies plus California sites	5	.9948	.9916	.878
4	28 taxospecies plus 22 sites	4	.9269	.9922	.924

[a] Proportion of initial communality exhausted by the number of dimensions extracted.

[b] Proportion of mean square of raw correlation matrix exhausted by the number of dimensions extracted.

Digitized by the Hunt Institute for Botanical Documentation

Table 5. Similarity Levels at Which Sites and Different Categories Cluster.

Analysis number	OTU's in each analysis	Sites	Species	Section
1	22 sites	.64 to .74		
2	28 taxospecies plus 3 sites	.74 to .81	.64 to .84	.64 to .84
3	28 taxospecies plus California sites	.76 to .84	.58 to .82	.58 to .82
4	28 taxospecies plus 22 sites	.76 to .88	.58 to .82	.58 to .82

Digitized by the Hunt Institute for Botanical Documentation

FIG. 1. SPAN diagram of 22 sites of section Sitchenses.

Digitized by the Hunt Institute for Botanical Documentation

Fig. 2. Phenogram of 22 sites of section Sitchenses.

Digitized by the Hunt Institute for Botanical Documentation

Fig. 3. SPAN diagram, set 1, of 28 taxospecies plus 3 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 4. SPAN diagram, set 2, of 28 taxospecies plus 3 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 5. SPAN diagram, set 3, of 28 taxospecies plus 3 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 6. Phenogram of 28 taxospecies plus 3 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 7. SPAN diagram, set 1, of 28 taxospecies plus all California sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 8. SPAN diagram, set 2, of 28 taxospecies plus all California sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 9. Phenogram of 28 taxospecies plus all California sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 10. SPAN diagram, set 1, of 28 taxospecies plus all 22 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 11. SPAN diagram, set 2, of 28 taxospecies plus all 22 sites.

Digitized by the Hunt Institute for Botanical Documentation

FIG. 12. Phenogram of 28 taxospecies plus all 22 sites.

Digitized by the Hunt Institute for Botanical Documentation

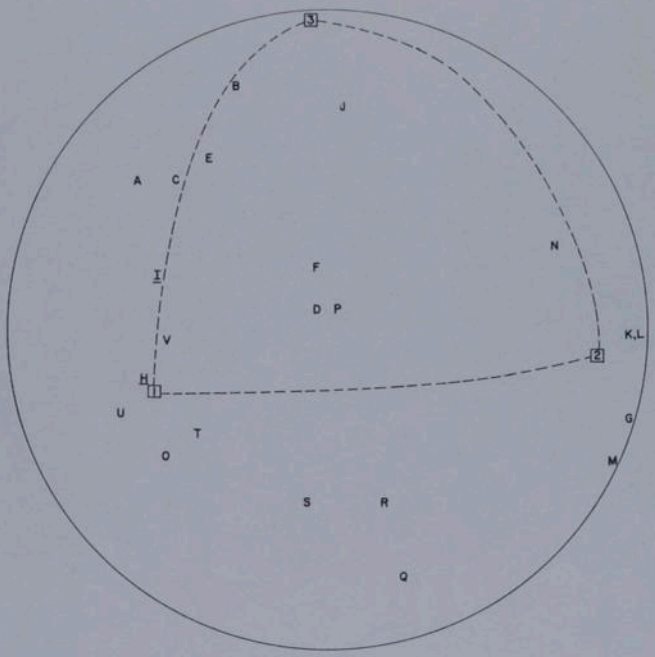


Fig 1
Cruciata

Digitized by the Hunt Institute for Botanical Documentation

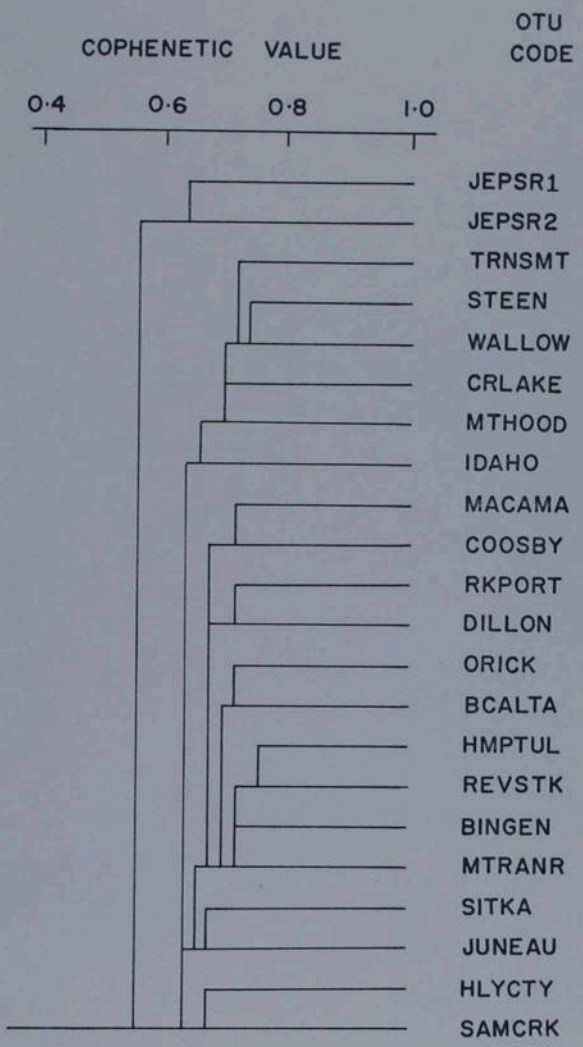


Fig 2
CROVE-LLO

Digitized by the Hunt Institute for Botanical Documentation

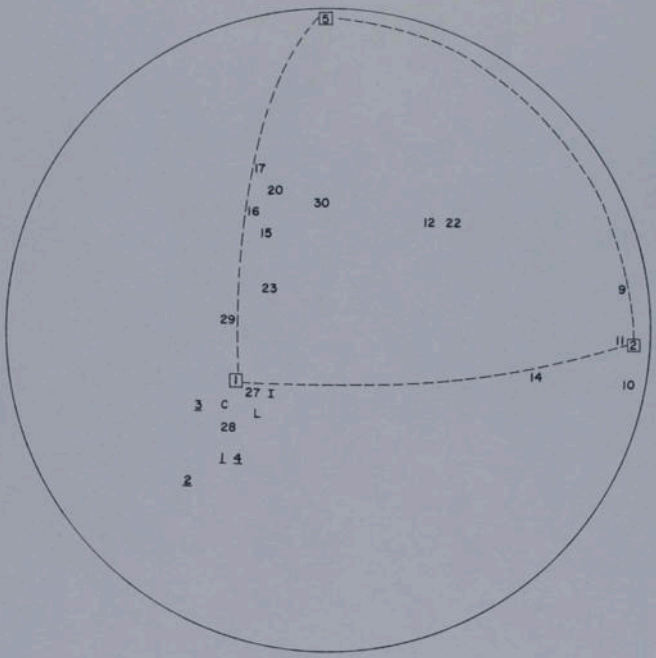


FIG 3
CROVELLO

Digitized by the Hunt Institute for Botanical Documentation

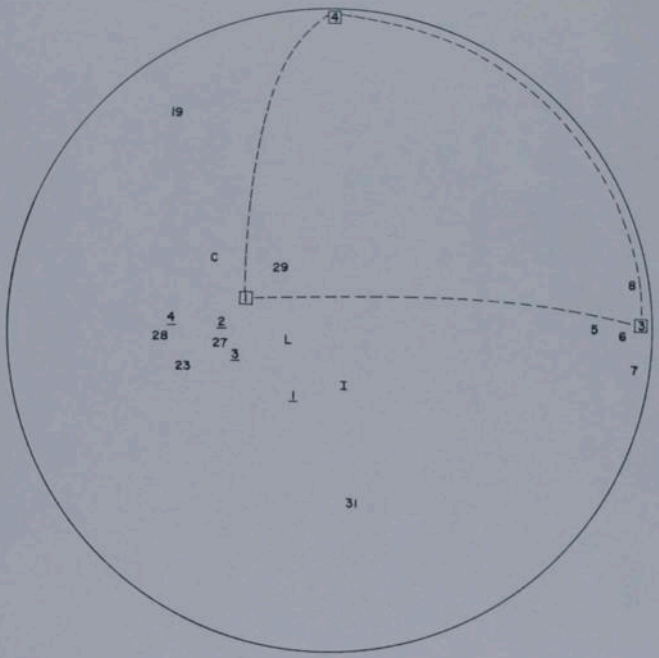


FIG. 4
Crowell

Digitized by the Hunt Institute for Botanical Documentation

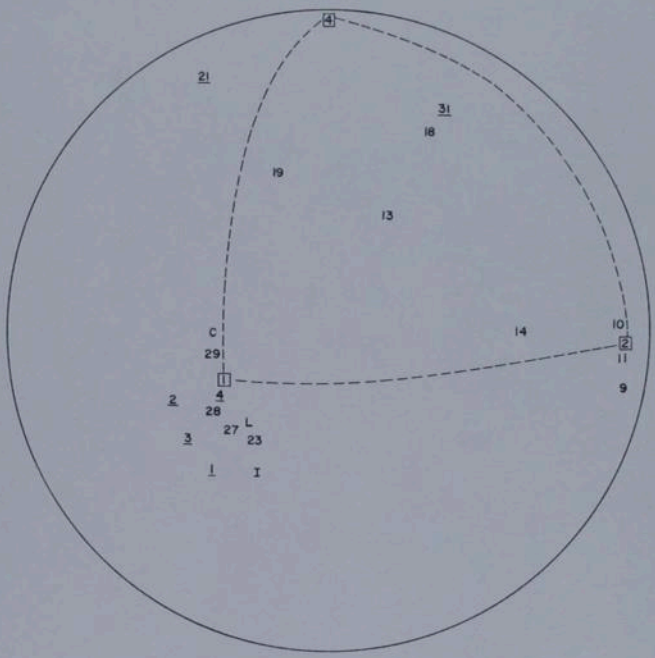


Fig. 5
Cavello

Digitized by the Hunt Institute for Botanical Documentation

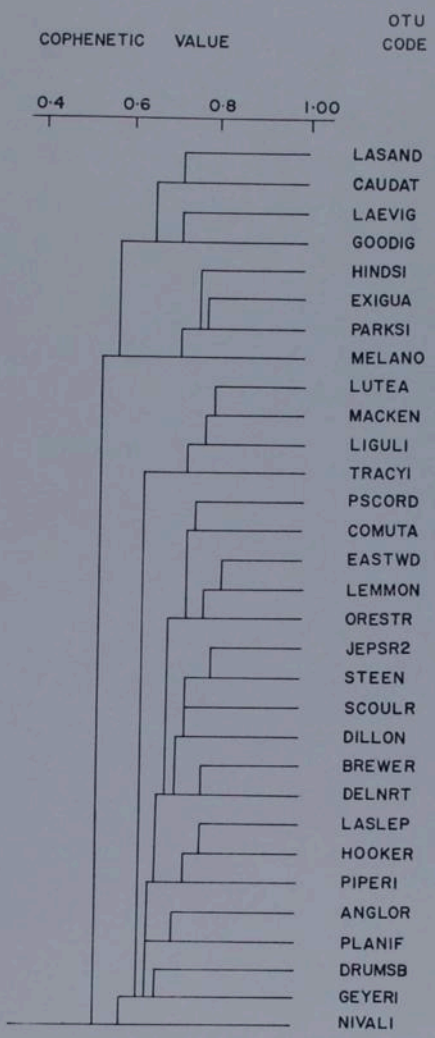


Fig 6
Crowell

Digitized by the Hunt Institute for Botanical Documentation

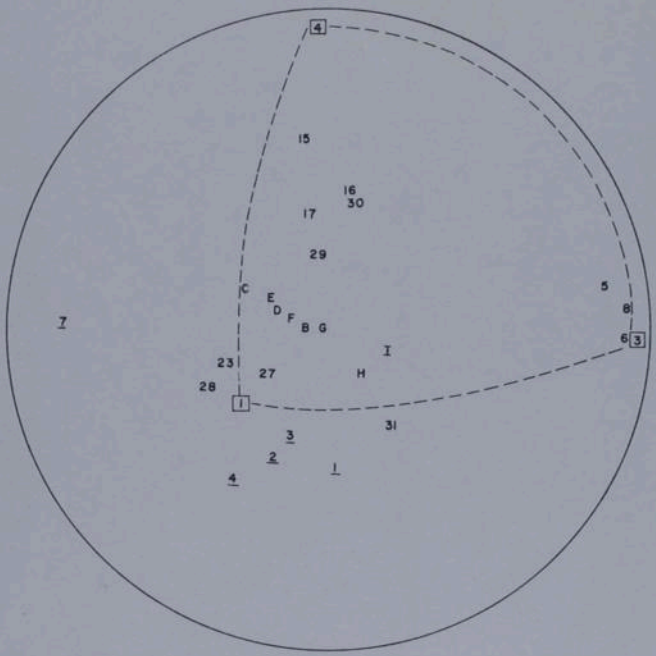


Fig 7
Crowell

Digitized by the Hunt Institute for Botanical Documentation

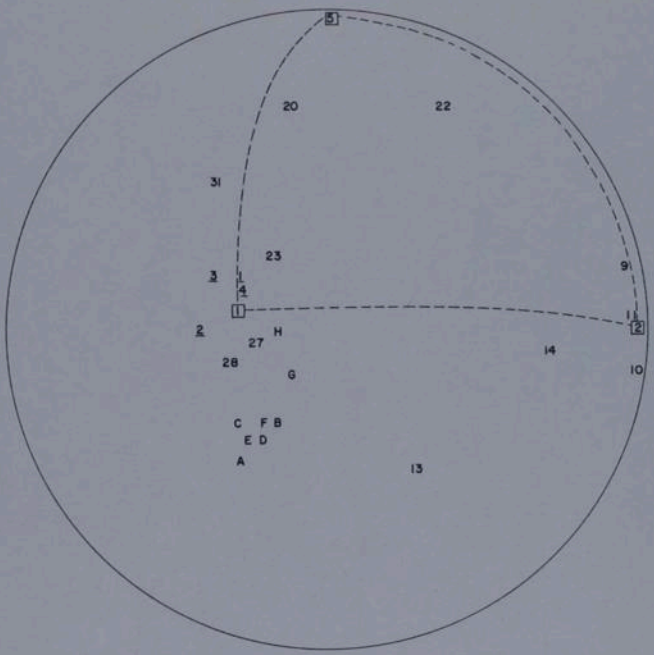


Fig 8
Crowell

Digitized by the Hunt Institute for Botanical Documentation

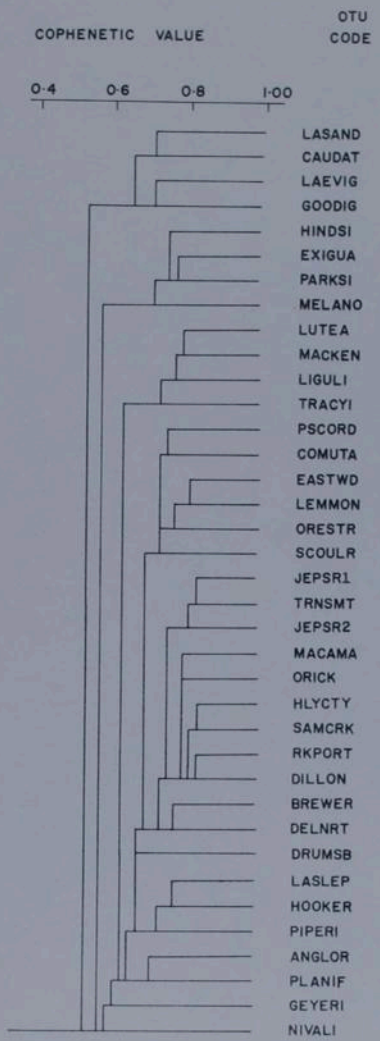


Fig 9

Digitized by the Hunt Institute for Botanical Documentation

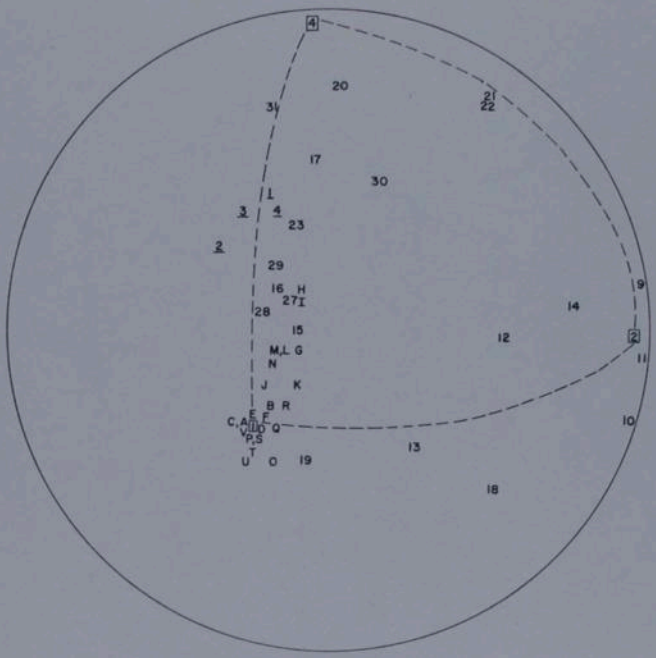


Fig 10
Crowella

Digitized by the Hunt Institute for Botanical Documentation

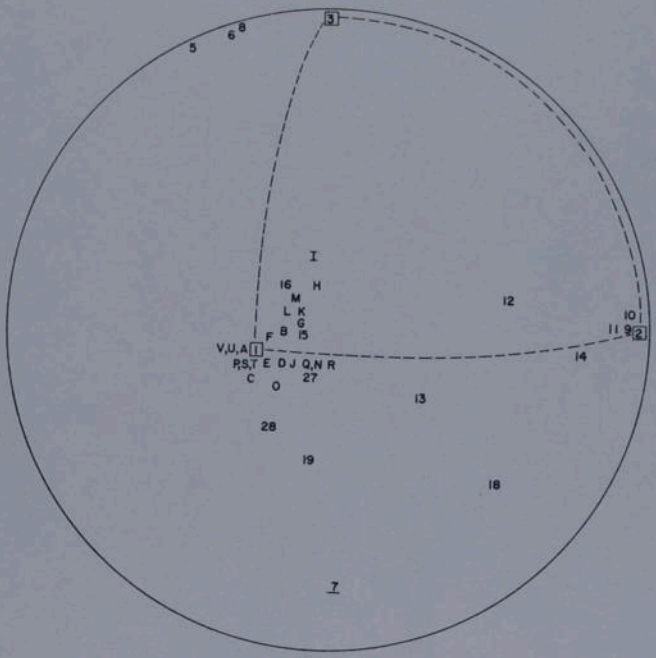


Fig 11
CRAVELLO

Digitized by the Hunt Institute for Botanical Documentation

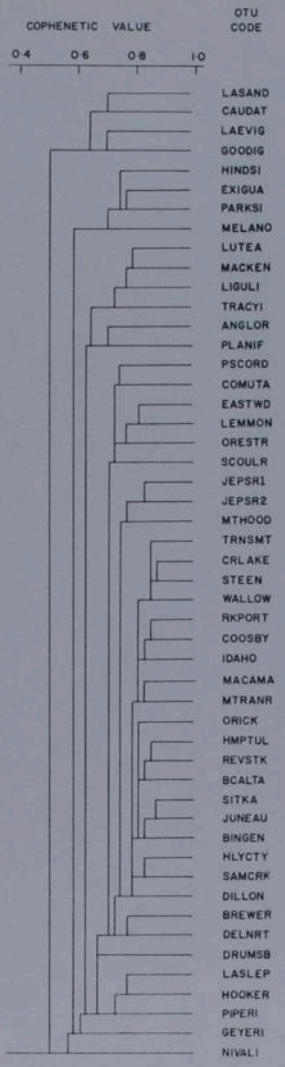


Fig 12
Crowell

Digitized by the Hunt Institute for Botanical Documentation

The American Midland Naturalist

FOUNDED BY J. A. NIEUWLAND, C.S.C.

JOSEPH A. TIHEN, EDITOR

UNIVERSITY OF NOTRE DAME
Notre Dame, Indiana 46556

DEC 1 1967

28 November 1967

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

Dear Dr. Rogers:

The American Midland Naturalist has just received a manuscript submitted by Theodore Crovello, entitled "Effects of Change of Characters in Numerical Taxonomy." The paper is 36 typed pages in length, of which 12 are occupied by a table listing the characters used in the different analyses.

Because we have no one on our editorial board with a primary interest in methods in and usage of numerical taxonomy, we must ask for an "outside" reviewer for this paper. Would you be willing to look it over for us? We're particularly interested in any comments you would offer concerning general worth of the paper, validity of the methods and conclusions, and so on - in other words, the content of the paper. Is it suitable for publication? What changes should be made, if any? We would not expect you to concern yourself with details of form and style, although we would of course welcome any suggestions along those lines also if you'd care to make them.

If you do not feel you can review this paper for us at present, we'd appreciate your suggestions as to some other person or persons whom we might appropriately ask to do so. An addressed postcard is enclosed for your convenience in replying.

We'll greatly appreciate any help you can give us.

Sincerely,

Joseph A. Tihen

Joseph A. Tihen
Editor

JAT/jt

Encl.

Referred to A.M. Rhodes
Dept Horticulture
Univ. Illinois
Urbana

Digitized by the Hunt Institute for Botanical Documentation

13 December 1967

Dr. John R. Reeder, Editor in Chief
Brittonia
Department of Biology
Yale University
New Haven, Conn. 06520

Dear John,

Thanks for your letter of the 27th. I'll go ahead with Ted Crovello.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

DEC 1 1967

Yale University *New Haven, Connecticut 06520*

DEPARTMENT OF BIOLOGY

Osborn Memorial Laboratories

November 27, 1967

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

Dear Dave:

I have your letter of November 21 concerning Ted Crovello's paper. I am sorry that he guessed that you were one of the reviewers. Actually, we took great pains, as we always do, to see that reviewer's names are not given to the author unless, of course, the reviewer signs his sheet of comments and indicates that he does not wish to remain anonymous. In the case of your appraisal, your name was not given to Mr. Crovello and you were referred to merely as "reviewer #1." I think he just guessed from your comments, method of expressing what you felt about the paper, etc.

Please tell Ted Crovello that I certainly am not prejudiced against him or his paper. As I indicated to him, I felt that I had asked competent specialists and their advice was rather on the negative side. If he wants to revise his paper in line with the suggestions of you and the other reviewer, I certainly would be happy to consider it again.

It seems to me that the best procedure would be as follows: tell him to revise his paper and send it to you. When you feel that it is satisfactory and something that we should publish, I will be quite happy to accept it without getting any further opinion from another referee. As I indicated to Crovello, the other reviewer was less critical, but did say that there seemed to be very little return for a tremendous amount of work. It was also suggested that perhaps BRITTONIA was not the most appropriate journal.

In summary then, if Ted Crovello fixes up his paper to your satisfaction, send it along and I shall see that it is published in our journal.

Again, I am sorry that the author was able to identify you so easily, but I am sure it was due to the fact that your comments, even though unsigned, bore a distinct "Dave Rogers" trademark.

Best regards,

John

John R. Reeder
Editor-in-Chief
BRITTONIA

JR:sc

Digitized by the Hunt Institute for Botanical Documentation

21 November 1967

Dr. John R. Reeder
Editor-in-Chief, BRITTONIA
Osborn Memorial Laboratories
Yale University
New Haven, Conn. 06520

Dear John:

I recently saw Ted Crovello, and he gave me hell for the review I gave him on the paper submitted to you. After having cooled him down, I told him that he should try again. But I told him that he'd have to say things differently from the way he did, and clarify his points. He did not emphasize, as I recall, that if you use a different number of OTU's (jargon for objects) that you would get differing types of classifications. He was interested in getting this across in the paper. At that point, I told him he should indeed rewrite and resubmit.

Would you be willing to take another look at his paper if he decided to reorganize it? I told him I was pretty sure that you would, but that I would check with you for him. Let me know, and I'll pass the message on to him.

I am a little unhappy that he found out that I was one of the reviewers, particularly when I was as harsh with his paper as I was. How did it happen? Or can you remember the circumstances? It really doesn't matter now, however.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

Taximetrics Laboratory

Armory 101

21 August 1967

Dr. John R. Reeder
Editor-in-Chief, BRITTONIA
Osborn Memorial Laboratories
Yale University
New Haven, Conn. 06520

Dear Dr. Reeder:

Herewith are my comments. There are some that do not fit the nine points of your suggested critique form and these are added at the end.

1. This material has not been published previously in this form.

2. This research does not warrant publication.

3. BRITTONIA is not the most suitable journal.

4. The illustrations are extremely confusing and could be improved by deletion.

5. No errors of fact, interpretation and/or calculation were noted.
The material is not presented logically from the standpoint of the taxonomist though it seems to be presented logically from the standpoint of some model descriptions.

6. Yes, methods were adequate.

7. Literary style is unclear. Errors in grammar are not noticeable. The material could be condensed.

8. There is familiarity with the literature; the bibliography is O.K.

9. I do not think the title is appropriate, but hasten to add that I do not know an appropriate title.

In addition to these nine points, the following can be stated:

Digitized by the Hunt Institute for Botanical Documentation

1. The amount of information conveyed in the paper does not require the amount of space taken to do so. We can restate the problem in terms as old as taxonomy, by asking whether the number of specimens examined has a direct influence on the conclusions reached in making a classification. Another problem can be restated: given correlated characters, what differences exist between two numerical taxonomic procedures.

2. The SPAN illustrations, which are intended to give a quasi-three dimensional pictorialization of the position of objects in a cluster, give very little information to the ordinary taxonomist.

3. The author is more interested in his methodology than in the information about the plants, and while this is a reasonable approach, I think his methodology is not clearly stated in terms meaningful to most systematists. The relationship between the methods employed and the decisions reached taxonomically are not at all clear. For example, the author states, in a sentence of conclusion, that "the difference is small, and centered mostly at the levels of clustering that involve somewhat unique OTU's." How small is "small", and how unique is "somewhat unique"? A numerical taxonomist, whose god is objectivity, is not permitted to make these subjective statements in conclusion.

4. I had a very difficult time reading this paper because of the number of undefined "jargon" type words. The author assumes that all his readers know "taxospecies," "phenograms" or "OTU's", etc. and these are not defined. Reference is made to Sokal and Sneath, 1963, for a definition of OTU, but, after having read that work, I am still in the dark about the meaning of the "operational" when attached to "taxonomic unit."

5. Use of the third person is clumsy and poor. In all cases he should say "I did ...." or "I published...." etc.

Sincerely yours,

David J. Rogers
Professor of Bio logy

Digitized by the Hunt Institute for Botanical Documentation

Yale University *New Haven, Connecticut 06520*

AUG 1 1 1967

DEPARTMENT OF BIOLOGY
Osborn Memorial Laboratories

July 31, 1967

Dr. David Rogers
Department of Biology
University of Colorado
Boulder, Colorado 80304

Dear Dave:

Enclosed is the manuscript by T. J. Crovello on CTU's in taxonomic study which you have agreed to review. I shall be pleased to have your critical comments and shall be especially interested to know your opinions as to the suitability of this paper for publication in BRITTONIA.

I am enclosing a list of "Suggestions to the Reviewers of Papers" which will indicate the sorts of things I hope you will look for. The author, of course, will appreciate a prompt decision concerning his manuscript so I hope you will have a chance to look it over reasonably soon.

Thanking you in advance for your help with this, I am

Sincerely,

John R. Reeder
Editor-in-Chief
BRITTONIA

JR:me
ENC:

Digitized by the Hunt Institute for Botanical Documentation

SUGGESTIONS TO THE REVIEWERS OF PAPERS

Reviewers of papers submitted to BRITTONIA are earnestly requested to prepare two (2) copies of their criticisms, one an unsigned carbon which can be sent to the author. Simple corrections may be made on the manuscript in soft pencil, but major changes should be noted on a separate sheet of paper.

In preparing your criticisms, please consider the following:

1. Has the material been published previously in the same or similar form?
2. Has the research been carried far enough to warrant publication?
3. Is BRITTONIA the most suitable journal for the publication of this material?
4. Are the illustrations essential? Could any of them be improved? Could they be combined or grouped more advantageously?
5. Are there errors of facts, interpretations, and/or calculations? Is the material presented logically?
6. Were the methods employed adequate to yield the results given?
7. Is the literary style clear and concise? Are there errors in grammar? Could the material be condensed?
8. Is there evidence of an adequate familiarity with the pertinent literature? Is the bibliography carefully prepared to conform to the style as set forth in current issues of BRITTONIA?
9. Is the title appropriate? Could it be shortened or otherwise improved?

Digitized by the Hunt Institute for Botanical Documentation

THE UNIVERSITY OF KANSAS · LAWRENCE, KANSAS · 66044

DEPARTMENT OF ENTOMOLOGY · SNOW ENTOMOLOGICAL MUSEUM

12 July 1967

Doctor John R. Reeder
Osborn Memorial Laboratories
Yale University
New Haven, Connecticut 06520

Dear Doctor Reeder:

Enclosed please find the ribbon copy, one carbon copy, the original figures and two copies of the figures of the manuscript, "The Effect of Change of Number of OTU's in a Numerical Taxonomic Study." Please consider this for publication in Brittonia. This work was begun while I was at Berkeley with Lincoln Constance and was expanded during this past year at Kansas.

The studies reported in the manuscript analyze the effect of change of the number of taxa on the results of numerical taxonomy. In this, the study is unique, as far as I know. But besides being a contribution to the techniques of numerical taxonomy, it is also a contribution to the taxonomy of the willows of California and to the willows of section Sitchensis in and out of California. For these reasons, it seems appropriate for submission to Brittonia.

Perhaps I should comment on the degree to which I describe one of the methods of clustering, Tryon's key communality cluster analysis. A reviewer may say it is too brief. I have not gone into greater detail for two reasons. First, the next level of detail would require more than just a paragraph or two, and I do not feel that the space consumed by it would justify its value. Secondly, Tryon has published expanded explanations of his method in widely available journals. Also, I have presented it in greater detail in both my thesis and in a much longer article being reviewed for University of California Publications in Botany. Of course, the analysis reported in the enclosed manuscript is not dealt with there. But one of the clustering techniques is.

After August 15, 1967, my permanent address will be:

Department of Biology
University of Notre Dame
Notre Dame, Indiana

Please let me know if I can be of further assistance. Thank you.

Sincerely,

Theodore J. Crovello, Ph.D.
Research Associate

TJC:cm

Enclosures

Digitized by the Hunt Institute for Botanical Documentation

FILE

UNIVERSITY OF COLORADO

BOULDER, COLORADO 80302

Taxonomy Laboratory
DEPARTMENT OF BIOLOGY
Armory 101

28 July 1967

Dr. John R. Reeder
Osborn Memorial Laboratories
Yale University
New Haven, Connecticut 06520

Dear John,

You may send me Dr. Crovello's manuscript entitled "The Effect of Change of Number of OYU's in a Numerical Taxonomic Study." I think that I know pretty well what is in it already, but I'll look it over and let you know. my opinion.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

JUL 28 1967

JUL 27 1967

Yale University *New Haven, Connecticut 06520*

DEPARTMENT OF BIOLOGY

Osborn Memorial Laboratories

July 25, 1967

Dr. David Rogers
Department of Biology
University of Colorado
Boulder, Colorado 80304

Dear David:

rec'd reply

I have a manuscript entitled "The Effect of Change of Number of OYU's in a Numerical Taxonomic Study" by T. J. Crovello. The work was started at the University of California but completed while the author was associated with Robert Sokal. This falls within an area in which I believe you have some experience and competence. Would you be willing to look the manuscript over and let me have your opinion as to its suitability for publication in BRITTONIA. If you feel that someone else would be more appropriate as a reviewer, I should appreciate your suggestions.

I shall look forward to your early reply.

Best regards,

John

John R. Reeder
Editor-in-Chief
BRITTONIA

JR:me

Digitized by the Hunt Institute for Botanical Documentation

Taximetrics Laboratory

Armory 101

27 July 1967

Miss Hazel Niehaus, Administrative Assistant
University of California Press
2223 Fulton Street
Berkeley, California 94720

Dear Miss Niehaus:

I enclose herewith my comments concerning Dr. Crovello's manuscript. The manuscript is coming along separately according to your instructions for mailing. The postage charge for the package is $1.55.

My social security number is 160-28-4983, I am a citizen of the United States and my home address is Route 2, Box 44 H, Fort Collins, Colorado 80521. I am not sure how much longer that will continue to be my address, for we are in the process of selling that place to move to Boulder; but for your purposes, I suppose that will do. Please send any future correspondence to this letterhead address.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

UNIVERSITY OF CALIFORNIA PRESS

BERKELEY • LOS ANGELES • NEW YORK

2223 Fulton Street
Berkeley, California • 94720

June 29, 1967

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

Dear Professor Rogers:

Thank you for agreeing to evaluate the Crovello manuscript. It is being sent to you under separate cover.

Enclosed herewith is a question form that may prove helpful in preparing your report. Please use it at your own discretion, however it does emphasize points of interest to the Editorial Committee of the Academic Senate when considering whether or not to publish the work.

In order to expedite the payment of your fee, would you kindly supply us with the following information at the time you send in your review:

Social Security Number
Citizenship
Home address

The Press will reimburse you for the postage required to return the manuscript.

We appreciate your willingness to assist us in this matter.

Sincerely yours,

Hazel Niehaus

(Miss) Hazel Niehaus
Administrative Assistant

P.S. Professor Rogers--I now find that the author has furnished only originals of his 23 figures. Kindly return these via insured post, say $150., at our expense, of course. Am I correct in thinking that the illustrations should accompany the manuscript to be judged properly? HN

Digitized by the Hunt Institute for Botanical Documentation

Reporton

A Numerical Taxonomic Study of the Genus Salix

Section Sitchenses

by

Theodore J. Crovello

1. A short precis of the content:

The paper exxentially describes a set of methods, either put together by the author, or his application to a new area of other workers' methods. The area of application is taxonomy, or more properly, biological classification. The author starts with a problem in the classification of a group of west coast willows (Salix), and discovers (before this work began) that he cannot employ existing techniques of classification to provide a classification which satisfies him. He therefore turns to methods of using computers in classification, and finds a set of methods which take their directions from various statistical methods (in the main, cluster analysis). The author proceeds then to apply his methods to the originally selected problem of several species of willows.

2. Is the subject worthy of study: yes.

3. Does it warrant the extent of the treatment given in this manuscript: yes, but you need not publish all of it - some condensing might be discreetly done - see my discussions later.

4. Extraneous or irrelevant items: pages 25-30 on sex differentiation is much too lengthy and wordy, and doesn't contribute that much to the argument. He should have made some biological argument, rather thann all the emphasis on this or that statistical problem. This latter should be mentioned only by name, with some reference to the voluminous literature on these techniques.

5. Original and significant contribution: While I have been involved in this field for some ten years, I would hesitate to cast judgment on Dr. Crovello's contribution because I could as easily be wrong as right. It is original, Whether it

Digitized by the Hunt Institute for Botanical Documentation

is also significant, I am unwilling to say. It may be. I probably will not employ his methodologies, but I may get some benefit by having the discussion published and in my library. If we can equate significant with "useful" I will say that the paper is useful.

6. Scholarly objectivity: In some respects, it is objective, in others not. It is objective in that the methods selected were chosen because the author considered them best for his problem. It is not objective in that the author tries to consider that "numerical taxonomy" is something superior to any other kind of taxonomy. This is evident both in his introduction and in his conclusion. To my mind, he has not entirely proved the validity of this position, but is acting somewhat defensively. It seems that he is championing a point of view which is not really helpful to his work. Those who have accepted the ideas of "munerical taxonomy" need no further convincing, but those who might, but haven't, will be antagonized by the conclusion's "chip on the shoulder" approach.

But the argument is deeper. The subject of taxonomy is broad, and this approach only tackles part of the subject. Taxonomy is in need of many such techniques, and to indicate that "numerical taxonomy" is "better than" some other undefined approach is useless.

7. Notes applicable - none included.

8. Material well organized: Satisfactory.

9. Style clear and understandable: Not always. Author used too many initials without reference back to the initial point of use. Furthermore, there are many undefined terms, perhaps well-known to the author, but not to the readers. These have generally been marked in the manuscript, and I will not copy them here.

10. There are some needs for editorial work, catching obvious grammatical errors, spellings, etc. I wouldn't do much to the reorganization of the data.

11. I think the study is good.

12, 13, 14. I recommend the study be published, after the modifications mentioned above.

Digitized by the Hunt Institute for Botanical Documentation

- Taximetrics Laboratory

June 26, 1967

Miss Hazel Niehaus
Administrative Assistant
University of California Press
2223 Fulton Street
Berkeley, California 94720

Dear Miss Niehaus:

I will be glad to review A NUMERICAL TAXONOMIC STUDY OF THE GENUS SALIX, SECTION SITCHENSES. Please send it to the following address: Taximetrics Laboratory, Department of Biology, Armory 101, University of Colorado, Boulder, Colorado 80302.

Sincerely yours,

David J. Rogers
Professor of Botany

DJR/ch

Digitized by the Hunt Institute for Botanical Documentation

UNIVERSITY OF CALIFORNIA PRESS

BERKELEY • LOS ANGELES • NEW YORK

2223 Fulton Street
Berkeley, California • 94720

June 22, 1967

Dr. David Rogers
Colorado State University
Fort Collins, Colorado

Dear Dr. Rogers:

We are in search of a reviewer for a manuscript submitted for consideration in our Botany Series. You were suggested as a person especially qualified to serve us as a critic by a member of the Panel of Advisory Editors.

The manuscript is entitled A NUMERICAL TAXONOMIC STUDY OF THE GENUS SALIX, SECTION SITCHENSES. It consists of 81 MS pp. plus 23 figures, and includes 7 tables. The author is Theodore J. Crovello, and this is a revised version of his Ph.D. thesis.

We can offer you a fee of $40 for this service. Your name, of course, would not be revealed to the author.

We would be grateful if you would agree to review this for us. May we hear from you soon -- within a week or so?

Thank you.

Sincerely yours,

Hazel Niehaus

(Miss) Hazel Niehaus
Administrative Assistant

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION
WASHINGTON, D.C. 20550

May 16, 1968

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

REF: B8-2359R

Dear Dave,

We would greatly appreciate your assistance in the evaluation of the enclosed proposal. The Foundation strives to insure that every proposal is fairly and critically evaluated. To achieve this goal, we rely heavily on experts in each field.

Your critical review of the merits of the project will be of great value. Aspects which may enter into such an evaluation include the relative importance, feasibility, and conceptual framework of the project. You may also wish to consider the capacity of the investigator to conduct successfully the contemplated project, as evidenced by his description of developments in the field; his presentation of the proposed project; and his training and research contributions. If the facilities described seem inadequate, this should be noted.

Please rate the proposal at the bottom of the evaluation sheet according to the scientific merit of the project. Budgetary aspects should not enter into the rating, but of course your comments on the budget are welcome.

Your comments on the project will be treated as confidential. The proposal and one copy of the evaluation sheet may be retained for your file. The proposal itself should be considered as a privileged document.

We would like to thank you in advance for the valuable assistance you are providing in making our proposal review process a judicious one.

Yours sincerely,

Paul D. Hurd, Jr.
Kenton L. Chambers
Program Director for
Systematic Biology

Enclosures

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION

PROPOSAL RATING SHEET

Reviewer David J. Rogers

Proposal No.: B8-0963R
Investigator: Robert H. Flake
Institution: University of Texas
Please return to: Systematic Biology
if possible by: 25 JAN 1968

Comments (*Continue on additional sheet if necessary*)

. The most significant deficiency in the proposal is that the author has too little actual advice and contact with the systematic biologists for whom the methods are to be developed. I would rather see a co-principle investigator, a systematic botanist, actually written into the proposal. As now proposed, Turner is added, almost as an after-thought, as "consultant." The type of work Flake has proposed demands that there be much more inter-change between the two disciplines than is indicated.

On the whole, I believe that Flake can do a workman-like job on the type of work he has proposed. He has a sufficient history of effort in this field, which will serve well in developing programs as indicated. The proposal, unfortunately, gives very little detail of the exact direction to be taken, or the actual programs which will be used. Much work is needed to make the computer a tool for systematics, and this proposal may add some needed light in the areas of characterization and prediction.

Facilities are more than adequate, they are ideal. The CDC 6600 is one of the most powerful computers in the country. The budget is modest, and seems well structured for the proposed research.

OVERALL RATING
- ☐ EXCELLENT
- ☒ VERY GOOD
- ☐ GOOD
- ☐ FAIR
- ☐ POOR

Signature of Reviewer: D J Rogers

Date: 23 January 1968

Other suggested reviewers (*optional*):

NSF Form 9-32 July 1967

"Reviewers' comments are solicited and received in confidence by the National Science Foundation, an agency of the U.S. Government, and will be given maximum protection from disclosure permitted under applicable laws, including 5 U.S.C. 552 (The Freedom of Information Act)."

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION
WASHINGTON, D.C. 20550

9 JAN 1968

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

Ref: B8-0963R

Dear Dave,

We would greatly appreciate your assistance in the evaluation of the enclosed proposal. The Foundation strives to insure that every proposal is fairly and critically evaluated. To achieve this goal, we rely heavily on experts in each field.

Your critical review of the merits of the project will be of great value. Aspects which may enter into such an evaluation include the relative importance, feasibility, and conceptual framework of the project. You may also wish to consider the capacity of the investigator to conduct successfully the contemplated project, as evidenced by his description of developments in the field; his presentation of the proposed project; and his training and research contributions. If the facilities described seem inadequate, this should be noted.

Please rate the proposal at the bottom of the evaluation sheet according to the scientific merit of the project. Budgetary aspects should not enter into the rating, but of course your comments on the budget are welcome.

Your comments on the project will be treated as confidential. The proposal and one copy of the evaluation sheet may be retained for your file. The proposal itself should be considered as a privileged document.

We would like to thank you in advance for the valuable assistance you are providing in making our proposal review process a judicious one.

Yours sincerely,

Ken

Kenton L. Chambers
Program Director
for Systematic Biology

Enclosures

Digitized by the Hunt Institute for Botanical Documentation

B8 0963 R

Proposal for Support of Research Entitled

NUMERICAL CLASSIFICATION STUDIES FOR SYSTEMATIC BIOLOGY

Submitted to

National Science Foundation

by

The University of Texas at Austin

Austin, Texas

November 7, 1967

Faculty Initiator:
Professor R. H. Flake

Digitized by the Hunt Institute for Botanical Documentation

Proposal to the

National Science Foundation

from The University of Texas at Austin

for Support of Research Entitled:

Numerical Classification Studies for Systematic Biology

Abstract

A new approach for numerical classification has been developed to aid in biological classification problems, particularly at the infraspecific level or lower. The main goal of the proposed research is to develop numerical methods for assessing the quality of a classification produced by this new approach.

Introduction

Numerous attempts* have been made to utilize the capabilities of the modern high-speed digital computer to aid in biological taxonomic studies. The initiator of this proposal became interested in computer-aided classification methods in 1962 while at Washington University (St. Louis), where as a research associate in the computation center he was involved in assisting a graduate student in the Botany Department who was doing research for his doctoral dissertation on a systematic study of genera of the subtribe Oncidiinae of the Orchidaceae. The method of numerical taxonomy proposed by Rogers and Tanimoto (Rogers and Tanimoto, 1960; reference in Appendix manuscript) was employed in that study. While the study of Oncidiinae was being pursued, a research psychiatrist on the staff of the Washington University Medical School proposed a study of psychiatric patient data utilizing the Rogers and Tanimoto approach. The results of both numerical studies were somewhat unsatisfactory. However, the exercise did give the participants new insight into the nature of numerical classification problems. The graduate

---

*For brief survey, see manuscript in Appendix.

Digitized by the Hunt Institute for Botanical Documentation

student completed the requirements for his Ph.D. and joined Rogers' group, then at the New York Botanical Gardens. The initiator of this proposal became more seriously interested in this area of research and, after joining the faculty of the University of Texas, a collaborative effort with Prof. B. L. Turner of the Botany Department produced a new approach for numerical classification (See manuscript in Appendix). A computer program has been developed embodying this numerical classification technique.

Subsequent application of this numerical tool to the analysis of chemical data from natural populations of the plant species Baptisia nuttalliana revealed a geographical structuring of the populations which was not apparent to a taxonomist who had analyzed the same data without the aid of a computer. This new numerical classification method appears particularly useful in such infraspecific studies, especially where the characters differentiating populations are subtle and the multiplicity of character variations makes morphological as well as chemical systematic studies without the aid of a computer most difficult, if not impossible.

Proposed Research

The main goals of the proposed research are:

1) Develop a method for estimating the predictive value of this recently developed approach for numerical classification.
2) Develop a method of identifying "key characters", that is, the characters "most" responsible for a heirarchical structure obtained from a given set of data.
3) Investigate ways to assess the relative value of the characterization* of data.
4) Apply this method to selected systematic problems.

The first three goals are not thought of here as being unrelated; in fact, the approach envisioned will be iterative in nature, involving the analysis of possibly all these questions simultaneously. In addition, the investigation of the properties of the new numerical classification scheme will be carried out in parallel with the application of this approach in a companion study of

---

*The recent work of Estabrook (Estabrook, G. F., Taxon Vol. 16, April, 1967) on characterization appears relevant here.

Digitized by the Hunt Institute for Botanical Documentation

natural populations of Juniperus virginiana,[†] a widespread species which is notoriously variable across its range. A unique feature of this research will be the intended close relationship between the assumptions in the theoretical study of numerical classification developments and the available or obtainable information from the empirical data in the population study.

Equipment

The only significant item of equipment necessary for the proposed study is the computer. A Control Data Corporation 6600 Computer located in the University of Texas Computation Center is available for this research.

The CDC 6600 system consists of the following major items: one central processor, ten peripheral and control processors, large central memory, display console and keyboard, magnetic tapes, card equipment, printers, and card punch. The 6600 computer executes on the average over 3 million instructions per second. The memory consists of 131,072 words of rapid access, magnetic core storage located in 32 independent banks. The central processor has ten arithmetic and logic units to add, multiply, divide, long add, shift, branch, Boolean, and increment; 24 operating registers; and has floating-point and fixed-point arithmetic. The peripheral and control processors have 4095 words of magnetic core storage; 12 input/output channels; and the following instructions: logical, branch, input/output, central processor access, and central memory access.

Initial software for the CDC 6600 computer system is scientifically oriented, with FORTRAN as the basic language.

Period of Performance

It is recommended that the proposed grant, if awarded, be made effective for a period of two years, beginning June 1, 1968, or some other mutually agreeable date.

[†]Rudloff, E. Von, Irving, R. and Turner, B. L., 1967, Re-evaluation of Allopatric Introgression Between Juniperus ashei and J. virginiana Using Gas Chromatography. Am. J. Bot. 54:660 (Abstract).

Digitized by the Hunt Institute for Botanical Documentation

Personnel

Principal Investigator:	Robert H. Flake
Born:	Little Rock, Arkansas, March 31, 1934
Education:	D.Sc., Washington University, St. Louis, Missouri, June 1962 Major: Electrical Engineering (Automatic Control), Minor: Physics A.M., Physics, Washington University, January 1960 B.S., Engineering Physics, Washington University, June 1956
Experience:	Assistant Professor of Engineering Science at Washington University (St. Louis) from June 1962 to September 1964; Assistant Professor of Electrical Engineering at The University of Texas since January 1966. In these capacities taught undergraduate and graduate level courses in automatic control, conducted research in automatic control theory and in the applications of computers to various taxonomic problems in medicine and biology. Senior Development Engineer and subsequently Consultant for the Aero-Division of Honeywell, Inc., St. Petersburg, Florida, from September 1964 to April 1967. In this capacity assisted in developing calibration and alignment techniques and navigation computational algorithms for inertial navigational systems.
Professional & Scientific Societies:	Institute of Electrical and Electronics Engineers, American Institute of Aeronautics and Astronautics International Association of Plant Taxonomists, Sigma Xi.
Current Professional Interests:	Guidance and control theory, numerical classification

Digitized by the Hunt Institute for Botanical Documentation

Publications: "Volterra Series Representation of Time-Varying Nonlinear Systems," Proceedings of the Second International Federation of Automatic Control Congress, Basel, Switzerland, 1963.

"Volterra Series Representation of Nonlinear Systems," Institute of Electrical and Electronics Engineers Transactions, Part II, 1963, pp. 330-335.

(with J. Zaborszky) "On Fitting with Functional Polynomials," Transactions of the American Society of Mechanical Engineers Journal of Basic Engineering, Volume 89, Series D, Number 2, June 1967, pp. 399-405.

(with J. G. Hartwell) "A Reference Orbit Useful in Integrating Between Activity Spheres." American Institute of Aeronautics and Astronautics Journal, Volume 5, Number 3, March 1967, pp. 599-600.

(with J. R. Cox) "Filter Design for the Average Response Computer," 19th Annual Southwestern Institute of Electrical and Electronics Engineers Conference Record, April 1967.

(with J. H. Satterfield) "A Computer Classification of Psychiatric Patients," 6th Annual Southeastern Regional Meeting of the Association for Computing Machinery and National Meeting of Society for Biomedical Computing, Volume I (Preprints), June 1967.

Current Support and Pending Applications

The initiator of this proposal does not have any current grant support for research in numerical classification. This proposal is not being submitted to any other agency.

Nonsolicitation Statement

Offeror represents that it has not employed or retained any company or person (other than a full-time bona fide employee working solely for the offeror) to solicit or secure this grant and that it has not paid or agreed to pay to any company or person (other than a full-time bona fide employee

Digitized by the Hunt Institute for Botanical Documentation

working solely for the offeror) any fee, commission, percentage, or brokerage fee, contingent upon or resulting from the award of this grant, and agrees to furnish information relating thereto as requested by the contracting officer.

Digitized by the Hunt Institute for Botanical Documentation

Estimated Budget for First Twelve Months
(June 1, 1968 through May 31, 1969)

			NSF Grant	Grantee Share
A.	Salaries and Wages			
	1. Senior Personnel			
	a. 1 - Principal Investigator (two summer months @ 9 mo. academic rate of 12,000)	2 Summer Months	$ 2,667	
	2. Other Personnel			
	d. 1 - Graduate Student	4 1/2 Academic Mos. 3 Summer Mos.	4,242	
	Res. Sci. Asst. III, 1/2 time-9 mos., full-time 3 summer mos. @full-time rate of 565 per month			
	f. 1 - Pre-Baccalaureate Student Approximately 650 hours @ 1.70/hr. for data coding and key punching	3 Calendar Mos.	1,100	
		Total Salaries and Wages	$ 8,009	
B.	Fringe Benefits		260	
		Total of A + B	$ 8,269	
D.	Expendable Supplies (Computer supplies, magnetic tapes, tab cards and paper)		250	
E.	Travel-Domestic (See Budget Notes)		560	
F.	Publication Costs		600	
G.	Other Costs			
	1. CDC-6600-Computer Charges (three hours @ $200 per hour)		600	
H.	Total Direct Costs		$10,279	
I.	Indirect Costs			
	1. On Campus 40.5% of salaries and wages		3,244	
J.	Total Costs		$13,523	
	Less Grantee contribution for indirect costs (excess over 20% of TDC exclusive of computer charges-$600)		-1,308	$ 1,308
Estimated Budget for First Twelve Months			$12,215	$ 1,308

Digitized by the Hunt Institute for Botanical Documentation

Estimated Budget for Second Twelve Months
(June 1, 1969 through May 31, 1970)

			NSF Grant	Grantee Share
A.	Salaries and Wages			
	1. Senior Personnel			
	a. 1 - Principal Investigator (two summer mos. @ 9 mo. academic rate of 13,000)	2 Summer Mos.	$ 2,889	
	2. Other Personnel			
	d. 2 - Graduate Students	9 Academic Mos. 6 Summer Mos.	8,682	
	2 - Res.Sci.Asst.III (1 -1/2 time-9 mos., full-time 3 summer mos. @ full-time rate of 592/mo.) (1 - 1/2 time-9 mos., full-time 3 summer mos. @ full-time rate of 565/mo.)			
	f. 1 - Pre-Baccalaureate Student-Approximately 650 hrs. @ 1.70/hr. for data coding and key punching	3 Calendar Mos.	1,100	
		Total Salaries and Wages	$12,671	
B.	Fringe Benefits		523	
		Total of A + B	$13,194	
D.	Expendable Supplies (Computer supplies, magnetic tapes tab cards and paper)		250	
E.	Travel - Domestic (See Budget Notes)		560	
F.	Publication Costs		600	
G.	Other Costs			
	1. CDC-6600-Computer Charges (three hours @ $200 per hour)		600	
H.	Total Direct Costs		$15,204	
I.	Indirect Costs			
	1. On Campus 40.5% of salaries and wages		5,132	
J.	Total Costs		$20,336	
	Less Grantee contribution for indirect costs (excess over 20% of TDC exclusive of computer charges - $600)		-2,211	$ 2,211
Estimated Budget for Second Twelve Months			$18,125	$ 2,211
Estimated Budget for Twenty-four Months			$30,340	$ 3,519

Digitized by the Hunt Institute for Botanical Documentation

ESTIMATED BUDGET-NOTES:

A. Salaries and Wages

Salaries paid from grant funds at The University of Texas at Austin conform to the rates approved by the Board of Regents for Salaries paid from regular University funds.

B. Fringe Benefits

Federal Social Security(4.4% of 1st $6600 of each salary; 4.9% after January 1, 1969); Workmen's Compensation Insurance (0.50% of applicable salaries) coverage, as may be required by University regulations under the University's self-insuring plan. Under University accounting procedures these items are direct costs.

E. 1. Domestic Travel

For attendance of one technical meeting and one trip to Washington, D. C., to discuss progress of research with NSF. Transportation expense including transportation by plane or rail and Pullman, or by private car at an allowance of 10¢ per mile in lieu of actual expense, plus per diem at not to exceed $17.00 per day on the quarter-day basis.

Digitized by the Hunt Institute for Botanical Documentation

Administrative Information

Further information relating to the technical portions of this proposal may be obtained from Robert H. Flake, Department of Electrical Engineering, The University of Texas at Austin, telephone: Area Code 512,GR1-1014(office), WA6-0717(home). Contractual matters should be referred to the University's Office of Sponsored Projects, Post Office Box 7726, Austin, Texas 78712, telephone Area Code 512, GReenwood 1-1353.

Submitted by:

Original Signed by
Robert H. Flake
Assistant Professor
Department of Electrical Engineering

APPROVED:

ORIGINAL SIGNED BY
JENS M. JACOBSEN
Jens Jacobsen
Executive Director
Office of Sponsored Projects

Original Signed by
A. W. Straiton
Chairman
Department of Electrical Engineering

I certify that the distribution of costs between the direct and indirect categories as shown in the proposal conforms to the usual accounting practices of the institution for all Federally supported or sponsored research, and to the distribution used by the cognizant Federal audit agency.

Original Signed by
James H. Colvin
Vice-President for Business Affairs

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX

NUMERICAL CLASSIFICATION FOR TAXONOMIC PROBLEMS

R. H. Flake

Department of Electrical Engineering

and

B. L. Turner

Department of Botany

The University of Texas, Austin, Texas, U.S.A.

A new procedure for numerical classification has been developed to aid in biological classification problems, particularly at the infraspecific level or lower. The numerical classification concepts that form the basis of this procedure are identified with biological taxonomic concepts. This numerical procedure appears to be consistent with the classification methods employed by most taxonomists.

1. Introduction

Numerous recent attempts (Sneath, 1957; Rogers & Tanimoto, 1960; Sokal & Sneath, 1963; Wirth, Estabrook & Rogers, 1966; Estabrook, 1966; Rubin, 1967; Needham, 1961; Parker-Rhodes, 1961) have been made to utilize the capabilities of the modern high-speed computer to aid in the classification problems associated with the areas of biological taxonomy and information retrieval.

The central problem in these computer classification procedures is the determination of subgroups, usually referred to as clumps or clusters,

Digitized by the Hunt Institute for Botanical Documentation

from a given set of objects that mutually share a majority of their characteristics, while not sharing many characteristics with the other objects not contained in the subgroups. In some classification procedures, particularly those associated with biological taxonomy, a hierarchy of the subgroups is also obtained.

The majority of the numerical approaches used to generate subgroups may be divided into the two categories of sorting algorithms and clump searching algorithms. Most of the procedures developed in the area of numerical taxonomy belong to the sorting algorithm category, whereas the clump searching algorithms were developed largely for information retrieval classifications.

In the sorting algorithm, the calculation begins by computing a table of similarity measures between all distinct pairs of objects. The hierarchy of subgroups is then generated by a sorting strategy (Lance & Williams, 1966) applied to these measures. These subgroups are thereby defined only in the operational sense that the subgroups obtained are the result of applying the particular sorting strategy on the calculated similarity measures. However, in the clump searching algorithm, a clump is usually defined as a set of objects with associated properties expressed in terms of the object similarity measures, and the clump searching algorithm is developed in order to search efficiently for the clumps of objects represented in the data.

A new numerical classification procedure associated with the category of clump searching algorithms and tailored particularly for biological

Digitized by the Hunt Institute for Botanical Documentation

classification problems is presented in this paper.

A discussion of the principles of biological classification is outlined in the next section. The third section introduces the new numerical classification concepts and discusses their relationship to the biological classification concepts presented in the preceding section. A computer program has been developed for the proposed numerical classification procedure, and the basic approach underlying the programmed search algorithm is outlined in the appendix.

2. Principles of Biological Classification

The morphologically oriented taxonomist is often asked to explain his method for arriving at a given decision or group of decisions involved in classification of taxa and their designation as to rank. Most taxonomists answer that organisms are assigned to a category according to the taxonomist's ability to discriminate between them. They are ranked subjectively, based on his best judgment as to what a species or some other category might be. In general, rank is assigned to groups according to the number of features which distinguish between the assemblages. When two or more assemblages are distinguished with difficulty one from the other, they are called varieties or subspecies of a single species; those which are relatively easily distinguished, but clearly related, are called species of the same genus; when two groups of assemblages can be readily recognized and each possesses a few peculiar kinds of characters, these are called genera, and so forth up the scale to where very

Digitized by the Hunt Institute for Botanical Documentation

unusual or wierd assemblages with a combination of many unique features might be called different families, orders, or even divisions.

This may sound non-scientific and rather arbitrary to those who have never attempted the exercise, particularly where the taxa in question are closely related and are sorted in the form of pressed plants or as preserved specimens in jars.

But under natural conditions the problem is not so ill-posed. Features that mark the organisms are easily recognized, variation can be readily evaluated and this related to enviromental features and, if the populations are large, literally thousands of individuals can be examined, albeit superficially, to test the validity of the classified assemblages.

Morphological characters, the basis for most taxononic classifications, are known to be under genetic control. The taxonomist classifies a given population or group of populations as a specific or infraspecific taxon whenever he can detect an assemblage of individuals which seem to share a common gene pool (i.e., individuals which exchange genes among themselves preferentially). The number of genes that differ between two assemblages or gene pools determines the rank that will be assigned those particular taxa.

The determination of the number of genes associated with the character states distinguishing populational assemblages is a laborious and time-consuming task, but this can be approached experimentally and such calculations have been attempted for at least a few taxa

Digitized by the Hunt Institute for Botanical Documentation

(Charles and Goodwin, 1943; Clausen and Hiesey, 1958). Because of the problems inherent in the experimental approach, most taxonomists are content to work on the assumption that genetic similarities or dissimilarities are reflected by morphological similarities or dissimilarities: the more character states shared by two assemblages, the more closely related; the more distant their relationship, the fewer the character states shared.

If the character differences which distinguish between assemblages are indeed a reflection of genetic differences (and most experimental work indicates that this is so), then it should be possible to assign biologically meaningful numerical values to the attributes which characterize these assemblages. These data could then be used in computer programs designed to cluster those individuals or populations (assemblages) which share the largest number of characters. Further, as will be indicated below, numerical treatments are useful in that some objective value can be assigned to the "distances" or "gaps" between clusters and, as a result, classifications of a more nearly uniform nature can be achieved.

No new insights into the problem of assignment of rank to an assemblage (or taxon) will be offered in the discussion that follows; for whether two very distinct species are to be recognized as belonging to the same or different genera (or families!) is largely a matter of convenience or personal preference of the taxonomist. Numerical methods cannot resolve such a problem since the assignment of ranks is largely subjective. However, only rarely is the taxonomist faced with this kind

Digitized by the Hunt Institute for Botanical Documentation

of question. More often the problem resolves itself to relating one or more assemblages to other assemblages so that some spatial ordering among the assemblages (or taxa) is obtained. In short, using all the characters available to him, the taxonomist attempts, usually with rapid scanning by eye and mental assimilation of the visual pattern data, to form "natural groups" and relate them one to the other by making multiple comparisons between and among all the individuals or assemblages under consideration. "Natural groups" are biological aggregations of populations or individuals which, when compared on all their characters, are more closely related one to the other than they are to any other group or population of individuals.

Biological aggregations, once recognized as "natural groups", are usually assembled into formal categories on purely phenetic grounds (i.e., without giving weight to the characters selected for study) or on phyletic grounds (where certain characters, usually selected a posteriori, are given more weight than others in arriving at the arrangement). The philosophy behind these two kinds of taxonomies has been tediously explored (Sokal and Camen, 1965) and it need only be noted here that most taxonomists accept phyletic treatments as preferable to those which are purportedly phenetic only (Turner, 1967). Therefore, careful consideration should be given to the matter of character weighting.

Among plant taxa in general it is well known that certain characters are genetically more variable than others and that, given a multitude of

Digitized by the Hunt Institute for Botanical Documentation

different climatic regions, selection can result in a rapid divergence of such characters in selected populations, the taxonomy of which may depend upon these characters. For example, leaf size, shape and texture are sufficiently variable in most taxa so that plant taxonomists tend to weigh these characters as less significant to phyletic studies than those characters of the flower. Vegetative features are more plastic than floral features and consequently there is a greater likelihood that rapid divergence or perhaps convergence might lead to mistaken relationships, particularly where different phyletic offshoots contribute taxa to similar environments and superficially similar habits are achieved. In any case, differing spatial arrangements of aggregations can be expected whenever the characters concerned are weighted differently. Thus, a worker who does not weigh floral characters as more important than those of the leaf may produce a classification that is different from that worker who weighs vegetative characters at the expense of the floral.

The taxonomist, if he chooses to weigh characters, must do so according to his experience and judgment. The numerical analysis of such data utilizing computers can produce classification models with greater precision, objectivity and a better evaluation of the recorded data than classifications produced by taxonomists without computer assistance. What should be emphasized here, however, is that weighting is inherent in the selection and in the numerical coding process (i.e., the assignment of states) for any character; consequently it is a problem

Digitized by the Hunt Institute for Botanical Documentation

that plagues both phyletic and phenetic worker alike. The computer does not relieve the taxonomist of his responsibility as a taxonomist; he must select and evaluate the states he assigns to the characters concerned. In fact it can be said that a numerical classification method will be only as successful as the ability of the worker concerned to intelligently use this new tool; computers will enable good taxonomists to produce better classifications, but poor taxonomists will not profit from this approach.

It is possible that, in the future, numerical procedures will be developed such that character weighing can be included in the computer program itself. This is especially likely where statistical data from natural or experimentally grown populations are obtained. For example, if the mean and variance is determined from the measurements of characters from numerous individuals of a given species throughout its range, it should be possible to develop an algorithm which will weigh characters in a biologically reasonable fashion according to their variance. This is essentially what most taxonomists do; they subjectively give more weight to characters which have less variation and de-emphasize characters which have a large (seemingly random) variation.

Farris (1966) and others have suggested that the numerical approach, especially where statistical data are available for populations, offers much potential for the resolution of taxonomic problems. This

Digitized by the Hunt Institute for Botanical Documentation

is particularly true at the infraspecific level or lower where the character differences which distinguish populations are subtle and the multiplicity of character variations make the standard morphological approach most difficult if not impossible.

The problem of classification at the specific or higher levels is not so difficult since, as indicated above, the taxonomist can order these taxa with considerable efficiency; what numerical taxonomy adds here is greater precision as to the density (or cohesiveness) of an aggregate and the calculation of a "gap" or "distance" between itself and a related taxon.

We believe that the numerical classification program developed from the concepts presented in the next section is consistent with the taxonomic approach employed by most workers. Nevertheless, the application of this approach will yield meaningful results only if the taxonomist has sufficient knowledge of his data to achieve a system of weighting which reflects the phyletic history of the characters studied.

3. Numerical Classification Concepts

The term numerical classification refers here to a procedure that identifies clusters of similar objects from a given object-attribute (character state) data matrix, and orders these clusters into a hierarchy. In order for this task to be performed with the aid of a computer, the concepts of a cluster and a hierarchy of clusters must first be defined in

Digitized by the Hunt Institute for Botanical Documentation

quantitative terms, and the definitions chosen should be consistent with the goals of the classification.

Quantification of these concepts requires a numerical representation or characterization of the objects to be classified. As previously noted, some form or other of character weighting is inherent in defining a character and in assigning a numerical representation to a character state. Thus, careful consideration regarding the purposes of the classification should be given to the implications of which characters should be utilized in the numerical representation of the objects, and of the relative weight assigned to the characters that are employed.

Then, with the numerical characterization of the objects accomplished, it is necessary to define a quantitative measure of the number of attributes any object shares with any other object, since this is the basis of the general concept of a cluster of similar objects. Several suitable numerical measures of similarity have been described in the literature (Sokal & Sneath, 1963).

The quantification of a classification procedure also requires a definition of a cluster that is associated with the notion of a "group of similar objects". The definition chosen should lend itself to the construction of a hierarchical arrangement of such clusters, if the method is to be applied to biological classification problems. One definition of a cluster that leads to the development of a hierarchical arrangement is most naturally described in terms of the notion of a

Digitized by the Hunt Institute for Botanical Documentation

"fuzzy" set (Bellman, Kalaba, & Zadeh, 1966).

In the usual mathematical language, a collection of all elements which satisfy some stated property is called a set. Thus, a given element is either a member of a set or it is not a member according to whether or not it satisfies the property defining the set. This notion of a set has been the basis for defining clusters in previous clump searching numerical classification approaches. However, there are many significant "groups"* of objects in biological classification problems that do not uniformly possess any (known) property or attribute that could be used to define a set. A concept that appears to be more relevant for describing such "groups" is called a fuzzy set.

A fuzzy set A in a space $X = \{x\}$ is represented by a membership (characteristic) function f_A, defined on X, and takes values in the interval [0,1], with the value of f_A at x, $f_A(x)$, representing the "grade of membership" of x in A. Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A.

The notion of a fuzzy set of sets or fuzzy class is a natural way of describing the concept of a cluster of similar objects introduced here. A fuzzy class may be defined as follows: Let $\underset{\sim}{O}$ be a class of subsets of the elements of X. A fuzzy class, $\underset{\sim}{A}$, is a fuzzy set of the sets in $\underset{\sim}{O}$ with a characteristic function $f_{\underset{\sim}{A}}$, defined on $\underset{\sim}{O}$, that takes on values in the interval [0,1].

---

*Used in this context in a colloquial sense and should not be confused with the mathematical concept of a group.

Digitized by the Hunt Institute for Botanical Documentation

Then, the notion of a cluster of similar objects is defined in terms of a fuzzy class as follows: Let $\underset{\sim}{C}$ be a fuzzy class of the sets in $\underset{\sim}{O}$, where

$\underset{\sim}{O}$ = { o : o is a subset of X containing two or more elements }

with a membership function $f_{\underset{\sim}{C}}(o)$ equal to the average of the measures of similarity of all distinct pairs of elements in o. The numerical value of $f_{\underset{\sim}{C}}(o)$ is called the cluster similarity of the cluster o. The clusters may be ordered according to their cluster similarity, and a hierarchy of partitions can be constructed by forming aggregations of the ordered clusters.

Let $\underset{\sim}{O}_s$ = { o : o ε $\underset{\sim}{O}$, $f_{\underset{\sim}{C}}(o) \geq s$ }, and let $I_{\underline{A}}$ be the index set of the sets contained in a subclass $\underset{\sim}{A}_s$ of $\underset{\sim}{O}_s$ with the following properties: for every subset $I \subseteq I_{\underline{A}}$,

$(\cup_I o_i) \cap (\cup_{I_{\underline{A}} - I} o_i) \neq \phi$, and $(\cup_{I_{\underline{A}}} o_i) \cap (\underset{\sim}{O}_s - \cup_{I_{\underline{A}}} o_i) = \phi$.

Then, the class $\underset{\sim}{A}_s = \{o_i : i \varepsilon I_{\underline{A}}\}$ is called an aggregation at the level s. The collection of aggregations at a level s forms a partition of $\underset{\sim}{O}_s$.

A hierarchy of partitions is constructed by ordering aggregations according to their level s. The formation of this hierarchy has the property that any object (cluster element) appearing in an aggregation belongs to a cluster in the aggregation that has a cluster similarity at least as large as that of the aggregation level, and its similarity to any object not in the aggregation is less than this level. In this sense, each object in an aggregation at a given level is a member of a "natural group"; that is, "groups" of similar objects, or clusters, that are separated from or dissimilar to objects in other clusters. The "gap"

Digitized by the Hunt Institute for Botanical Documentation

between two aggregations appearing at a certain level may be defined as the difference between this level and the greatest lower level at which the two aggregations intersect.

Another important classification concept mentioned in the previous section was the density of an aggregation. The density of an aggregation is defined as the ratio of the number of distinct pairs of objects in a cluster in the aggregation containing the largest number of objects, divided by the total number of distinct pairs of objects in the aggregation.

Having chosen these classification concepts, a reasonably efficient search algorithm was programmed to yield a hierarchy of "natural groups" of objects given an object-attribute data matrix. The basic ideas underlying the programmed algorithm search strategy are discussed in the appendix.

The numerical aid to classification presented here has been applied to the analysis of chemical data from natural populations of the plant species Baptisia nuttalliana. As a result, a geographical structuring of the populations was discovered, which was not apparent to a taxonomist who had analyzed the same data by the conventional taxonomic approach (Horne, 1965; Flake, et. al, in press).

This investigation was partially supported by Research Grants GM-11111-04A1 from the National Institutes of Health, National Science Foundation Grant Number NSF-GB-5448X and the Department of Defense's Joint Services Electronics Program (U.S. Army, U. S. Navy, and U. S. Air Force) under Grant AF-AFOSR-766-67.

Digitized by the Hunt Institute for Botanical Documentation

REFERENCES

BELLMAN, R., KALABA, R. & ZADEH, L. (1966). J Math. Anal. & Appl. 13, 1.

CHARLES, D. R. & GOODWIN, R. H. (1943). Am. Natural. 77, 53.

CLAUSEN, J. & HIESEY, W. W. (1958) "Experimental Studies on the Nature of Species." (IV) Genetic Structure of Ecological Races. Carnegie Inst. Wash. Publ. 615.

ESTABROOK, G. F. (1966). J. Theoret. Biol. 12, 297.

FARRIS, J. S. (1966). Evolution. 20, 587.

FLAKE, R. H., et al. (in press).

HORNE, D. B., (1965). Doctoral Dissertation. The University of Texas, Austin, Texas.

LANCE, G. N. & WILLIAMS, W. T. (1966). Nature. 212, 218.

NEEDHAM, R. M. (1961). "A Method for Using Computers in Information Classification", Report M. L. 152, Issued by Cambridge Language Research Unit, Cambridge, England.

PARKER-RHODES, A. F. (1961). "Contributions to the Theory of Clumps", Report M. L. 138, Issued by Cambridge Language Research Unit, Cambridge, England.

ROGERS, D. J. & TANIMOTO, T. T. (1960). Science, 132, 1115.

RUBIN, J. (1967). J. Theoret. Biol. 15, 103.

SNEATH, P. H. A. (1957). J. Gen. Microbiol. 17, 201.

SOKAL, R. R., & CAMEN, J. H. (1965). Syst. Zool. 14, 176.

SOKAL, R.R., & SNEATH, P. H. A. (1963). "Principles of Numerical Taxonomy", San Francisco: W. H. Freeman.

TURNER, B. L. (1967). J. Pure & Appl. Chem. 14, 189.

WIRTH, M., ESTABROOK, G. & ROGERS, D. J. (1966). Syst. Zool. 15, 59.

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX: The Search Algorithm Strategy

The definition of a cluster states that any subset (containing two or more objects) of the set of objects being examined is a cluster. Thus, if there are N objects (plants or animals, taxon, etc.) being considered in the numerical data, then there are $2^N - (N + 1)$ clusters of objects present in the data.

The experience of the authors indicates that the number of objects considered in the numerical treatment of biological data is of the order of one hundred. If N = 100, then $2^N - (N + 1) \simeq 10^{30}$, which is much too large a number of clusters to enumerate, even for highspeed computers. Thus, if only the identification of the "best" clusters (those with highest cluster similarity) are needed for the purposes of the numerical classification, then it appears imperative to develop a more efficient way of locating the "best" clusters than simply computing the cluster similarity of all $2^N - (N + 1)$, and sorting out the "best" clusters from this (usually) enormous list.

The following property of a cluster is the basis of a relatively efficient search procedure for locating the "best" clusters.

If a cluster o_ℓ containing $n(\geq 3)$ elements, has a cluster similarity $f_{\underset{\sim}{C}}(o_\ell) = \alpha$, $\alpha \in [0,1]$ then there exists a cluster o_K with n-1 elements such that $o_K \subset o_\ell$, and $f_{\underset{\sim}{C}}(o_K) \geq \alpha$. This assertion is

Digitized by the Hunt Institute for Botanical Documentation

easily verified by the following simple argument: Let $\{S_{ij}\}$ be the set of symmetric ($S_{ij}=S_{ji}$) similarity measures of each pair of objects under consideration. Then the cluster similarity of a cluster o_ℓ containing n objects is

$$f_{\underset{\sim}{C}}(o_\ell) = \frac{2}{n(n-1)} \sum_{i \in I(o_\ell)} \sum_{j \in I_i(o_\ell)} S_{ij}$$

where $I(o_\ell) = \{i: i \text{ is an index of an object in } o_\ell\}$ and $I_i(o_\ell) = \{j: j \in I(o_\ell) \text{ and } j > i,\ i \in I(o_\ell)\}$.

$$\text{Let } S_\alpha = \sum_{\substack{j \in I(o_\ell) \\ j \neq \alpha}} S_{\alpha j} = \inf_i \sum_{\substack{j \in I(o_\ell) \\ j \neq 1}} S_{ij}$$

$$\text{Then } f_{\underset{\sim}{C}}(o_\ell) = \frac{2}{n(n-1)} \left[S_\alpha + \sum_{i \in I(o_K)} \sum_{j \in I(o_K)} S_{ij} \right]$$

where o_K is the cluster composed of the n-1 objects remaining after object α is removed from o_ℓ. We wish to show that $f_{\underset{\sim}{C}}(o_\ell) \leq f_{\underset{\sim}{C}}(o_K)$. In order to accomplish this, we set

$$f_{\underset{\sim}{C}}(o_\ell) + \mathcal{E} = f_{\underset{\sim}{C}}(o_K)$$

and calculate $\mathcal{E}$. A straightforward calculation yields

$$\mathcal{E} = \frac{4}{n(n-1)(n-2)} \left[\sum_{i \in I(o_\ell)} \sum_{j \in I_i(o_\ell)} S_{ij} - \frac{n}{2} S_\alpha \right]$$

Now, since S_α is the minimum of the sum of the off-diagonal elements of any column of the matrix of similarity measures for the objects o_ℓ, then

$$\mathcal{E} \geq \frac{4}{n(n-1)(n-2)} \left[\frac{n}{2} S_\alpha - \frac{n}{2} S_\alpha \right] = 0$$

Which proves the assertion.

Digitized by the Hunt Institute for Botanical Documentation

The search strategy for locating the "best" clusters is simply to calculate an ordered list of some predetermined length of the "best" clusters containing n objects (beginning with n=2), and generate a list of "best" clusters containing n + 1 objects by adjoining single objects to the clusters in the ordered list, accepting any new cluster of size n+ 1 that has a cluster similarity at least as large as that of the last cluster in the previous list. The new clusters of size n+1 are then ordered according to the cluster similarity and the process is repeated until all clusters that have a cluster similarity above the minimum cluster similarity in the first list (n=2) are found. All of the clusters are then ordered in one list according to their cluster similarity. The generation of the hierarchy of aggregations from the list of ordered clusters is accomplished by a routine enumeration process.

Digitized by the Hunt Institute for Botanical Documentation

A SUPPLEMENTARY APPENDIX FOR RESEARCH PROPOSAL
NUMBER B8-0963R ENTITLED NUMERICAL
CLASSIFICATION STUDIES FOR
SYSTEMATIC BIOLOGY

Expanded Discussion of the Proposed Research

The main goal of the proposed research is to develop methods for assessing the quality of a classification produced by the recently developed approach for numerical classification. The best classification is the classification with the most predictive value, and results from using properly characterized data. It is proposed to develop numerical methods for assessing the relative value of the characterization of data, and for estimating the predictive value of a classification.

Characterization

The quality of a classification depends on how well the data reflects or characterizes the pertinent biological information. The approach envisioned here for investigating the relative value of data characterization will involve the calculation of the relative "weight" of the coded characters in a set of taxonomic data in order to better inform the biological worker about the character weighting implied by any of a variety of methods of coding. This will make it possible to easily implement a coding process that reflects the biological worker's estimate of the proper a priori character weighting. This is a relatively well-defined aspect of the characterization problem which may not require much investigation to obtain a satisfactory approach, particularly

Digitized by the Hunt Institute for Botanical Documentation

in view of the recent work of Estabrook*, which appears to be relevant here.

Another aspect of the characterization problem involves the identification of the "key characters", or the characters most responsible for a heirarchical structure obtained from application of the numerical classification method to a set of taxonomic data. This information will be valuable to the biological worker in assigning a posteria weight to the taxonomic characters, for re-processing the data if it appears desirable.

The approach envisioned in this portion of the characterization problem involves the dual classification, that is, the numerical classification of the characters themselves, rather than the taxa. Some preliminary investigation with this type of approach has been carried out, and the results are encouraging.

Prediction

Prediction with reference to a classification refers to two sets of phenomena**, previously unused characters and previously unknown taxa. A taxonomist generally works with a small fraction of the total number of characters of an organism. If an entirely new set of characters is tested against a classification and this does not lead to any major revision, then

---

* Estabrook, G. F. (1967). "An Information Theory Model for Character Analysis", Taxon 16, 86.

** Mayr, Ernst. (1965). "Numerical Phenetics and Taxonomic Theory." Syst. Zool. 14, 73.

Digitized by the Hunt Institute for Botanical Documentation

the existing classification is a good one and it is said to have a "high predictive value" in relation to the previously unused characters. Similarly, if newly discovered taxa fit easily into the existing classification, then the classification has a high predictive value.

The goal of our investigation in this area will be to quantify these concepts and develop a method for estimating the relative predictive value of heirarchical classification structures.

These two problem areas, characterization and prediction, are not independent. The investigation of one will inherently involve the other. In addition, the investigation of the quality of the new numerical classification scheme will be carried out in parallel with the application of these methods to selected systematic problems.

Systematic Applications

1. Baptisia. The program(s) developed will be applied to a number of systematic problems, especially those involving the genus Baptisia. This taxon has been studied intensively by comparative phytochemists at the biochemical systematic laboratories in the Department of Botany and there exists a plethora of chemical data both at the specific and infraspecific level. For example, all 16 species of the genus have been carefully surveyed for flavonoids and approximately 80 compounds have been structurally determined. Two

Digitized by the Hunt Institute for Botanical Documentation

of the species, B. nuttalliana[†] and B. leucophaea[††] have been studied populationally so that infraspecific variability from over a broad geographical region might be studied. Preliminary treatment of some of the B. nuttalliana data (almost 40 flavonoid characters varying in 20 states amoung 640 plants collected by random sampling from throughout the range of the species; Horne, 1965) has revealed interesting correlations (referred to in the first appendix). In addition, it is hoped that a program can be developed to treat what is perhaps the most intensively studied natural hybrid swarm to date, over 1000 plants of B. leucophaea and B. sphaerocarpa and their hybrid derivatives in a 10-acre tract, each plant tagged and studied for its flavonoid components.[†††] By a computer treatment of the approximately 50 variables in this population it should be possible to detect the extent of hybridization and backcrossing since the species concerned do not share common flavonoids.

2. Juniperus. Professor B. L. Turner and Dr. Ernst von Rudloff, Organic Chemist at the Prairie Regional Laboratory, Saskatoon, Saskatchewan, Canada,

---

[†]Horne, D. B. (1965). "The Distribution of Flavonoids in Baptisia nuttaliana and B. lanceolata and their Taxonomic Implications." Doctoral Dissertation, The University of Texas, Austin, Texas.

[††]Brehm, K. P. (1962). "The Distribution of Alkaloids, Free Amino Acids, Flavonoids, and Certain Other Phenolic Compounds in Baptisia leucophaea Nutt. var. laevicaulis Gray, and their Taxonomic Implication." Doctoral Dissertation, The University of Texas, Austin, Texas.

[†††]Baetcke, K. P. (1966). "The Structure of an Interspecific Hybrid Population of Baptisia leucophaea and B. sphaerocarpa Based on the Use of Chemical Markers." Doctoral Dissertation, The University of Texas, Austin, Texas.

Digitized by the Hunt Institute for Botanical Documentation

in collaboration, have compiled an enormous amount of terpenoid data, both qualitative and quantitative, for the two species, J. ashei and J. virginiana. The taxa are readily distinguished by their chemistry, there being approximately 60-70 characters which distinguish between the species. Since the species reportedly hybridize over a wide range (in fact this is acclaimed by many workers to be the best documented case for allopatric introgression) it ought to be possible to detect gene flow between the species, if such is occurring. Treatment of the data by preliminary computer programs indicates that little if any hybridization is occuring;[††††] however, the data do suggest that there is a clinal structuring in the populations from Northeast to Southwest across the eastern United States. An intensive sampling program to test this possibility will be undertaken during the winter of 1968. These data will be collected in the form of samples from specific populational sites (i.e. 10 plants collected from each populations, the latter taken at approximately 100 mile intervals in a transect from Austin, Texas, to Washington, D. C.).

3. Other problems. In addition to the treatment of the chemical data from Baptisia and Juniperus, the programs will be applied to morphological data both separately and in combination with the chemical. The program will also be available to other workers who might wish to evaluate their own data. For example, Dr. Donald E. Stone of Duke University has essentially the same kind of quantitative data for the fatty acids of the genus Carya and he has expressed an interest in collaborating with the principal investigator for the

---

[††††] von Rudloff, Irving and Turner, (1967). "Re-Evaluation of Allopatric Introgression Between Juniperus ashei and J. virginiana using Gas Chromatography Am. J. Bot. 54, 660 (Abstract).

Digitized by the Hunt Institute for Botanical Documentation

treatment of these data. Needless to say, the data will also be available to those graduate students working with a wide range of organisms and characters, be they morphological or chemical.

ADDITIONAL PERSONNEL

Professor B. L. Turner, Director of the Herbarium and Chairman of the Department of Botany at The University of Texas, has agreed to act as a systematic consultant for this proposed research program.

Digitized by the Hunt Institute for Botanical Documentation

REFERENCE LIST IN THE FIRST APPENDIX REVISED
TO SHOW FULL TITLES OF ARTICLES CITED

BELLMAN, R., KALABA, R. and ZADEH, L. (1966). "Abstraction and Pattern Classification." J Math. Anal. & Appl. 13, 1.

CHARLES, D. R., and GOODWIN, R. H. (1943). "An Estimate of the Minimal Number of Genes Differentiating Two Species of Golden-Rod with Respect to Other Morphological Characters." Am. Natural 77, 53.

CLAUSIN, J. & HIESEY, W. W. (1958). "Experimental Studies on the Nature of Species." (IV) Genetic Structure of Ecological Races. Carnegie Inst. Wash. Publ. 615.

ESTABROOK, G. F. (1966). "A Mathematical Model in Graph Theory for Biological Classification." J. Theoret. Biol. 12, 297.

FARRIS, J. S. (1966). "Estimation of Conservatism of Characters by Constancy within Biological Populations." Evolution. 20, 587.

FLAKE, R. H., et. al. (in press) "Populational Analysis by Methods of Numerical Systematics."

HORNE, D. B., (1965). "The Distribution of Flavonoids in Baptisia nuttaliana and B. Lanceolata and their Taxonomic Implications." Doctoral Dissertation. The University of Texas, Austin, Texas.

LANCE, G. H. & WILLIAMS, W. T. (1966). " A Generalized Sorting Strategy For Computer Classifications." Nature. 212, 218.

NEEDHAM, R. M. (1961). "A Method for Using Computers in Information Classification," Report M. L. 152, Issued by Cambridge Language Research Unit, Cambridge, England.

PARKER-RHODES, A. F. (1961). "Contributions to the Theory of Clumps," Report M. L. 138, Issued by Cambridge Language Research Unit, Cambridge, England.

ROGERS, D. J. & TANIMOTO, T. T. (1960). "A Computer Program For Classifying Plants." Science, 132, 1115.

Digitized by the Hunt Institute for Botanical Documentation

RUBIN, J. (1967). "Optimal Classification into Groups: An Approach For Solving the Taxonomy Problem." J. Theoret. Biol. 15, 103.

SNEATH, P.H.A. (1957). "The Application of Computers to Taxonomy." J. Gen. Microbiol. 17, 201.

SOKAL, R. R., & CAMEN, J. H. (1965). "The Two Taxonomies: Areas of Agreement and Conflict." Syst. Zool. 14, 176.

SOKAL, R. R., & SNEATH, P.H.A. (1963). "Principles of Numerical Taxonomy," San Francisco: W. H. Freeman.

TURNER, B. L. (1967). "Plant Chemosystematics and Phylogeny." J. Pure & Appl. Chem. 14, 189.

WIRTH, M., ESTABROOK, G. & ROGERS, D. J. (1966). "A Graph Theory Model for Systematic Biology, with an Example for The Oncidiinae (Orchidaceae)." Syst. Zool. 15, 59.

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION

PROPOSAL RATING SHEET

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

Proposal No.: B8-2292R
Investigator: Robert A. Davidson
Institution: Catholic U. of America
Please retutn to: Systematic Biology
if possible by: 29 MAY 1968

Comments *(Continue on additional sheet if necessary)*

A summary of this request seems to be:

1. Taxonomy, particularly phylogenetic taxonomy, is a very complex affair which the human can do because he is a phylogenetic product.

2. Davidson has a Mark I computer program called WHALE. It isn't very good yet, and maybe he can make it better (or it can make itself better) by some mysterious process which is not known to him or to us. "Making it better" is apparently intended to mean - to do some job of classifying. Just what WHALE does, we are not told.

3. The genus Froelichia is too complex, either for WHALE I or II, so what is known about the genus will be written up on a "best we can do" level, and then shelved. Instead, Davidson will pick a well-behaved genus such as Eustoma which can tell him (or WHALE II) what he ought to do.

4. What he intends (or ought to do) is not clear to me, and for this reason, I cannot but say that he should go back and rewrite this proposal. Incidentally, I never did find out what "cybernetic systematics" is.

OVERALL RATING
- ☐ EXCELLENT
- ☐ VERY GOOD
- ☐ GOOD
- ☐ FAIR
- ☒ POOR

Signature of Reviewer:

Date: May 21, 1968

Other suggested reviewers *(optional)*:

NSF Form 173, January 1968

" Reviewers' comments are solicited and received in confidence by the National Science Foundation, an agency of the U.S. Government, and will be given maximum protection from disclosure permitted under applicable laws, including 5 U.S.C. 552 (The Freedom of Information Act)."

Supersedes NSF Form 9-32 which may be used until supply is exhausted

Digitized by the Hunt Institute for Botanical Documentation

NATIONAL SCIENCE FOUNDATION
WASHINGTON, D.C. 20550

May 15, 1968

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

Ref: B8-2292R

Dear Dave,

We would greatly appreciate your assistance in the evaluation of the enclosed proposal. The Foundation strives to insure that every proposal is fairly and critically evaluated. To achieve this goal, we rely heavily on experts in each field.

Your critical review of the merits of the project will be of great value. Aspects which may enter into such an evaluation include the relative importance, feasibility, and conceptual framework of the project. You may also wish to consider the capacity of the investigator to conduct successfully the contemplated project, as evidenced by his description of developments in the field; his presentation of the proposed project; and his training and research contributions. If the facilities described seem inadequate, this should be noted.

Please rate the proposal at the bottom of the evaluation sheet according to the scientific merit of the project. Budgetary aspects should not enter into the rating, but of course your comments on the budget are welcome.

Your comments on the project will be treated as confidential. The proposal and one copy of the evaluation sheet may be retained for your file. The proposal itself should be considered as a privileged document.

We would like to thank you in advance for the valuable assistance you are providing in making our proposal review process a judicious one.

Yours sincerely,

Paul D. Hurd, Jr.

Kenton L. Chambers
Program Director for
Systematic Biology

Enclosures

Digitized by the Hunt Institute for Botanical Documentation

B8 2292 R

THE CATHOLIC UNIVERSITY OF AMERICA
Washington, D. C., 20017

Reference No. 68-45

Research Proposal:

Biologic, chemotaxonomic and biometric studies on Froelichia and Eustoma as backdrops for the pursuit of cybernetic systematics

Desired Starting Date: 1 September 1968
Duration of Requested Support: Two Years

Submitted to:

National Science Foundation
Washington, D. C.

Dr. Robert A. Davidson,
Principal Investigator

Dr. J. P. O'Connor, Dean,
Graduate School of Arts and Sciences

Dr. Dale C. Braungart
Head, Department of Biology

Digitized by the Hunt Institute for Botanical Documentation

Introduction

The history of science suggests that within given sets of phenomena there always exist "golden thread" phenomenologic subsets, or components, of relatively simple activity or behavior or mechanism which, when discovered, permit an amazingly encompassing aspect of these phenomena to be viewed at a level that we consider far more sophisticated than was possible prior to their disclosure. Thus, e.g., Mendelian particulate genetics, the orbital model of the atom, the quantum theory of energies, the Watson-Crick and related models of biologic replication and information coding, have had profound affects upon the conceptual worlds in which they participate. Each of these"breakthroughs" has one property in common: a notional framework is provided within which events are seen to occur with mathematically rigorous predictability.

Viewed from a slightly different perspective, it is clear that those things which man feels he understands best, can rely on most, can build upon and extrapolate from most confidently, are the things which follow "laws" which can be expressed mathematically. Thus, in large part, physics, surveying, chemistry, rocketry, birth control, hydraulics, etc. rely basically upon arithmetically expressible foundations. Many of these foundations are in a sense superficial and transparent and it seems logical to conclude that these very properties were responsible for man's understanding of them at a relatively early stage in scientific history. In other words, some phenomena are more "shrouded" than others; generally those least shrouded are reduced to mathematics earliest, those most veiled elude this level of reduction longest. That the laws of gravity should be perceived prior to the rules of protein templating is fully expected.

Let us concede that many relatively simple sets of phenomena, or systems, are governed by integral components, or system subsets which can be mathematically portrayed. Thus, planetary systems can be understood via the ramifications of mass and gravity "subsets"; electronic and magnetic phenomena depend upon phenomenologic subsets of kinetic energies; an understanding of DNA action depends in part upon the concept of permutation of base pair members; organismal photoperiodic response has a component that can be reduced to simple physical properties. Clearly, these and other relatively simple systems depend for their internal coordination and order upon the interactions between simpler system subsets whose properties can rather accurately be described by appropriate mathematics. If relatively simple systems are so dependent, then it surely must follow that the most complex systems are similarly regulated. The very complexity of the complex system is the "shrouding" which hides the elementary "laws" which govern its order, but these rules certainly must exist if the complex system exists. If order were not inherent within

Digitized by the Hunt Institute for Botanical Documentation

the system it would crumble to chaos (and could not have developed in the first place).

A crude simile may help to erase the gibberish tone of the above commentary. Let us exemplify a "relatively simple system" as "the driving of a nail into a piece of wood with a hammer". The "system subsets" which comprise this system are relatively few, conspicuous and well-ordered: pick up hammer, place point of nail to wood, draw hammer back, speedily bring hammer to head of nail, repeat last two operations until nail's head is flush with wood's surface. The substructure of this system is such that even a child or chimpanzee can learn it. There are few possibilities for feedback nuances which demand critical deviations from the obvious course of events. There is considerable systematic laxity such that mistakes can be made (e.g., missing the nail) or deviations from the norm can occur (e.g., striking at different rates or with varying force) without a concomitant systemic breakdown. Nonetheless, the system's success will depend upon reasonably correct performance and ordering of its integral subsets.

Similarly, we can liken a "relatively complex system" to the performance of a delicate piece of brain surgery. Its system subsets include all the activities necessary to anesthetize the patient, enter his brain case, etc.,etc., etc. The substructure of this system is ultimately almost as simple as driving a nail but is more varied. Again, a child or chimpanzee could probably learn to perform each integral activity upon which the orderly system of brain surgery depends. It is the critical importance of veiled feedback, the systematic rigidity of operations,and the complexity of relations and order between simple components that prevents either child or chimp from hanging out his shingle.

Thus, nail-driving and brain surgery depend equally upon interrelations and ordering of integral subsets. The fact that one activity is simple while the other is complex does not relate so much to the simplicity of ultimate subsets but to their variety and relationship complexity. Surely the same is true of all biologic phenomena.

Thence, according to the present view, all aspects of biology must be ultimately comprised of complexly related aspects of the physical world which govern matter and energy. The latter aspects can be mathematically portrayed. Thus, all biologic aspects should be potentially mathematically portrayable. Mathematic models of biologic phenomena are efforts directed toward this goal.

Digitized by the Hunt Institute for Botanical Documentation

Systematics, Biology, Numbers, and Froelichia

This theoretic viewpoint, that the most complex and elusive biologic matters simply must be ultimately reducible to a network of simple, mathematically expressible properties has materialized at several levels and aspects of work. Through the confidence and support of the offices of Systematic Biology, National Science Foundation, we have been enabled to study a large number of populations of Froelichia, to collect representative samples, and to analyze these materials (with the aid of competent research assistants) via various approaches. Just as some geneticists have been more interested in genetics than in fruit flies, we generally have been more interested in what Froelichia can suggest about evolutionary systematics than in Froelichia per se. We have used this inherently "difficult" taxon mainly to study mathematic models and analyses which are to us of considerable interest. Some of this work on Froelichia has been reported and is the primary or subsidiary subject matter of the following papers:

1965. Davidson, R. A. Photoperiodism in Froelichia floridana (Nutt.) Moq. (AMARANTHACEAE). Ecology 46: 520-524.
Synopsis: Studies suggest that flower initiation in plants from Iowa and Texas populations of Froelichia floridana (Nutt.) Moq. is promoted by short-day photoperiods and retarded by long-day photoperiods. However, the degree of long-day retardation is less for northern plants than for southern ones. These differences probably are directly adaptive and seem to relate most directly to reproductive timing phenomena which is governed by differential environmental conditions that occur before or after the summer solstice (southern populations and northern populations, respectively). The photoperiodic-ecologic relationships for F. floridana are different from any known to have been previously reported and yet seem basic enough to have an unrecognized general importance in plant distribution and evolution, including speciation. The methodology and rationale of an approach to the investigation of this phenomenon is outlined.

1966. Davidson, R. A. & R. A. Dunn. A new biometric approach to systematic problems. BioScience 16: 528-536.
Synopsis: During the past few years the forefront of systematics has been occupied, in part, by efforts to devise mathematic models which will aid in the objectivization of taxonomic classification, analysis, and interpretation. The most widely publicized methods, operating within a concept of equal character weighting, have been severely criticized from several mathematic and biologic viewpoints. This paper gives an original analytic mathematical model for classification problems which, by its integral operations, weights characters in accordance with demonstrated probability distributions. Examples are provided which utilize data from a computer-simulated evolutionary sequence and from a "real-life" problem of plant migration and derivation in the genus Froelichia. The model provides for the preferential acceptance of pertinent hypotheses [the correctness of this choice has recently been substantiated by chromatographic information not included in the original data operated upon by the model]. The model's methodology is outlined in language which attempts to relate logically concepts and methods from both biology and mathematics. We feel tnat this approach may have foundations more theoretically sound than several

Digitized by the Hunt Institute for Botanical Documentation

alternative (and also yet relatively untested) methods that currently are available.

1967. Davidson, R. A. & R. A. Dunn. A correlation approach to certain problems of population-environment relations. Amer. Jour. Bot. 54(5): 529-538.
Synopsis: A model which visualizes a species as a set of biotype subsets is useful when considering the degree of fitness and homogeneity of populations in relation to environmental stress. Under founder principle concepts we expect that the number of such subsets tends to decrease as local populations encounter new selective forces in environments and habitats other than those occupied by predecessor populations (assuming no hydridization). Relations between biotypes, considered from the point of view of "variation associated with character correlation" or correlated response and polygene systems, implies that the magnitude of product-moment correlations between characters may be used as a general measure of selective stress impinging upon populations. Magnitude is defined as the per cent of the total possible pairs of characters which correlation coefficients are significantly different from zero. The direction of magnitude expected under the founder principle is from low in an undisturbed or ancestral situation to high in an alien one. This hypothesis is demonstrated in the genus Froelichia and its relation to more inclusive evolutionary ideas is discussed.

1968. Davidson, R. A. Non-Mendelian systematics. In, R. Nardone (ed.), Mendel Commemorative Symposium. In press. Catholic Univ. of Amer. Press, Washington, D. C.
Synopsis: The visible impact of Mendelism on taxonomic practice has not been as extensive as was once expected. It is at the level of foundational taxonomic philosophy, interpretation, and extrapolation that genetics has had its main effects. These conclusions follow simply from the fact that while systematic relationships must be dependent upon the total genetic histories of organisms, most taxonomic questions unfortunately cannot be effectively answered by presently available genetic methodologies. While hope decreases for an all-inclusive "objective" taxonomy via genetics, the demands of progress insist on an increase in the scientific rigor of systematic practice. These events seem mainly responsible for the widespread interest in the new interdisciplinary development of biometric systematics. This paper argues the theoretical correspondencies between population genetics, phylogeney, and parametric multivariate statistics. Research on the genus Froelichia is presented to illustrate, corroborate, and extend concepts based upon these correspondencies.

1968. Davidson, R. A. & R. A. Dunn. Computer simulation of certain forms of evolutionary change: a preliminary report. Taxon 17: 3-10.
Synopsis: A 3-dimensional model based upon a modified discriminant function analysis is considered to be a valuable aid to concepts of evolutionary relations between Froelichia species. To test whether

Digitized by the Hunt Institute for Botanical Documentation

the model's configuration is a reflection of multivariate similarities and differences rather than merely an artifact, a computer program was designed to simulate the kind of evelutionary change which characterizes Froelichia and many other "closely related" taxa. A 3-d discriminant function model based upon five simulated taxa accurately portrayed the relationships expected by programming. This result, plus others not cited here, are taken as initial support for the validity of such models when applied to real biologic data.

1968. Dunn, R.A. & R.A. Davidson. Pattern recognition in biologic classification. Pattern Recognition. In press.
Synopsis: In lieu of any knowledge of an a priori classification, biologic objects of the environment (viz. higher plants and animals) are independently classified by certain diverse human cultures and by some animals with a remarkable uniformity which can be interpreted to reflect evolutionary relationships between the objects classified. This phenomenon of congruent hierarchical classifications or conceptualizations suggests an inherent "mental program" selectively attained to meet a survival-oriented demand. Although envisioned, this mental program as yet is operationally unknown and mysterious. Since the procedure is undefined, it is of theoretic interest and practical importance to examine the same kind of classificatory results as obtained by methodology that is specifiable. Temporarily suspending the direct search for the "mental model of biologic classification", this paper discusses the attributes of a modified discriminant function approach to evolutionarily reflective taxonomies. The procedure is applied to two sets of data, one pertaining to species of the genus Froelichia, the other to Centaurium.

This was an invited paper and is one which hopefully will bring various theoretical and practical problems of systematics to a reading audience comprised mostly of engineers, computer scienticists, and others interested in the fundamental aspects of pattern recognition. We may note here tersely that at one sense and level, systematics is ultimately a complex problem in pattern recognition.

Digitized by the Hunt Institute for Botanical Documentation

The following dissertations also are Froelichia-based or utilize Froelichia data to exemplify mathematic theory:

1965. Dunn, R. A. Mathematical methods in discriminatory analysis and some applications. M. A. Dissertation. The Catholic Univ. of America Libraries, Washington, D. C.

This is an important "pure math" consideration of some of the problems inherent in discriminant function approaches to biologic discrimination. Froelichia data is used to illustrate various aspects of "power" within discriminant function methods.

1967. Kolli, Sarveswararao. Biochemical and floral anatomical studies in the genus Froelichia (AMARANTHACEAE). Ph.D. Dissertation. The Catholic Univ. of Amer. Libraries, Washington, D. C.

This is a particularly valuable study in that it presents biochemical profiles (phenolic compounds) of Froelichia populations that were previously analyzed by various multivariate methods. The correspondencies between mathematically and biochemically-based assessments of relationships within and between populations is quite remarkable in some cases. Kolli's study extends one of my own (unpublished) that was conducted under the direction of the late Dr. Ralph Alston.

Other papers and dissertations which are aspects of our total research effort can be noted in Appendix I, Biographic Information, Principal Investigator. It may be of some slight interest to note that the trend of work in our laboratories has largely been an outgrowth of the extremely puzzling and "mentally overpowering" complexity of Froelichia variation. It is too early for a reasonably valid assessment of the worth of this work, some of which is rather bizarre; however, it has, we feel, attracted a relatively wide interdisciplinary range of apparent interest.

Digitized by the Hunt Institute for Botanical Documentation

Cybernetic Systematics as Dependent upon Phylogenetic Systematics

As evidenced by the preceding synopsis of the paper, "Pattern Recognition in Biologic Classification", at some point there began to develop an idea in which already we have invested quite heavily and can only vaguely see a reasonable ending point (as long as matters seem promising). The idea can be labeled "Cybernetic Systematics". This is merely a shorthand way of alluding to the goal of developing a model of biologic classification that is so general and so definitively specified that it can be used to formulate computer "software" (synomously, programs) which will enable the computer not only to arrive at the same taxonomic conclusions as does man, but, additionally and more importantly, to do so via processing procedures which can be argued to be identical (or reasonably so) with those used by man's intellect.

From this point of view, man's classifications of his biologic world are themselves biologic phenomena, or "phenomenologic sets" in the sense of the introductory comments, and should also be susceptible to reduction to the form of a mathematical model.

According to widely accepted tenets, an initial goal in many modern systematics studies is to learn as much as is possible, or practical, about the biology of the oganisms of interest. This involves the acquisition of information related to ecology, morphology, cytology, physiology, breeding systems, chemical properties, biogeography, genetics, etc. Once this information is acquired it is collapsed, by intuition or reason, in ways such that most important and largest differences are shown by hypothesized greatest evolutionary distance implied in a natural classification judged to be appropriate. Similarly, least important and smallest differences are reflected by hypothesized least evolutionary distance implied in the same classification scheme.

As currently envisioned, the cybernetic attack on systematic theory and practice includes, and ultimately will of necessity expand (perhaps to the point of electro-mechanical photographic analysis) these probes for information. However, the sphere of interest would be extended to include how this information is coded, "packaged", and processed up to and inclu the development of relationship decisions and the attainment of hierarchical, purportedly phylogenetic, classifications.

Such an expansion of interest, if successfully pursued to the point of operational model acquisition, should help us to understand not only the processes leading to the particular kind of concept called "natural classification"[1], but also, inherently, to understand the "evolutionary meaning" of the biologic information via which such concepts are generated. At the present time no one knows how such precepts are formed. But formed they

1 The writer's views on the nature of "natural classification" are included in Davidson & Dunn 1968.

Digitized by the Hunt Institute for Botanical Documentation

are, and if the orderly relations between ultimate subsets of the system, "phylogenetic concepts", could be discovered, the computer could be commanded to do the same-- for surely the mind is, at least at some level, nothing more nor less than a biologic computer.

Elsewhere we have written (Davidson & Dunn, 1968): "A recent book (Solbrig, 1966) opens):
'Approximately half a million species of living plants and twice as many species of animals have been found and described to date...' The statement implies the following historical sequence: 1) Hundreds of billions of living organisms must have been perceived by men to exist 'out there.' 2) By some intellectual process, which must involve generalization, a vastly smaller number of sets, each called a species, has been found to subsume groups of these biologic objects. 3) By dint of classification, relationships between members of these sets also have 'been found'. This classificatory development may fail to impress the mid-twentieth century biologist who expects much more from his colleagues and their sophisticated machinery than the 'merely descriptive' sort of science inferred. However, if he decided to create a machine that would duplicate man's achievement of classifying the organic world--then he would become impressed, if not overwhelmed. Though the human activity is taken for granted, he would find no broadly unified, operational theories nor comprehensive set of algorithms that confidently could be programmed into his machine even though other components were technologically available. What has been done [and continues to be done both with and without data from critical experimental procedures] would seem almost impossible to duplicate mechanically for so many essential concepts remain fuzzily defined."

Unfortunately, perhaps, the view is much greater even than this. Although I have not as yet been able to argue matters convincingly, I am tentatively satisfied that the following points are somehow interrelated and valid:

1) All data entering into ideas that pertain to natural classifications (including those available to Darwin) are equally important (e.g., flower color is "as important" as is reproductive incompatibility) --until the probability distributions of these data (as "calculated" organism to organism) are learned. This learning is one of the first steps leading to an ultimate weighting of characters. The probability distributions of characters will be probabilistically, functionally related to the actual phylogenies of the organisms considered. These probability distributions give rise in our minds (as in, e.g., Aristotle's who had concepts of "low" and "high" organisms) to ideas of "evolutionary primitiveness vs. advancedness", and ultimately are incorporated in hierarchical biologic classification.

Digitized by the Hunt Institute for Botanical Documentation

2) All concepts are hierarchical structures of relationships.

3) All learning is a matter of relationship understanding; thus, learning is concept formation.

4) While hierarchical,"natural" biologic classifications are concepts based upon learned information distributions, their structure is mediated by an inherent mental "program" selectively acquired to produce this particular kind of classification instead of other kinds[2]--of which a nearly infinite number theoretically exist. Other conceptual structures can, and do, of course, exist, but these are of basically different kind: their defining properties are given a priori. Such is not the case in biologic classification. There are no "signs from God" available to Aristotle, Linnaeus, or modern man which tell, e.g., that whales are more closely related (phylogenetically closer, more essentially similar, etc.) to cattle than they are to fish. There is no complete fossil record known which "proves" that succulent euphorbs are "nearer" to milk purslanes than they are to cacti. Yet there has been a historical, nearly universal agreement as to the correctness of these and similar ideas. Thus, biologic classifications of the kind man judges as "probably phylogenetic" are particularly "pure" forms of complex conceptualizations which depend originally upon a totality of non-preselected information (selection may come later).

Thus we arrive at the strong possibility that the exercise of biologic classification uniquely provides semi-controlled conditions via which one of man's most mysterious biologic properties can be studied: namely, his power of conceptualization. We believe that this power is directly related to the selective effects of the phylogenetic relationships between organisms in his evolutionary environment.

In our view, then, such broad topics of concept formation, learning, generalization, etc. depend ultimately upon phylogeny. Thus, it should follow that biologic systematists are, by a surprising turn of events, well equiped to study such phenemena. While these matters are the traditional domain of philosophers, experimental psychologists, educators, etc., is there any reason to believe they should restrictively stay there? There are some indications that perhaps they should not. Idea formation and learning, etc., have been discussed by some of the world's greatest scholars for centuries. There have been massive experimental attacks on these problems for decades. Yet, "It is somewhat startling to have to admit that we still know very little about the learning process." (Uttal, 1967). In spite of several brilliantly conceived, mathematically oriented models of learning, man has not yet been able to program the most sophisticated computers to play the simple game of checkers well enough so that wins actually may be expected (see, e.g., Samuel, 1967).

[2]This point has been preliminarily argued in Davidson (1967).

Digitized by the Hunt Institute for Botanical Documentation

We feel that there is the strong possibility that the cognitive psychologists, etc., may be impeded in their progress to understand the workings of human and non-human intellects by deeply entrenched experimental simplifications which have resulted in their studying the wrong things. We have said elsewhere (Dunn and Davidson, 1968): "In our view biologic classifications (those judged 'non-artifical') constitute a unique wealth of examples against which the results of operational classification models can be compared. Primacy is their major recommendation, for it must be recognized that higher animals, including man, did not develop their capacities of categorization and specification through selective encounters with triangles vs. tilted triangles, vs. circles vs. ellipses or through confrontations with letters and words [e.g., signs saying FOOD; DANGER; TIGER; SAFE; DEER; MATES JUST AHEAD]...The principal selective components of the environment were for any animal all other organisms, each presenting a unique physical stimulus pattern for perception and interpretation. When man classifies plants and animals he is classifying the kind of objects which selectively gave rise to his general ability to classify. By studying the results of man's processing of the information contained in organismal patterns, we hopefully may gain insights into how this processing is accomplished, not only with respect to organisms, but to all other objects and events."

The WHALE Model as Our Initial Approach Toward Cybernetic Systematics

The foregoing and related ideas began to materialize some two years ago. Thinking was hence directed toward the formulation of a beginning operational model. Clearly, it is imperative that the model be compatible with what is known about such diverse matters as retinal physiology, color perception, energy transfer along neural networks, learning, efficient vs. inefficient computer programming, evolution, genetics, ethnobiology, experimental psychology, mathematic logic, probability theory, sleep physiology, etc., etc., etc.

Eventually, a model which seemed theoretically required was envisioned. The next problem was to convey this model to computer programmers such that its conceptual structure could be converted to computer language. These efforts were thwarted largely by the model's complexity plus my failure to bring matters to the absolutely operational, algorithmic level required. I hence decided that I must learn to program if the model was to get off the ground. I did that and am now teaching a course in "Computer Programming for Biologists". The bulk of last semister's research effort was devoted to personally programming our initial model of phylogenetic classification. For lack of a better name, this model has been called the WHALE.[3]

WHALE, Model I, has been applied recently to a series of bacteria data that has been reported by Eugene Kennedy of our department (Auletta & Kennedy, 1966). Through a cooperative effort with a local non-profit medical research foundation, $WHALE_I$ also has been used to analyze data pertaining

---

3 An early, non-general and non-computerized version of this model is presented in Davidson & Dunn (1968).

Digitized by the Hunt Institute for Botanical Documentation

to a problem of differential disease diagnosis. Both results were encouraging and will be discussed in a subsequent paper.

The foregoing studies were chosen principally because both utilized information originally gathered in binary form: a certain bacterium either does or does not produce phosphatase, indole, reduce nitrate, etc; a certain patient either does or does not show a fever, have elevated leucocyte counts, etc.

Our next problem was to extend the generality of WHALE to permit the utilization of both this form of data as well as the quantitative, continuous measurement form generated by most biologic studies. At that time, no satisfactory method was known by which this generalization might be obtained without concomitant serious loss of information. We believe that we have now validly solved this problem and programming leading to a more sophisticated WHALE, Model II, currently is in progress.

The foregoing comments extend directly to the following topics.

Froelichia and Cybernetic Systematics

Most of the problems posed by NSF Research Proposal, "Systematic Biometry, with Special Reference to Froelichia Moench" (submitted and approved, 1965) have been resolved or clarified to a point which permits "informed speculation". A particular problem suggested in this proposal is now seen to be theoretically impossible to solve (to be the subject of a separate paper). Certain problems have been reported upon and can best be pursued in other laboratories using different research materials. The really significant problems of Froelichia which remain are those which can be approached only by specific, concentrated efforts involving extensive field and laboratory work. I hope to parcel these out to interested M.S. and Ph.D. candidates.

Probably less than 25% of "interesting" information currently available on Froelichia is now in print. Until the development of the WHALE model, a formal revision of Froelichia was assigned a low priority rating and was intentionally deferred. There were two related reasons for this: 1) Any study of methology using Froelichia data had the potentiality of refining our understanding of various biological and systematic relationships included within the genus; 2) any study more general than Froelichia was, ipso facto, considered "more important" than Froelichia. The general areas of pattern recognition, systems control, multivariate analysis, biologic models, etc. are extremely fast moving and there is pressure to keep up with the ideas of others and to record one's own. The general conclusions outlined in a recently noticed Russian book (Yablokov, 1966) are remarkably reminiscent of those found in Davidson & Dunn (1967). However, the gestation period currently in effect for WHALE has demanded that the revision of Froelichia be given top priority. The reason for this is simply related to our forthcoming attempt to analyze all available data pertinent to a phylogenetically reflective classification of Froelichia via WHALE$_{II}$, and subsequent versions of the WHALE model. If the WHALE analysis can classify these materials, specimen by specimen, population sample by

Digitized by the Hunt Institute for Botanical Documentation

population sample, species by species, subgenus by subgenus, in such an intuitively phylogenetically reasonable way, our case will be far stronger if we can relate the WHALE analysis to a published (or 'in press') revision than if we are able only to refer to "what we think".

This point touches upon a cluster of very complex matters that relate to the problem of erecting and testing models of the kind we seek. We cannot adequately sum up the whole story as we see it, but certain points can be made without getting into excessive semantic difficulty.

The WHALE model incorporates a small number (perhaps 25) of integrally related calculational, basic proceedures. Each of these major steps has perhaps a half dozen available options any one of which will do approximately the same thing. The only rule available for the selection of the basic proceedures and options has been the logic that has seemed to accrue from the various scholarly disciplines which WHALE attempts to conceptually include. This logic has been in no case clear-cut and absolute. We have merely tried to "outguess" matters at a multitude of ambiguous points of departure. (Clearly, then, the element of luck,or near luck, is ever present in the development of the model.)

More importantly, we must note that the only way that we can ever hope to evaluate the validity of the model, whose stated purpose is to organize germane information into phylogenetically reflective taxonomies, is to compare its results with those provided by non-computer means.

For this reason, regardless of the extent to which we are carried into the realms of mathematics and computer technology, we can never foreseeably release our reins from the "biosystematics" of organisms eventually to be considered by the computer. Any less inclusive view would be inherently erroneous and would immediately doom the pioneering facets of our work to failure. In problems such as those we are attempting to resolve, there are clearly two kinds of failure possible: actual and apparent. It is the latter kind that is most fearsome. If something is no good, it simply is no good. Fine! If something seems no good, this evaluation stemming from a related view that is in itself wrong, we are in real trouble.

Thus, if WHALE does classify phylogenetically, we can think that it does so only if our ideas on the phylogeny of organisms mechanically classified are firm and correct. We cannot expect this unless our biologic approaches to classification are as complete and well-informed as possible. If WHALE is a valid model we can expect it to classify "better" than does man solely because the model will not forget, and because it is not subject to "psychological whims". However, the only way "better" could be evaluated (or, for that matter, distinguished from "more poorly") would be to engage in subsequent biologic studies.

If WHALE produces results for Froelichia that deviate widely from those we think are true, we will need to unravel the program's logic in an attempt to discover an error. Perhaps other operations will need to be added, perhaps some must be deleted, perhaps other options must be chosen.

Digitized by the Hunt Institute for Botanical Documentation

Quite likely the amount of information provided will have to be increased --even up to the point of somehow imputting "what the plants look like". (This in itself is a giant problem.) Perhaps, instead, we will need to look more critically at biologic aspects of materials in an effort to discover new features which would tend to support the model's unexpected results.

Some readers of this proposal may be tempted to suggest that a model of biologic classification might better be developed utilizing data independently gathered by workers in other laboratories. This practice would eliminate bias. Most of the refutation of this only illusionarily sound idea is contained in the preceding paragraph. Summarily, only intimate familiarity can provide the "feedback" link between the developing conceptual model and materials. Only this feedback can trigger the idea, "the model may need a change perhaps here...." or "perhaps if we study this population more intensely...." Etc.

We hope to develop gradually a completely general model via materials well known to us. If we can arrive at a point of complete satisfaction with the model in terms of theory and personal practice, we can then seek critical tests via cooperative efforts with outside workers. These logically would include Drosophila experts.

Specific Support Requested. I. Froelichia.

I hope that the preceding commentary provides a fairly comprehensive view of our "systematics philosophy". We firmly believe that the practice of systematics is an almost incomprehensibly complex business. It is this complexity that has engendered the rather meaningless concept of "taxonomic mystique". On the other hand, this concept summarily embraces a tremendous range of problems that we consider to be of great potential singificance. These are the problems that gradually have become our major concern. It seems obvious that these issues can be reasonably attacked only by "wearing many hats". The funds specifically requested are those which will permit me to wear a "hat of biology and systematics" but concomitantly to concentrate most of my time and energies mainly within the realms of mathematics, cognitive psychology, computer programming and technology, etc. I simply cannot maintain learning and productivity in all these areas that must go on all at once if our research posture is to be furthered.

Through current generous NSF support, I was fortunately able to employ, last September, Mrs. Carolyn Unzicker. I have known Mrs. Unzicker since 1954 when we both were graduate students in the Department of Botany, State University of Iowa. She received her M.S. degree there in 1956, concentrating in cytology under the direction of Dr. Henry Dean. Her husband has been empolyed as a physicist at the Naval Research Laboratory, Washington, D. C., since 1955. She has longed to be productively involved in botanical studies since her children became of school age. Mrs. Unzicker has proved her ability, sincere interest and loyalty during the past months. She has "no genes" for mathematics and could not care less about numbers. In a very real sense this is an advantage, for the funds requested in this proposal are largely to permit me to retain Mrs. Unzicker for at least a two year

Digitized by the Hunt Institute for Botanical Documentation

period during which time she can act as "the biological extension" of my feeble mind. If our work goes well and justifies continued support, Mrs. Unzicker would hope to attach herself to our program as a relatively permanent and professional assistant.

Under my supervision Mrs. Unzicker has been engaged, during the past few months, in the production of the formal revision of Froelichia. Considerable progress has been made. I believe that the manuscript can be completed within (at most) a year from this date. We will incorporate into this revision the rather large amount of diverse information that I have gathered during the past years. Gaps are becoming obvious as we progress and are being filled as they occur. The revision, if excessively large, hopefully can be publisher over several journal issues or as a small book published by The Catholic University of America Press. It is impossible to accurately predict publication costs, if such should arise. A request for supplementary funds may be submitted at a later date.

Exit Froelichia.

Froelichia has served our research purposes surprisingly well. As mentioned earlier, its very complexity was the ill wind or stroke of fortune (depending on one's viewpoint) that has carried us into rather unorthodox areas. However, there are definite disadvantages inherently present when Froelichia is viewed solely as research material:

1) Individuals of all taxa normally are self-pollinated. Occasional cross pollination may occur via ovipositing insects. Additionally, one syrphid fly has been observed to visit flowers, apparently feeding, was captured and was found to carry Froelichia pollen (morphologically very distinctive). Without much doubt these features are largely responsible for the tangled complexity of variation. A reasonably direct genetic resolution of this maze seems an excessively difficult problem for, in spite of repeated attempts, the floral manipulations required for genetic tests are invariably too severe to be withstood by plants. Flowers always abort prematurely.

2) Froelichia materials are difficult to work with cytologically. Unambiguous counts are very elusive. Detailed karyotype studies are nearly an impossibility. Chromosome numbers are uniformingly constant, apparently, throughout the genus.

3) According to Dr. Ralph Alston (deceased), Froelichia materials yield notably "messy" chromatograms. He led me through a purification procedure that is quite time consuming and only partially effective in "cleaning up the spots". In spite of satisfying results already obtained, biochemical profiles of Froelichia individuals, populations and taxa may inadvertently include more error than would be case for most plant groups.

Digitized by the Hunt Institute for Botanical Documentation

4) Froelichia is a rather non-descript plant which commands little general attention. Additionally, most of its "critical characters" are not readily visible and thus are not those which enter into "field identifications" and similar facets of taxonomic practice that engage many biologists. Both of these features mitigate against an early evaluation by peers of our revisionary efforts. Hence, even if WHALE is a valid model accurately reflecting our revision of Froelichia, probably only a few biologists over the next few years will be truly qualified to judge the validity of our revision. Clearly, the soundness of the model's structure can only be judged by the apparent accuracy of its results. Few persons will have the actual experience with materials needed to reliably confirm or deny this accuracy. This problem, inherent in all revisionary studies, is, of course, partially overcome by the information detail incorporated in the published end product.

Research Material Desiderata.

Assuming gradual progress and refinement in the computer aspects of our interests, we need research materials having rather well defined properties. These can be crudely outlined as follows:

1) The taxon should be comprised of a relatively small number of taxa of lesser rank. An important aspect of the WHALE model is its ability to "learn". Hence the theoretical structure of the model is violated if it is "told" anything other than its algorithms for learning and classification. For example, we cannot inform a computer that flower color is or is not a critical character for delimiting taxa. The computer must "decide" this for itself. Computationally active memory areas in computers are only so large and this has become a problem even in the case of $WHALE_{II}$, which I consider to be still rather primitive. If we can rewrite $WHALE_{II}$ to further and further stages of refinement, working memory storage availability increasingly will become a major problem. Hence, the "universe" (i.e. the most inclusive taxon) considered by the model must, at least for the time, be of manageable size.

2) The taxon should be restricted to continental North America. This will increase the possibility of comprehensive coverage. Herbarium specimen materials should be more or less "complete" for all taxa. Field work can be relatively efficient, detailed and inexpensive. It has become apparent that the general structure of any classification depends upon the total information contained. Gaps in information will generate classifications that in part are dictated solely by these hiatuses.

3) Materials should provide clean chromatograms.

4) Materials should be cytologically easy to approach.

Digitized by the Hunt Institute for Botanical Documentation

5) Plants should bear many-seeded fruits. A useful property with respect to certain genetic studies.

6) Plants should have short life cycles. Ditto, genetics.

7) Plants should have large, uncomplicated flowers. Ditto, genetics.

8) Plants should occur in well defined, conspicuous populations. Ditto, genetics. Also, such materials are easily spotted in the field. The danger of non-productive field work by the specialist is greatly reduced. Widespread geographic coverage can be accomplished efficiently, if this collecting approach is required. By dint of the same conspicuousness, other biologists can be expected to have been drawn unconsciously to the same materials, and, through their natural curiosity, to have become aware of some of the taxonomic relations and problems inherent in the group. Publications treating conspicuous taxa must inherently reach, in a meaningful way, a larger audience than publications treating relatively unnoticed ones.

9) Plants should not possess selfer or apomictic breeding systems. Such systems generally result in variational patterns that are overly complex for purposes currently at hand. The role of biologically important isolating mechanisms is often excessively obscured in non-outbreeding groups.

Primarily, we need materials whose variational patterns can be explained at a level as nearly directly genetic as possible. The inferential distance between phenotypically expressed patterns and underlying biologic rationale should be short and relatively direct. For example, as matters now stand, WHALE should effortlessly denote suspected hybrids. If we can biologically confirm such cases we have a strong argument. If we only can say that the computer-suspected individuals are, in fact, "putative hybrids", we have almost no argument at all.

Specific Support Requested. II. Eustoma (GENTIANACEAE).

Cognizant of the major disadvantages of Froelichia and foreseeing the kind of materials desired for subsequent work, I some time ago spoke of these problems to Dr. Lloyd Shinners. Dr. Shinners suggested that someone should look more closely at Eustoma. He has had an active interest in this group for many years and has published on it (Shinners, 1957). To date 13 population samples have been measured for 34 characters (some 3400 separate measurements have been taken). Preliminary studies suggest that all desiderata previously mentioned are met by Eustoma (cytologic studies excepted; none yet has been conducted). An additional advantage lies in that geographic areas occupied by Eustoma harbor some of the more interesting Froelichia problems which remain. Another gentian, Centaurium, has been intensively studied in our laboratories (Dunn, 1967), and the tentative decision has been made to gravitate toward the GENTIANACEAE as a major research taxon.

As soon as the Froelichia revision is completed I hope to commence a full scale assault on Eustoma. The routine of this work will be the

Digitized by the Hunt Institute for Botanical Documentation

responsibility of Mrs. Unzicker.

Summary of Specific Support Requested.

No request is made here for support of pioneering computer aspects of research. Funds are requested to support our continued biologic efforts. Work on Froelichia is to be phased out via a formal revision. The revision will bring together all facets of our work that I consider of possible interest. Work initiated toward a biochemically-, cytogenetically-, biometrically-oriented revision of Eustoma is to be continued.

Facilities.

General. --Equipment for routine procedures in chemotaxonomy are on hand. Funds are requested for supplies only. Quality optical equipment for cytology work also is extant. A modern high-speed electronic desk calculator recently was provided to me by department funds. Extensive biometric work can be accomplished by computer programs already developed. Minimal funds are requested for payment to our Computing Center for operating time. A small greenhouse and caretaker are present. Space, library, herbarium and standard laboratory facilities are adequate.

Personnel. --I am required to teach one course per semister. My research time is adequate. Funds are requested to provide me with a salary during summer periods and thus free me from teaching. While I have requested salary for only two monghs during summers, it has been my practice to devote full time to research throughout nearly four summer months. I am requesting salary for Mrs. Unzicker for two full time years. She is qualified to work with considerable independence at a professional, or nearly professional level.

Field Work. --I hope to have critical problems in Eustoma clearly delimited by spring, 1969. I believe that the number of these problems will be considerably less than was true of Froelichia and am guessing that an adequate amount of materials and information can be gathered during a single 6-week period of concentrated field effort. Funds for travel are hence restricted to summer, 1969. At least part of the collecting will be done in "circular fashion", pinpointing and collecting from populations in the bud and flower stage with return for plants in prime seeding condition.

Other Sponsors. --This proposal has been submitted to no other possible sponsors.

Digitized by the Hunt Institute for Botanical Documentation

REFERENCES CITED

Auletta, A. E. and E. R. Kennedy. 1966. Deoxyribonucleic acid base composition of some members of the Micrococcaceae. Jour. Bact. 92: 28-34.

Davidson, R. A. 1967. A cybernetic approach to classification: preliminaries. Taxon 16: 3-7.

Davidson, R. A. and R. A. Dunn. 1967. A correlation approach to certain problems of population-environment relations. Amer. Jour. Bot. 54: 529-538.

Davidson, R. A. and R. A. Dunn. 1968. A cognitive view, and a probability-oriented model of biologic classification. Symposium on the use of data mechanization and computers in clinical medicine. New York Academy or Science. In press.

Dunn, R. A. 1967. A revision of the genus Centaurium of the continental United States. Ph.D. Dissertation. The Catholic Univ. of Amer. Libraries, Washington.

Dunn, R. A. and R. A. Davidson. 1968. Pattern recognition in biologic classification. Pattern Recognition. In press.

Samuel, A. L. 1967. Some studies in machine learning using the game of checkers. II. Recent progress. IBM Jour. Res. & Dev. 11:601-617.

Shinners, L. H. 1957. Synopsis of the genus Eustoma (GENTIANACEAE). Southwest. Nat. 2: 38-43.

Solbrig, O. 1955. Evolution and systematics. Macmillan, New York.

Uttal, W. R. 1967. Real-time computers . Technique and applications in the psychological sciences. Harper & Row, New York. p. 234.

Yablokov, A. 1966. Variability of Mammals. Nauka, Moscow. (In Russian, with English summary.)

Digitized by the Hunt Institute for Botanical Documentation

APPENDIX I. BIOGRAPHICAL INFORMATION, PRINCIPAL INVESTIGATOR

Name:

Robert A. Davidson.

Birhday & Birthplace:

19 May 1927; Keokuk, Iowa.

Education:

1945. Diploma, Davenport Public High School, Davenport, Iowa.

1951. A.B., Augustana College, Rock Island, Illinois.

1952. M.S., State University of Iowa.

1957. Ph.D. State University of Iowa.

Military Experience:

1945-1947. Warrant Officer, Purser/Pharmacist Mate, U.S. Maritime Commission.

1957-1959. Administrative Officer/Research & Development Officer, Special, U.S. Air Force, School of Aviation Medicine, Randolph Air Force Base, Texas. Final rank, 1st Lt.; principal investigator, "Botanic Aspects of Pollinosis".

Positions Held:

1952-1957. Graduate Assistant, State University of Iowa.

1958-1960. Assistant Professor, State University of Iowa (replacing faculty member on leave).

1960-1961. Taxonomic Botanist, U.S. Department of Agriculture, Beltsville, Md.

1961. Visiting Lecturer in Botany, University of Wisconsin.

1961- Department of Biology, The Catholic University of America. Current rank, Professor.

Research Grants:

1960. Recipient of National Science Foundation Grant. Revoked upon entry into the U.S. Department of Agriculture.

Digitized by the Hunt Institute for Botanical Documentation

Research Grants: (continued)

1962. Recipient of National Science Foundation Research Grant: "Biometrics of variation and cytotaxonomy of Froelichia Moench (AMARANTHACEAE)".

1965. Recipient of National Science Foundation Research Grant: "Systematic biometry, with special reference to Froelichia Moench".

Field Experience:

1952-1968. United States, Mexico, Panama.

Teaching Experience:

Biostatistics

Computer Programming for Biologists

Economic Plants

General Botany

Local Flora

Numerical Methods in Systematics

Plant Geography

Plant Taxonomy (Elementary and Advanced)

Digitized by the Hunt Institute for Botanical Documentation

Societies and Organizations:

American Institute of Biological Sciences
American Association for the Advancement of Science
American Society for Plant Taxonomy
Botanical Society of America
California Botanical Society
Gamma Alpha Scientific Fraternity
Iowa Academy of Sciences
International Association for Plant Taxonomy
New England Botanical Society
Pattern Recognition Society
Society for the Study of Evolution
Society of Sigma Xi
Southwestern Association of Naturalists

Publications and Theses:

1952. The Senecioneae, Cynareae, and Cichorieae of Iowa. Masters Thesis. State University of Iowa Library, Iowa City. 122 pp.
1952. The Senecioneae, Cynareae, and Cichorieae of Iowa. Proc. Iowa Acad. Sci. 60: 98-111.
1957. The flora of southeastern Iowa. Ph.D. Thesis. State University of Iowa Library, Iowa City. 755 pp.
1957. Comments on the proposed flora of Europe. Taxon 6(8): 221-222.
1959. The vascular flora of southeastern Iowa. State University of Iowa Studies in Nat. History 20(2): 1-102.
1960. A proposal related to taxonomic research materials. Taxon 9(7): 197-198.
1960. Plant communities of southeastern Iowa. Proc. Iowa Acad. Sci. 67: 162-173.
1960. Botanic aspects of pollinosis in the area of San Antonio, Texas. School of Aerospace Medicine, Brooks Air Force Base, Texas. 59 pp.
1961. Preliminary review of the genus Lycopersicon. U. S. D. A. Misc.Publ.
1961. Variance in herbarium specimen identification and other considerations based upon the preparation of a local flora. Rhodora 63(752): 207-214.
1961. Preliminary reports on the flora of Wisconsin. No. 45.Amaranthaceae-Amaranth Family. Wisc. Acad. Sci., Arts & Letters 50: 75-87, J. Sauer, 1st author.
1963. Initial biometric survey of morphological variation in the Cirsium altissimum--C, discolor complex. Brittonia 15(3): 222-241.
1965. Photoperiodism in Froelichia floridana (Nutt.) Moq. (AMARANTHACEAE). Ecology 46(4): 520-524.

Digitized by the Hunt Institute for Botanical Documentation

1966. A new biometric approach to systematic problems. BioScience 16: 528-536. Rosalie A. Dunn, 2nd author.
1967. A cybernetic approach to classification: preliminaries. Taxon 16: 3-7.
1967. A correlation approach to certain problems of population-environment relations. Amer. Jour. Bot. 54(5): 529-538. Rosalie A. Dunn, 2nd author.
1967. IOPB Chromosome number report. Taxon 16(1): 65-66. J. Mangaly, 1st author; Rosalie A. Dunn, 3rd author.
1968. Non-Mendelian systematics. In R. Nardone (ed.), Mendel Commemorative Symposium. In press. Catholic University of America Press, Washington, D. C.
1968. A cognitive view, and a probability-oriented model of biologic classification. Symposium on the use of data mechanization and computers in clinical medicine. New York Academy of Sciences. In press. Rosalie A. Dunn, 2nd author.
1968. Computer simulation of certain forms of evolutionary change: a preliminary report. Taxon 17: 3-10. Rosalie A. Dunn, 2nd author.
1968. Pattern recognition in biologic classification. Pattern Recognition. In press. Rosalie A. Dunn, 1st author.

Dissertations Directed (Catholic Univ. of America)

1966. Graffe, A. J. A study of morphological variation within the Cirsium altissimum-Cirsium discolor complex. M. S.

1966. Hebert, H. A revision of the genus Carphephorus (COMPOSITAE). M. S.

1966. Mangaly, J. K. A cytotaxonomic study of the herbaceous species of Smilax: Section Coprosmanthus. Ph. D.

1967. Dunn, R. A. A revision of the genus Centaurium of continental United States. Ph. D.

1967 Kolli, S. Biochemical and floral anatomical studies in the genus Froelichia. Ph. D.

1967. Steiber, M. A flora of Ann Arundel County, Maryland. M. S.

1968. Bellmer, E. H. Distribution, variation, and chromosome numbers in the Appalachian shale barren endemic Eriogonum allenii Watson. Ph. D.

Digitized by the Hunt Institute for Botanical Documentation

Budget for Two Years

		First Year		Second Year		Total
Personnel:						
Salaries -						
Robert A. Davidson, Ph.D.						
Prin. Investigator -A.Y. 20%	(A)	$ 2,600.	(A)	$ 2,600.	(A)	$ 5,200.
- Summer - 2 mos.		2,889.		2,889.		5,778.
Research Assistant - 100%		7,200.		7,500.		14,700.
Fringe Benefits -						
Pension Premiums	(A)	172.	(A)	172.	(A)	344.
		216.		216.		432.
Social Security Taxes	(A)	69.		69.		138.
		316.		330.		646.
Total Personnel		$13,462.		$13,776.		$27,238.
Travel:						
Field work - 8,000 mi. @ 10¢		800.		-		800.
Subsistence - 45 da. @ $10 - ea.		450.		-		450.
Supplies and Materials:						
Chemicals for cytology, chemotaxonomy; photographic and copying supplies & services; shipping costs for herbarium materials, etc.		200.		200.		400.
Books:		100.		100.		200.
Data Processing ($35.00 per/hr.)		300.		500.		800.
Total Direct Costs		$15,312.		$14,576.		$29,888.
Indirect Costs:						
51.91% of direct salaries	(A)	1,350.	(A)	1,350.	(A)	2,700.
		5,238.		5,392.		10,630.
Total Costs		$21,900.		$21,318.		$43,218.
(A) C.U. Participation						
Funding:						
National Science Foundation		$17,709.		$17,127.		$34,836.
Catholic University of America		4,191.		4,191.		8,382.
Percentages:						
National Science Foundation		80.9%		80.4%		80.6%
Catholic University of America		19.1%		19.6%		19.4%

Digitized by the Hunt Institute for Botanical Documentation

18 January 1968

Dr. David M. Prescott
Editor, BioScience
PSR Building 1
Campus

Dear Dave:

Goodall's paper is in good shape. I recommend that it be published essentially without change. However I hope he can do something to improve the illustration. The printout which he photographed is much too small to be legible.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm

Digitized by the Hunt Institute for Botanical Documentation

BioScience

Scientific Editor: DAVID M. PRESCOTT
Executive Editor: ROBERT S. LEISNER
Managing Editor: MARGARET U. CHAMBERS

AMERICAN INSTITUTE OF BIOLOGICAL SCIENCES

3900 WISCONSIN AVE., N.W., WASHINGTON, D. C. 20016 • 202—362-6733

January 11, 1968

Dr. David Rogers
Biology Department
Campus

Dear Dave:

Could you help me with evaluation of the enclosed manuscript by D. W. Goodall entitled "Identification by Computer" submitted to BioScience? Is the material of sufficient interest and substance to merit publication?

With thanks,

Sincerely,

David M. Prescott bm

David M. Prescott
Editor, BioScience

DMP:bm
Enclosure: manuscript
CC: Dr. Robert S. Leisner

Digitized by the Hunt Institute for Botanical Documentation

JAN 12 1968

Yale University *New Haven, Connecticut 06520*

DEPARTMENT OF BIOLOGY

Osborn Memorial Laboratories

January 10, 1968

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

Dear Dave:

I appreciate very much your thoughtful comments concerning people who think they know something about numerical taxonomy in general, and your specific reviews of the two papers in particular. You need not apologize for your tardiness, actually, you had two of them and the holidays got in the way besides. Furthermore, I don't know many others I can "parasitise" for reviews of these types of papers.

I note your remark that you "agonize over these reviews." I appreciate that, because this means to me that you have put a great deal of thought into them. Papers on numerical taxonomy are very traumatic for me, since I am not particularly knowledgeable in this field. On the one hand, I don't want to allow a lot of "garbage" into BRITTONIA, but neither do I want to give the impression that we just won't accept any such papers.

Again let me say that I am very grateful for your help, both this time and in the past. It bothers me that I find it necessary to ask your help so often, but I really don't know what else to do. Could you perhaps give me the names of a few others whom you think might be competent judges?

All good wishes,

John

John R. Reeder
Editor-in-Chief
BRITTONIA

JR:sc

Digitized by the Hunt Institute for Botanical Documentation

5 January 1968

Dr. John R. Reeder, Editor-in-Chief
BRITTONIA
Department of Biology
Yale University
New Haven, Connecticut 06520

Bear John:

Finally herewith, the two papers you sent me so long ago. I could claim all sorts of excuses for not returning them sooner, but you've heard 'em all.

I do agonize over these reviews. I get several of my crew to read them and we have discussions about the merit, correctness, value, etc. One of our troubles is that we've been doing these things long enough to know, at least, that most of the problems exist, and have developed some background which many of the people first getting on to the use of computers seem to discover in these papers. All this is in way of saying that it is difficult not to merely wave a hand and say that, from our Olympian seat, your paper isn't worth doing. But this won't help a bit.

I do wish that some of our colleagues would be a little more sophisticated about this work, and if they don't know how to answer some particular statistical or mathematical problem themselves, they would take the necessary steps to communicate with a sympathetic mathematician. For in this work there are at least three separate disciplines to master if the machinery is ever to be useful, and many of these papers are really the results of no mastery of any of the three.

First of all, the authors of the Cucurbit paper apparently think that the various methods now employed for numerical taxonomy were specifically designed for numerical taxonomy, and unfortunately, this simply is not the case. The authors have not thought out what each of the various statistical methodologies were intended for in the first place, and whether or not they can be forced to apply to their own work. The are not alone - many others "accept" some statistical procedure or other without knowing its function. This is a rather difficult task, I admit, and one that I learned only after some rather hard-won experience, but learn it I did, and they should too. Really nothing is contributed by taking methodologies and sort of "shooting crap" with them. Maybe, the attitude seems to be, if we take enough of these things, we can build a bridge with them. If the results look "close to" our feelings, then they must be "right. But for such work, this is costly, time-consuming, and not scientific.

Digitized by the Hunt Institute for Botanical Documentation

Secondly, the paper by the Bobisuds (on computer key construction). It has some problems, as you'll see intthe enclosed comment sheets, but there is an even greater one that we think should be considered - in the long range of things. First of all, most of the work here for keys intends to use the computer as a sort of substitute for the human, or to put it another way, to make the computer do a key the same way a human does - that is, a sequential processing of the key elements. This is terribly wasteful of t the computer's time, and of the human who must attend the hardware while this type of key is running. Computers can do identifications in a different manner than we can, and more efficiently, so we should be addressing the development of computer based identification systems to the simultaneous consideration of characters, rather than a sequential consideration. This idea is not new with us, but is in the thinking of several people, Sokal among them. We will have a paper out on the subject fairly soon, but in such a sophisticated way that it will have to be put into the Journal of the Association of Computing Machines. It will be spelled out for biologists, probably in Creator.

I hope that I have said some meaningful words hereffor you. I want to encourage people to work with computers, and I certainly want to help you get them into Brittonia, if at all possible. But I did want to get some general statements to you, not specifically to the authors of these papers.

Sincerely,

David J. Rogers
Professor of Biology

DJR:gm
Enc.

Digitized by the Hunt Institute for Botanical Documentation

Critique of Bobisud: Computer Construction of Taxonomic Keys

The authors propose a "methodology for key construction suitable for either hand use or use on a computer" (p. 1). The also construct an example using the methodology.

This paper is weak. First, it is not constructed from a systematic point of view.. Perhaps this represents some of the traditions of the field in which they work. Nonethe less, there is no theoretical framework which can be used to justify the "methodology." Accordingly the authors present an example of a small study.

Two of the concepts with which the author deals are considered with "black or white" precision. Characters are either "correlated" or "not correlated." For a group of ten objects some characters may correlate, but for studies of hundreds of objects, characters may "almost" correlate, but perfect correlation is very imporbable. In such cases this function "Total Value (cf. p. 7) will be taken over only one character. This is not necessarily a weakness but is to point out that the authors' sampel study may not be representative of larger studies.

Also, "Total Value" which determines the order of the characters in the key, could be improved. As it stands, the "Total Value" of several "lousy" correlated characters could exceed the "Total Value" of one excellent character. This situation could be remedied by constructing Total Value in a different way (see last part of this critique).

The point of the counter example is not that Bobisud's method is "incorrect" but that if heuristic numerical methods are to be used, the authors have not picked the best possible ones.

Finally, there is a cut and dried distinction made between ACTIVE and INACTIVE characters. But if a character varies negligibly (say only over 1% of the objects, and call it "INERT") then to use it (and its correlated characters) as a splitting character, would give negligible information in the keying-out process; [The "information" gained thereby would be very close to zero.] hence would cause a longer key to be constructed. Although traditional key constructors may ignore criticism from this viewpoint, some modern authors point to this as a desirable criterion to make the keying-out process shorter and more efficient. [See, for example, Osborne, New Phytologist, Vol. 62, pp. 144-160 (referred to by Shetler).]

It is to be hoped that future papers on utilization of the computer to construct keys may be more sophisticated, if not more systematic.

Digitized by the Hunt Institute for Botanical Documentation

Appendix

Define "Groupvalue" of a group of correlated characters K_1, K_2, ... K_n, on an object as the maximum of the values (over all characters in the group) for the given object.

i.e. Groupvalue (K_1, K_2,... K_n) = $\max_{K_1, K_2, \ldots K_n} \text{Val}(j,i)$

Groupvalue may be calculated for each object of the study. Then of any two groups of characters, G_1 and G_2, the number of times Groupvalue (G_1) exceeds "Groupvalue" (G_2) may be computed. If "Groupvalue" of one group exceeds "Groupvalue" of another group more than half the time, it is the better group of correlated characters.

Example: 4 correlated characters most of which have val = 2
1 character with val = 5

Then the above scheme picks out the 1 character in preference.

Digitized by the Hunt Institute for Botanical Documentation

Rhodes, et al., Cucurbita

The paper is, by current standards in this area, well-written. It should be published, in essentially its present form.

I am sorry that it does not have some of the following in it. First, there is no discussion about the structuring of the input data, and though the authors state that a list of the characters employed may be obtained by writing to someone, this does not tell anyone how the authors conceived of or structured the information which the various techniques were used on. Clearly, one may influence the various techniques by the way the character is employed, coded, etc.

Secondly, this type of testing of a few methods of similarity (or distance, or dissimilarity) measures and a few clustering techniques does not tell someone whether a particular method is "good" and therefore should be used to the exclusion of others in the process of classification. Many other types of similarity can be generated, and many other clustering techniques have been proposed. It is important that the authors know why a particular method does what it does mathematically. I do not know, as a result of reading this paper, whether any one of the several tested methods is more useful than others, except by default. For example, it seems that (p. 1, sentence in Abstract beginning on line 22) if the "size" or "shape" of an OTU dictates the method, I am at a loss if I do not know in advance what "size" or "shape" my OTU is. Therefore, is it not reasonable to say that none of the methods here employed is useful because they insist on this apriori knowledge?

But this does not say that "numerical taxonomy is no good" because the wrong methods are used. It should send the workers to seek assistance from mathematicians. That is a difficulty, which no one will deny.

Digitized by the Hunt Institute for Botanical Documentation

Rhodes, et al., Cucurbita

The paper is, by current standards in this area, well-written. It sh should be published, in essentially its present form.

I am sorry that it does not have some of the following in it. First, there is no discussion about the structuring of the input data, and though the authors state that a list of the characters employed may be obtained by writing to someone, this does not tell anyone how the authors conceived of or structured the information which the various techniques were used on. Clearly, one may influence the various techniques by the way the character is employed, coded, etc.

Secondly, this type of testing of a few methods of similarity (or distance, or dissimilarity) measures and a few clustering techniques does not tell someone whether a particular method is "good" and therefore should be used to the exclusion of others in the process of classification. Many other types of similarity can be generated, and many other clustering techniques have been proposed. It is important that the authors know why a particular method does what it does mathematically. I do not know, as a result of reading this paper, whether any one of the several tested methods is more useful than others, except by default. For example, it seems that (p. 1, sentence in Abstract beginning on line 22) if the "size" or "shape" of an OTU dictates the method, I am at a loss if I do not know in advance what "size" or "shape" my OTU is. Therefore, is it not reasonable to say that none of the methods here employed is useful because they insist on this aprɸori knowledge?

But this does not say that "numerical taxonomy is no good" because the wrong methods are used. It should send the workers to seek assistance from mathematicians. That is a difficulty, which no one will deny.

Digitized by the Hunt Institute for Botanical Documentation

Dec 4. 1967

Yale University *New Haven, Connecticut 06520*

DEPARTMENT OF BIOLOGY

Osborn Memorial Laboratories

November 29, 1967

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

Dear Dave:

Enclosed are the two manuscripts which you have agreed to review.

Thanking you in advance for your help, I am

Sincerely,

John

John R. Reeder
Editor-in-Chief
BRITTONIA

JR:sc
Enc.

Digitized by the Hunt Institute for Botanical Documentation