



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

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

Usage guidelines

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

Statement on harmful and offensive content

The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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of time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

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

About the Institute

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

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

Taximetrics Lab.

January 9, 1971

Dr. R. E. Schultes
Botanical Museum of Harvard University
Oxford Street
Cambridge, Mass. 02138

Dear Dick:

Sorry for the delay in returning this. The reason for the delay is that I took the liberty of corresponding directly with the Author on a number of corrections, which are now included. The author was very slow in sending me the corrections. I took this unusual step inasmuch as I knew the work first-hand, and the author personally.

The paper is a factual report of work done, and is correct in all details presented. I am sure that readers of EB will find the paper useful, and I hope that the system described, TAXIR, will become generally used by many similar endeavors.

Thanks for your patience.

Sincerely,

David J. Rogers,
Prof. of Biology

Encl.

OFFICE OF THE DIRECTOR



Aug 7/70

Dear Dave —

Would you care to act as
editorial "confident" for this article?
No hurry. Thanks,

Dick.



1
B

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE

REGIONAL PLANT INTRODUCTION STATION
WASHINGTON STATE UNIVERSITY
JOHNSON HALL 88
PULLMAN, WASHINGTON 99163

January 5, 1971

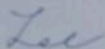
Dr. David J. Rogers
Taximetrics Laboratory
Department of Biology
Armory 101
University of Colorado
Boulder, Colorado 80302

Dear Dave:

I am sending you pages 5 and 9 with corrections as indicated in our telephone conversation of December 30, 1970. It is embarrassing to find these errors continually persisting even after proof-reading with a stenographer. I hope this will clear the thing up.

I don't know what kind of interference would produce the affect of garbage cans being thrown around, but it certainly terminated our conversation. If I interpret your last statements correctly, you were going to write a letter. If this is not what you said, please let me know.

Sincerely,



Leland W. Hudson
Horticulturist

enc.

UNITED STATES DEPARTMENT OF AGRICULTURE
AGRICULTURAL RESEARCH SERVICE

REGIONAL PLANT INTRODUCTION STATION
WASHINGTON STATE UNIVERSITY
JOHNSON HALL 56
PULLMAN, WASHINGTON 99163

December 8, 1970

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

Dear Dave:

Herewith please find correction pages for the TAXIR paper which you are reviewing for Economic Botany. If this does not fulfill the needed changes, let me know and I will get the corrections off to you, hopefully by return mail or as soon as I can make it, certainly much quicker than this time around.

I think Sam probably has told you that Willis Skrdla, of the North Central Regional Plant Introduction Station, has requested a TAXIR package for his use there at Ames, Iowa. I am not sure what crop he wants to use it for initially, but I am supposing it might be tomatoes, since that is one of the major crops. There is a type of information in this crop which we have not attempted to handle yet and that is genetic coding for genes. Gene symbols, as you know, are an accumulative mass of upper and lower case numbers, superscripts and subscripts. This is a computer impossibility and we will have to figure out some sort of a substitution for the genetic codes. Have you come up against this problem in your work there, and if so, do you have any ideas? Thank you for your kind attention, both to my paper and to these questions.

Sincerely,



Leland W. Hudson
Horticulturist

enc.

NATIONAL SCIENCE FOUNDATION

PROPOSAL RATING SHEET

Reviewer David J. Rogers

Proposal No.:

Investigator:

Institution:

Please return to:

If possible by:

PIB1583

TURNER & FLAKE

Univ. of Texas

Systematic Biology

January 22, 1971

Comments (Continue on additional sheet if necessary)

I cannot see the justification for this proposal. The problem has been worked over in considerable detail before, and while there was but a single-line transect made earlier, either the point is proved or sufficiently well-made that all that will be added is more proof. Were the work to demonstrate something new or different, or if the subject material in some other group of plants, it might be better justified.

In these times of restricted NSF budgets, I cannot help being influenced by the budget request. It does not seem to me to be realistic on the part of the investigators to expect that NSF will pay for a very expensive piece of equipment for this one project alone, nor pay the summer salaries of persons making the salaries indicated.

OVERALL RATING

- ☐ EXCELLENT
☐ VERY GOOD
☐ GOOD
☐ FAIR
☒ POOR

Signature of Reviewer:

Date: Jan. 12, 1971

Other suggested reviewers (optional):

Dr. M. T. Hall
Morton Arboretum

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. 522 (The Freedom of Information Act)."

Research Proposal Submitted to National Science Foundation

By

The University of Texas at Austin

Austin, Texas 78712

P1B1583

Title of Proposed Project: Computer-Aided Systematic Studies of *Juniperus virginiana* Populations

Co-Principal Investigators:

Name: B. L. Turner & R. H. Flake

Title: Professor & Chairman Assoc. Professor

Social Security No.: 456-20-4697 430-56-3450

Department: Botany Elec. Engr.

Date of Submission: December 23, 1970

New Proposal _____ or Renewal of NSF Grant No. GB-20713

Proposed Starting Date March 1, 1971 Amount Requested \$87,662 Number of Months 24

Endorsements:

Principal
Investigators

Department
Chairmen

Name: B. L. Turner & R. H. Flake B.L. Turner & A.W. Straiton Peter T. Flawn

Signature: <u>ORIGINAL SIGNED BY</u>	<u>ORIGINAL SIGNED BY</u>	<u>ORIGINAL SIGNED BY</u>
(BLT) Prof. & Chairman		Vice-President
Title: (RHF) Assoc. Prof.		for Academic
		Affairs ad interim
Telephone (BLT) 512 471-5262	(BLT) 512 471-5262	
Number: (RHF) 512 471-1014	(AWS) 512 471-1851	512 471-5966

Renewal Proposal for Continuance of
COMPUTER-AIDED SYSTEMATIC STUDIES
of Juniperus virginiana Populations

Abstract

New computer data-processing techniques have been developed to aid in systematic studies which have proven to be particularly useful in studies at the infraspecific level. Analysis of Juniperus virginiana populations sampled along a 1500-mile transect from northeastern Texas to Washington, D.C. using terpenoid data indicates a clinal pattern of variation in these populations. No biochemical evidence could be found to support the hypothesis that hybridization with J. ashei might be causing this variability in J. virginiana through introgression, as had been widely held.

The principal research effort proposed in this renewal application involves the extension of the analysis of J. virginiana populations from those along a single linear transect to populations sampled in a two-dimensional regional grid. This should permit the detection of peripheral perturbations through gene flow from the allopatric species J. scopulorum and J. horizontalis if, as has been reported, introgression from these taxa occurs. Further, considerable insight should be obtained as to the origin of J. virginiana and its southern segregate, J. barbadensis.

Collaborative studies with other workers studying populational problems especially in the development of data-processing techniques for treatment of quantitative populational data (in addition to the J. virginiana studies) are also envisioned.

Introduction

New computer data-processing techniques have been developed to aid in systematic studies which have proven to be particularly useful in the detection of phenotypic differentiation in infraspecific populations. Analysis of Juniperus virginiana populations sampled along a 1500-mile transect from northeastern Texas to Washington, D.C., using terpenoid data, indicates a clinal pattern of differentiation in which the most mutually similar populations appear in the Appalachian region and progressively more divergent populations were found in more distant regions as measured along the transect toward the southwest (Flake, VonRudloff and Turner, 1969). These populations were re-sampled a year after the first study and the pattern of differentiation was reconfirmed (Flake, VonRudloff and Turner, unpubl.). No biochemical evidence could be found to support the hypothesis that hybridization with J. ashei might be causing this variability in J. virginiana through introgression, as had been widely held.

J. virginiana is the most abundant and widely distributed conifer of tree size in the Eastern United States (Fowells, 1965); thus it is extant in a multitude of environmental regimes and it is easily accessible in populational form.

Prior to our populational studies of the J. ashei - J. virginiana complex (mentioned above) it was widely held that regional hybridization between these two very distinct taxa (belonging to different sections of Juniperus) constituted the best documented case for allopatric introgression (Anderson, 1953; Grant, 1958; Davis and Heywood, 1963). As indicated, this was not borne out by our studies nor those of Adams (cf. Adams and Turner, 1970).

A much more promising case of allopatric introgression is apparently that of J. virginiana with the closely related species, J. scopulorum (and/or J. horizontalis) as has been suggested by Fassett (1944a, b; 1945a, b, c) and more recently by Van Haverbeke (1968). The latter author, in particular, through his study of these taxa in the Missouri River Basin (Figs. 1 and 2), has stated that:

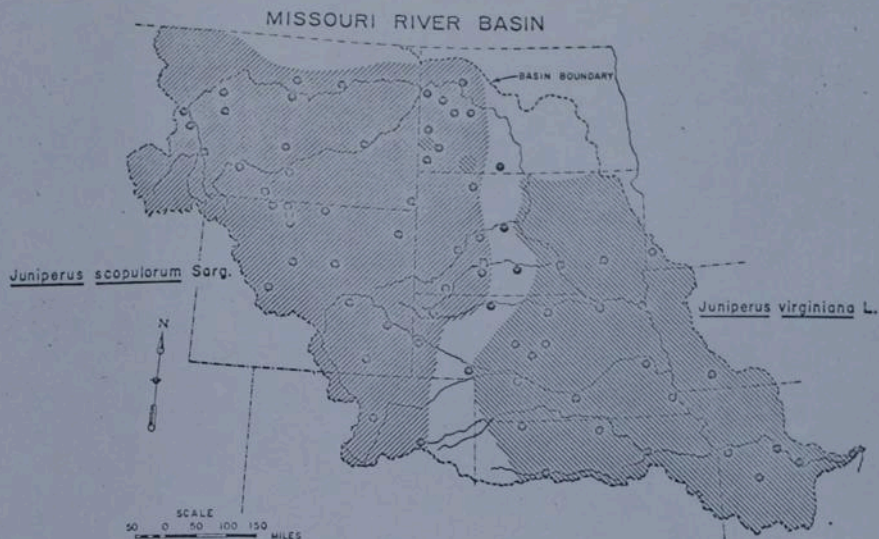


Fig. 1.—Missouri River Basin illustrating reported species ranges of *Juniperus scopulorum* Sarg. and *J. virginiana* L., major river drainages, and locations of collection sites.

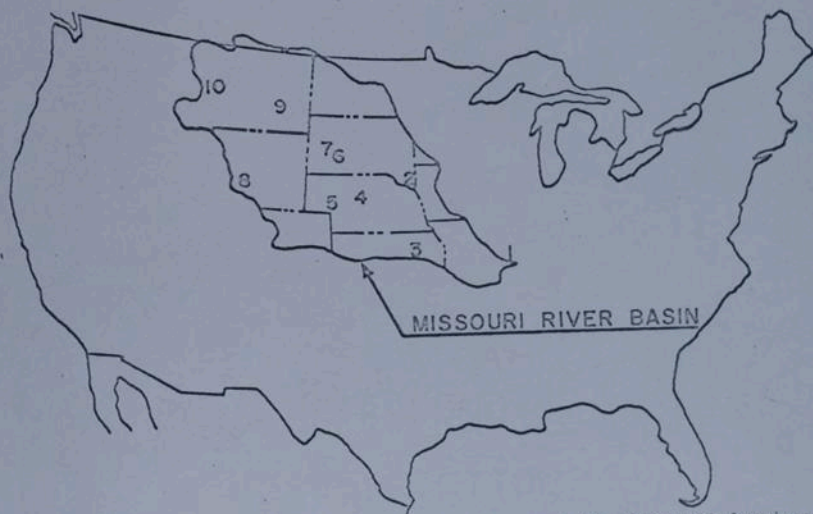


Fig. 2.—Transition in gross morphological appearance of *Juniperus* trees across the Missouri River Basin from the range of *J. virginiana* (east) to and into the range of *J. scopulorum* (west); also emphasizing the variation found within the *J. scopulorum* population throughout its range. (Numbers are keyed to a picture on following pages).

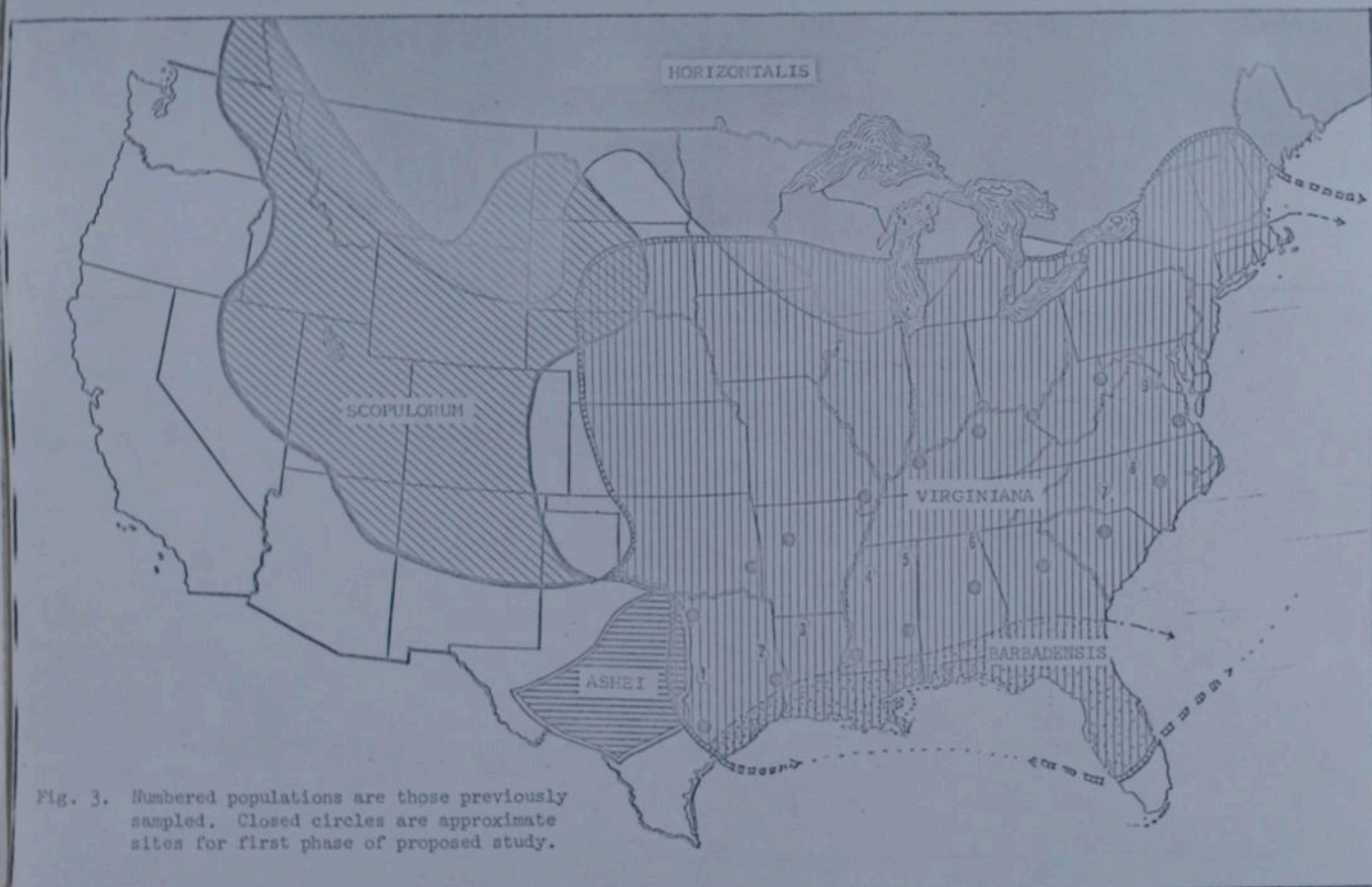
The entire *Juniperus* population within the Basin is apparently of hybrid derivation with neither of the extreme parental types being found. There is a trend of increasing hybrid index values (also percentage germ plasma values) from southeast to northwest over the Basin from the reported range of *J. virginiana* to and into the reported range of *J. scopulorum*. This condition may be the result of bilateral introgression between the two species. There was, however, a strong tendency toward bimodality within the population as demonstrated by the presence of two distributions in each of the three hybrid indices. This indicated the presence of two different species - *J. scopulorum* and *J. virginiana*.

There is apparently introgressive hybridization involving a third species (*J. horizontalis* Moench.) into the predominantly *J. scopulorum* population in Montana and northern Wyoming. This is shown, in part, by the widespread presence of depressed individuals and an increase in the magnitude of the variation in the hybrid indices toward the northwest. Also, the wide range of environmental systems present throughout this vast and varied topographic area along with the isolated nature of many of the stands (Fig. 1) undoubtedly accounts for much of the variation demonstrated within the predominantly *J. scopulorum* population.

While Van Haverbeke admits his data might be interpreted as constituting evidence for introgression between, in particular, *J. virginiana* and *J. scopulorum* he, nevertheless, suggests an alternate hypothesis:

As an alternative interpretation, it would seem that because of the greater diversity of the junipers in western North America, that *J. virginiana* was at some time derived from this area. It seems possible that with the inherent variability in the germ plasma ancestral to both *J. scopulorum* and *J. virginiana*, that propagules could flourish in sites toward the east. This could have initiated an eastward migration-propagule by propagule-which through mutation and selection eventually became what we now recognize as *J. virginiana*.

It should be noted that this latter evolutionary model is in direct conflict with that proposed by us (Flake, VonRudloff and Turner, 1969) in which we suggest that the Appalachian Region is the ancestral center for the origin of *J. virginiana* and its various races (Fowells, 1945; Hall, 1952). The numerical program using chemical data (40 to 60 terpenoid characters, each showing quantitative variation both within and between populations) which we have used successfully in our study of the *J. virginiana* - *J. ashei* complex (outlined below) should be especially suited for the kind of clinal problem suggested by Van Haverbeke.



Previous Investigations of Juniperus virginiana Populations

Previous work on Juniperus virginiana is summarized in the attached reprint and it need only be stated again here that the populations concerned were resampled a year later and the analysis of these data re-confirmed our earlier findings: that a clinal differentiation pattern was present, this being unrelated to introgression from J. ashei.

PROPOSED RESEARCH

The principal research effort proposed in this renewal application involves the extension of the analysis of J. virginiana populations from those along a single linear transect to populations sampled in a two-dimensional regional grid. The number of J. virginiana populations analyzed will be increased three-fold in this initial phase of the regional study over the number previously considered (Fig. 3). The regional study will entail the analysis of the distributions of approximately forty terpine compounds in six hundred individuals. Hybridization of J. virginiana with J. scopulorum, J. horizontalis and possibly J. barbadensis is considered probable and in either case could cause perturbations in the populational pattern detected in this first phase of the regional study.

The second phase of the proposed regional study will involve the extension of the regional sampling of the first phase either north toward areas of sympatry with J. scopulorum, J. horizontalis, or south into areas of sympatry with J. barbadensis (Fig. 3) depending on the pattern of differentiation detected in the initial regional study.

Collaborative studies with other workers studying populational problems in addition to the J. virginiana studies are also envisioned. For example, Professor J. Hanover of Michigan State University has an enormous amount of quantitative data for the terpenes in populations of various species in the genus Picea and collaborative studies of some of these data with the principal investigators of this proposal have been in progress for the past two years (Wilkinson, et al; 1971)

Unquestionable experience with the expanded study of Juniperus virginiana proposed here will permit the development of new conceptual approaches to populational problems at the infraspecific level; indeed, such studies should permit an extension of populational consideration at the supraspecific level, an approach which seems reasonable in view of the wide use of populational data by biosystematists generally (cf. Solbrig, 1970).

Literature Cited

- Adams, R. P. 1969. Chemosystematic and numerical studies in natural populations of Juniperus. Doctoral Thesis, The University of Texas, Austin, 240 pp.
- _____ and B. L. Turner. 1970. Chemosystematic and numerical studies of natural populations of Juniperus ashei Buch. Taxon 19:728-751.
- Fassett, N. C. 1944a. Juniperus virginiana, J. horizontalis and J. scopulorum I. The specific characters. Bull. Torr. Bot. Club 71:410-418.
- _____. 1944b. II. Hybrid swarms of J. virginiana and J. scopulorum. Bull. Torr. Bot. Club 71:475-483.
- _____. 1945a. III. Possible hybridization of J. horizontalis and J. scopulorum. Bull. Torr. Bot. Club 72:42-46.
- _____. 1945b. IV. Hybrid swarms of J. virginiana and J. horizontalis. Bull. Torr. Bot. Club 72:379-384.
- _____. 1945c. V. Taxonomic treatment. Bull. Torr. Bot. Club 72:480-482.
- Flake, R. H., E. VonRudloff and B. L. Turner. 1969. Quantitative study of clinal variation in Juniperus virginiana using terpenoid data. Proc. Natl. Acad. Sci. 64:487-494.
- Fowells, H. A. 1965. Silvics of forest trees of the United States. U.S.D.A. Handbook, 271, 762 pp.
- Hall, M. T. 1952. Variation and hybridization in Juniperus Ann. Mo. Bot. Gard. 39: 1-64.
- Solbrig, O. T. 1970. The phylogeny of Gutierrezia: an eclectic approach. Brittonia 22:217-229.
- Van Haverbeke, D. F. 1968. Population analysis of Juniperus in the Missouri River Basin. Univ. Nebraska Stud., Nov. Ser. 38:1-82.
- Wilkinson, R. C., J. W. Hanover, J. W. Wright and R. H. Flake. 1971. Genetic variation in the monoterpene composition of white spruce (Picea glauca). Jour. Forest Sci. (In press).

Equipment

There are two major items of equipment necessary to expedite the proposed study. The first is a large general purpose digital computer required to process the systematic data, and the second is a gas liquid chromatographic apparatus for the terpene analysis.

Digital 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.

Gas Liquid Chromatograph - It has been found that the acquisition of the precise data needed for the computer-aided analysis requires a high-resolution dual-detection analytical instrument fitted with capillary

columns and with a digital read-out integrator of the type found on the Varian 2100-40 instrument. The 600 individual samples will require over a thousand hours processing on a single GLC instrument, thus this additional instrument is needed to reduce the chemical analysis time to within feasible limits. The second year we plan to add two recorders and intregators to the instrument to double the data output rate.

Period of Performance

It is recommended that the proposed grant, if awarded, be made effective for a period of two years, beginning March 1, 1971, or some other mutually 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.

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; Associate 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: Computer-aided classification, control theory.

- Publications: R. H. Flake, "Volterra Series Representation of Nonlinear Systems," IEEE Trans. Part II, 1963, p. 330.
- R. H. Flake, "Volterra Series Representation of Time-Varying Nonlinear Systems," Proc. 1963 IFAC (Basel), 2, p. 91
- J. Zaborszky & R. H. Flake, "On Fitting with Functional Polynomials," ASME Trans., 89, Series D, 1967, p. 399.
- R. H. Flake & J. R. Cox, "Filter Design for the Average Response Computer," 19th SWIEECO Conf. Record, (Dallas), 1967.
- R. H. Flake & J. G. Hartwell, "A Reference Orbit Useful in Integrating Between Activity Spheres," AIAA J., 5, 1967, p. 599.
- R. H. Flake & J. H. Satterfield, "A Computer Classification of Psychiatric Patients," 6th S. E. ACM Conf. (Chapel Hill), 1 1967.
- R. H. Flake, "Numerical Classification of Biological Populations," (Abstract) Proc. Conf. on Engr. in Med. and Biol. (Houston), 1968.
- R. H. Flake & B. L. Turner, "Numerical Classification for Taxonomic Problems," J. Theoret. Biol. 20, p. 260.
- D. E. Stone, G. A. Adrouny & R. H. Flake, "New World Juglandaceae. II. Hickory Nut-Oils, Phenetic Similarities, and Evolutionary Implications in the Genus *Carya*," Am. J. Bot., 56, No. 8, 1969, p 928.
- R. H. Flake, E. von Rudloff & B. L. Turner, "Quantitative Study of Clinal Variation in *Juniperus virginiana* using Terpenoid Data," Proc. Natl. Acad. Sci., 64, No. 2, 1969, p 487.
- R. H. Flake, "Numerical Classification Techniques for Populational Studies at the Intraspecific Level", (Abstract) Proc. XI International Botanical Congress (Seattle) 1969.

- 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.

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: (last five years)

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)

Chromosome numbers in the Compositae. VIII. Mexican and Central American species. Southwest 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--Asteraceae). Brittonia 15: 316-331. (with D. Horne)

A taxonomic study of the genus Amphicarpaea (Leguminosae). The Southwest 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 Stenocarpa (Compositae-Heliantheae-Galinsoginae). The Southwest 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-Heliantheae). Madrono 18: 203-206.

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

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

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 canavinine 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. Cuatrecasas)

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.

1968

Miller, H. E., T. J. Mabry and B. L. Turner, 1968, Intraspecific variation of sesquiterpine lactones in Ambrosia psilostachya (Compositae). *Am. J. Bot.* 55: 316-324.

Flake, R. H. and B. L. Turner, 1968. Numerical classification for taxonomic problems. *J. Theoret. Biol.* 20: 260-270.

1969

Turner, B. L., 1969. Problems of Hybrid Nomenclature. *Proc. First Int. Citrus Symposium* 1: 451-453.

Skvarla, J. J. and B. L. Turner, 1969. Fine structure of Petrobineae (Compositae-Heliantheae) pollen walls. *Am. J. Bot.* 56: 418-419.

Turner, B. L., 1969. Chemosystematics: recent developments. *Taxon* 18: 134-151.

1970

Turner, B. L. 1970. Molecular approaches to populational problems at the infraspecific level. Chapter 10, IN *Phytochemical Phylogeny*, (Ed.) J. B. Harborne. pp 187-205.

Turner, B. L. 1970. Chromosome numbers in the Compositae. XII. Australian species. *Amer. J. Bot.* 57(4): 382-389.

Estimated Budget for First Twelve Months
(March 1, 1971 through February 29, 1972)

		<u>NSF Grant</u>
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 \$14,500)	3 Summer Mos. \$3,222	\$ 5,944
(co-principal investigator, B. L. Turner, 1 summer mo. @ 9 mo. academic rate of \$24,500)	\$2,722	
2. Other Personnel		
d. 2 Graduate Students	9 Academic Mos. 6 Summer Mos.	8,918
1 Res. Sci. Asst. I ($\frac{1}{2}$ time 9 mos., full-time 3 summer mos. @ full-time rate of \$539 per mo.)		
1 Res. Sci. Asst. III ($\frac{1}{2}$ time 9 mos., full-time 3 summer mos. @ full-time rate of \$650/mo)		
f. 1 Pre-Baccalaureate Student (Approx. 650 hrs. @ 2.00/hr for data coding and key punching)	3 Calendar Mos.	1,300
g. 1 Senior Secretary (15% time 12 mo. @ full-time rate of \$450/mo)		810
	Total Salaries and Wages	16,972
B. Fringe Benefits (5.2% of first \$7800 of each salary)		<u>528</u>
	Total of A & B	17,500
C. Permanent Equipment (See Budget Notes)		22,000
D. Expendable Supplies (Computer supplies, magnetic tapes tab cards and paper)		250
E. Travel Domestic (See Budget Notes)		1,910
F. Publication Costs		1,000
G. Other Costs		
1. CDC-6600-Computer Charges (3 hours @ \$200 per hr.)		<u>600</u>
H. <u>Total Direct Costs</u>		43,260

I. Indirect Costs

1. On Campus 43.5% of Salaries and
Wages (\$2,613) ~~\$7,383~~

Amount requested: 20% TDC exclusive of
Computer Charges (\$600) and Permanent
Equipment (\$22,000)

1,132

J. Total Estimated Budget for First
Twelve Months

47,392

The University of Texas at Austin will cost-share in this project in accordance with current Foundation policy.

Estimated Budget for Second Twelve Months
(March 1, 1972 through February 28, 1973)

		NSF Grant
A.	Salaries and Wages	
1.	Senior Personnel	
a.	2 Co-Principal Investigators 3 Summer Mos.	6,166
	(co-principal investigator, R.H. Flake, two summer mos. @ 9 mo. academic rate of \$15,000)	\$3,333
	(co-principal investigator, B.L. Turner, 1 summer mo. @ 9 mo. academic rate of \$25,500)	\$2,833
2.	Other Personnel	
d.	2 Graduate Students 9 Academic Mos.	
	1 Res. Sci. Asst. I ($\frac{1}{2}$ time 9 mos., full-time 3 summer mos. @ full-time rate of \$565/mo.) 6 Summer Mos.	9,338
	1 Res. Sci. Asst. III ($\frac{3}{4}$ time 9 mos., full-time 3 summer mos. @ full-time rate of \$680/mo.)	
f.	1 Pre-Baccalaureate Student 3 Calendar Mos.	1,300
	(approx. 650 hrs. @ \$2.00/hr. for data coding and key punching)	
g.	1 Senior Secretary 832	
	(15% time, total 12 mo. @ full-time rate of \$462/mo.)	
	Total Salaries and Wages	17,636
B.	Fringe Benefits	596
	Total of A & B	18,232
C.	Permanent Equipment (See Budget Notes)	14,000
D.	Expendable Supplies (Computer supplies, magnetic tapes tab cards and paper)	250
E.	Travel - Domestic (See Budget Notes)	1,910
F.	Publication Costs	1,000
G.	Other Costs	
1.	CDC-6600-Computer Charges (3 hrs. @ \$200/hr.)	600
H.	Total Direct Costs	35,992

I. Indirect Costs1. On Campus 43.5% of Salaries and
Wages (\$7,672)Amount requested: 20% TDC exclusive of
Computer Charges (\$600) and Permanent
Equipment (\$14,000)4,278J. Total Estimated Budget for Second
Twelve Months

40,270

Total Estimated Budget for Twenty-Four
Months

87,662

The University of Texas at Austin will cost-share in this project in accordance with current Foundation policy.

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 (5.2% of 1st \$7800 of each salary; 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.

C. Permanent Equipment

First year - 1 gas chromatograph type Varian 2100-40 four column with detectors (\$7,000), 2 Digital Integrator Type Varian 480 (\$12,000), 2 Recorders Type Varian 20 (\$2,000), Capillary Columns (\$1,000).

Second year - 2 Digital Integrators Type Varian 480 (\$12,000), 2 Recorders Type Varian 20 (\$2,000).

E. 1. Domestic Travel

For attendance of two technical meetings (\$500).

Field work, estimated 9,000 miles (\$1,410).

NATIONAL SCIENCE FOUNDATION

PROPOSAL RATING SHEET

Reviewer D.J. Rogers
Univ. of Colo., Boulder

Proposal No.: B018534
Investigator: A. J. Gilmartin
Institution: Smithsonian Inst
Please return to: Systematic Biology
if possible by: DEC 17 1969

Comments (Continue on additional sheet if necessary)

This proposed activities have been done many times already, and I can see no need for further "data diddling" with the "methods" listed. We already know that "the alpha taxonomist's subjective judgments about variability" are quantifiable, but not in the manner given by the investigator.

We also know, or can quickly determine, the effects of missing data, in any study.

It is time now to stop designing new numerical methods for taxonomy, and get on with the business of using those available on real problems in taxonomy.

OVERALL RATING

- ☐ EXCELLENT
☐ VERY GOOD
☐ GOOD
☒ FAIR
☐ POOR

Signature of Reviewer:

D.J. Rogers

Date:

Dec. 26, 1969

Other suggested reviewers (optional):

PROPOSAL TO THE
NATIONAL SCIENCE FOUNDATION

0018534-238E69

Date Submitted SEP 17 1969

1. Name and Address of Institution

Smithsonian Institution, Washington, D. C. 20560

2. Principal Investigator

Amy Jean Gilmartin, Visiting Researcher, Department
of Botany, Smithsonian Institution

3. Business Contact

Joseph H. Shealy, Contracts Office, Smithsonian Institution, telephone code 144-5957

4. Title of Proposal

The Quantification of Population Variability Within the
Angiosperms: A Numerical Phenetics Study

5. Period of Research

Two years

6. Abstract

The proposed two-year research project employs the EDP techniques of numerical phenetics (numerical taxonomy) to test the hypothesis that the alpha taxonomist's subjective judgments about variability are quantifiable and that phenetic studies of populational variability using a few individuals and many characters provide estimates of variability comparable to statistical ones using many individuals and few characters.

The current controversy between pheneticists and those less sympathetic toward numerical taxonomy arise, in part, from their different methodologies. However, if the statistical and the phenetic approaches to estimating variabilities are shown to produce results which are relatively similar then numerical phenetics will be applicable to problems that are not amenable to statistical analyses, in particular populations of many tropical species.

Estimates of two parameters among the members of a population, the mean phenetic distance (m.p.d.) and its variance will be used to define variability of populations. These estimate respectively, the similarity between the various members of the populations and their density or cohesiveness. The application of these procedures to populations is methodologically similar to that devised by the Principal Investigator in a completed study of the comparative variabilities within classically circumscribed taxa in the Bromeliaceae. The proposed study will extend this approach to the population level.

The results of three types of estimates of variability will be compared: statistical, phenetic and subjective. The test material includes four species of the Bromeliaceae, a tropical plant family; a woody, perennial species of the tropical subfamily Caesalpinoideae, and an herbaceous, annual Rosaceae. Comparisons of the quantifications of phenetic differences and variabilities will be made between comparable systems, for example, among populations of the bromeliad species. Lastly, comparisons of these quantifications of populational variability among different families having different dimensions (characters) will be of interest to theoretical taxonomy.

7. Introduction

The proposed research tests and evaluates a method for quantifying phenetic differences and variabilities in populations of several groups of Angiosperms. The variation within and between populations has been the subject for many studies, for example, Cook and Johnson (1968); de Wet (1968); Kannenberg and Allard (1967); Levin (1967); Lindsey and Vickery (1967); Long (1967); Marshall and Jain (1968); Read (1969); Seeligmann and Alston (1967) and numerous others. These studies have focused upon a few selected characters, three or four or rarely up to 16 (Goodman, 1968) or upon one category of characters as in the work of Seeligmann and Alston (1967) on flavonoids.

Estimates of variability using a few characters and a relatively large sample of the populations with scatter diagrams, polygrams or statistical analyses are especially applicable in studies for which numerous individuals of the population are accessible for random sampling (e.g., Ranunculus spp., many Gramineae species, temperate tree species, etc.). Unfortunately information from easily accessible stands of individuals are not always available to the systematist but in a phenetic study often

large numbers of characters can be obtained from each of a small number of individuals.

Crovello (1968a) in a study of *Salix*, section *Sitchensis* has pointed out that numerical taxonomy seems to be peculiarly suited to systematic studies of plants for which information about the breeding system is not readily available and for long lived species which are not amenable to uniform garden studies. That is, large numbers of characters and a numerical phenetics study should prove especially useful in systematic studies of perennial taxa which have their individual members scattered geographically.

Tropical trees and shrubs especially, often do not lend themselves to random sampling experiments. Because stands of individuals of many tropical species are not available and because there may be a kilometer or more in distance between individuals of the populations, random sampling of these types of populations for a statistical study often is impossible.

I hypothesize that numerical phenetics (numerical taxonomy) provides a method for sampling intra- and inter-populational variability which is particularly well suited to many tropical taxa for which statistical estimates of population variability are not possible. It is suggested that the method described in the following section is especially appropriate for alpha taxonomic studies of tropical perennial species.

Alpha taxonomic studies generally utilize a large number of characters when judgments must be made about circumscriptions and variability and often the taxonomist must base his decisions on a very small number of specimens. From these, he endeavors to learn the pattern of characters and their variability which he communicates in a subjective manner. Davis and Heywood (1963, p. 143) write concerning classical taxonomy and characters, "... even if only some of them are explicitly described, they may all be assessed at a glance ..." For a general purpose as opposed to a special classification, the maximum number of characters provides the superior type of classification. The classical taxonomist implicitly uses the maximum-character type of classification. It is hypothesized that the use of a large number of characters also often provides the best quantified estimate of the pattern of variability of populations.

Quantified estimates of the pattern of variability (beyond range estimates for some of the characters) have not always been included in alpha taxonomic studies. However, I assume that whenever it is possible quantified estimates of intra-populational variability should be made as these are more accurately communicated than subjective ones. In addition, it seems likely that biosystematic studies of the future will draw more of their data than at present from prepared floras, especially from computerized ones. Therefore, it seems advantageous to perfect a method for quantifying the pattern of variability when the usual scatter diagrams and statistical methods are not appropriate.

The method to be used here involves scoring a few individuals for a large number of characters. From data such as collected in my bromeliad studies (Table 1) and with the Computer Program (Table 2), the mean phenetic distance among the individuals is computed. The procedure is described in the next section for estimating two parameters of a biological population, the mean phenetic distance, hereafter called the m.p.d., and its variance. Gilmartin (1967; 1968b; 1969a, b) devised this approach in order to estimate intraspecific, intrageneric and intrafamilial variability during the course of an alpha numerical taxonomic study of the Bromeliaceae of Ecuador. I computed (Gilmartin 1968a, 1970) the m.p.d. and its variance to quantitatively define various classically derived taxonomic circumscriptions as part of the alpha taxonomic study.

The m.p.d. and its variance estimate respectively, the similarity between a taxon's members and its density or cohesiveness. The proposed study will provide estimates of these parameters for defining populations. Samples of two or more populations of single species will also be submitted to the computer program in order to obtain an estimate of the minimum m.p.d. and its variance within species. If it proves desirable the data will also be submitted to standard programs for printing out phenograms and cluster analyses, as in Crovello (1968a, b; 1969).

In order to evaluate the m.p.d. and its variance as parameters for assessing variability and homogeneity within groups such as populations, a comparative study will be made. Standard procedures for sampling variability within populations with five characters and 30 or more individuals will be carried out. The same populations will be subjected also to a numerical

phenetic analysis to compute the m.p.d. and its variance with 60 or more characters and 5, 10, and 15 individuals. It is hypothesized that the two types of estimates, a) statistical and b) numerical phenetical, will show comparable relative results.

At the same time, the classical method of the alpha taxonomist will be employed, subjective appraisal of variability. These subjective judgments will be made by a specialist in each particular taxonomic group. Variability will be subjectively rated on a scale of 1-5, 1 being considered to indicate "uniformity" and 5, "extreme variability." The subjective type of appraisal of the pattern of variability will provide a comparison and will relate the quantified results to the method which has been used in all alpha taxonomic studies. The subjective method is extremely useful but one for which it is particularly difficult to accurately communicate results.

Although I have previously employed the parameters, m.p.d. and its variance, in order to quantify the taxonomic constructs of alpha taxonomists (Gilmartin, 1969a, b); the method seems at least as appropriate for obtaining biosystematically significant quantifications of the variability within populations. In fact, Crovello (1968b) has utilized numerical phenetics for estimating the pattern of variation of 'site collections' (in part comparable with populations), although the m.p.d. and its variance were not employed precisely as proposed here.

The family Bromeliaceae is especially fitted to be the experimental material for this study because 1) it is a tropical family and individuals live at least several years, a trait typical of tropical shrubs and trees; 2) frequently entire individuals can be collected and herbarium specimens made of these affording a large array of characters for the phenetic analyses; 3) there is a wealth of alpha taxonomic information concerning this family in the monograph which L. B. Smith is currently preparing; 4) the Principal Investigator has previous experience with members of the family both in herbaria and in the field.

The problems of missing data and the effects of changes in the character array both quantitative and qualitative merit attention and the proposed research will also help to elucidate

these problems beyond the present state, for example, Crovello (1968a, b, 1969) and Gilmartin (1967, 1968b, and in preparation).

8. Materials and Methods

Materials

Members of the family Bromeliaceae will make up the bulk of the experimental material. For comparative purposes a woody perennial species (Leguminosae) and an herbaceous annual (Rosaceae) will also be subjected to population sampling.

The Bromeliaceae is a tropical, Angiospermous, Monocotyledenous family with which the Principal Investigator is familiar (Gilmartin, 1966, 1968a, 1970). Although previous work has been initially undertaken as a purely alpha taxonomic endeavor, subsequently it became evident that the members of this family offered valuable opportunities for elucidating the pattern of variability in general in tropical species. As a result intra group variability was quantified for several taxa in the family (Gilmartin 1969a, b) as an alternative to subjective judgments concerning variability.

Populations of two bromeliad species, Guzmania zahnii Mez and Vriesea pitteri Mez from 1200 meters altitude near San Vito in Southern Costa Rica will be sampled. These both represent mesophytic, tropical, perennial epiphytes.¹ Tillandsia prodigiosa Benth. (Bromeliaceae) will be sampled from a population between mile 48 and 51 Villa Union-Durango, Sinoloa, Mexico at 1700 meters altitude.¹ This population is in a semi-xeric, open, pine-oak forest. A population of Hechtia sp. (Bromeliaceae) from the southern portion of the Cape District, Baja California, between 300 meters and 900 meters altitude in strongly xeric conditions will also be sampled (Figure 1).

With some Bromeliaceae as with many tropical species, random sampling techniques such as taking every tenth plant are

1

The population samples of G. zahnii and V. pitteri were carried out with the help of 7 students from the OTS course during July 1968. Tillandsia prodigiosa in the state of Sinoloa, Mexico has been studied by the P.I. for three consecutive years (1966-69) and a population sample of 65 specimens has been completed. Herbarium specimens and photographs are in the private herbarium of the P.I.

not always feasible. Therefore, the population samples represent selection of all or most of the fertile specimens near the center of the populations.

The Bromeliaceae is a tropical largely new world plant family, which grows in both xeric and mesic habitats. The same characters (dimensions) will be applicable to the populations of each of the four species regardless of the habitat type. However, since missing data (lack of information concerning some characters) does occur, the effects will be evaluated in two ways: the method described and used by Crovello (1969) and that tested by Gilmartin (1967; 1969b and in preparation).

The woody species, Cassia (Brandege 1892a, b) was chosen because it is a member of a large cosmopolitan family and because it was available for random sampling in a tropical to subtropical area, the Cape District, Baja California. It grows in both xeric and mesic conditions, being much more robust in the latter. A population from near 900 meters altitude, Cape District, Baja California will be sampled and two other populations from 200 meters, one xeric and one mesic will also be sampled (Figure 1). The m.p.d. within each of the three populations of this species should afford interesting comparisons in quantified terms of the variability within xeric versus mesic populations of the same species and the possible effects of altitude upon variability of a woody perennial species in a tropical type habitat.

An annual, herbaceous Rosaceae, Potentilla sp. (Brandege, 1892a, b) was chosen to provide comparison with a temperate, short lived taxon and because it was abundantly available in the same general tropical area as the above, the Cape District, but growing in a temperate climate at 1800 meters altitude.

In addition to the population samples described above, general collections will be made in the Cape District, Baja California and the herbaria of the California Academy of Science, the University of California, Berkeley and the U.S. National Museum will be consulted. These herbaria are especially well represented with specimens of the taxa to be studied from the study areas.

Populations of Cassia sp. and Potentilla sp. will be sampled during the next two years, plants photographed and the herbarium specimens prepared. Although truly random sampling of most bromel populations is scarcely possible because of the limited numbers of mature plants available during any one flowering season, random sampling of populations of the other taxa is feasible and will be realized. Samples of populations of the respective species will be drawn by either taking every tenth plant along a transect or by taking all plants from lots near the center of the populations.

Methods

The unit of comparison, OTU (operational taxonomic unit) for the numerical phenetic analyses will be individual specimens. These will be collected from each of the above mentioned areas and will be scored with the appropriate character-array. In general, intra-plant variability will be handled by consistency in scoring the same relative positions of organs or parts. However, if variability with the individual plant is high then at least five measurements or determinations will be made for the characters in question and the means of these will be used for scoring the character-states. No conscious weighting of characters is planned although weighting inherent to the selection of the numbers of character-states will necessarily be present.

Format cards have been prepared for recording the information concerning the bromeliad species (Table 1). One hundred and four characters are included and these will be used or slightly modified. Similar types of data but appropriate to the other two major taxonomic groups will be gathered from the individuals of the samples from these other taxa.

Cognizance is taken of the problem of dioecism in Hechtia sp. and the linkage of some characters to one or the other of the sexes. Epidermal characters will be used particularly with the bromel species with which they were demonstrated to form a useful category of characters (Gilmartin, 1968b). Scoring for the characters of all groups will be done on standard scoring sheets from which the IBM cards will be punched.

For comparative purposes both standardized and non-standardized data will be employed. Previous work by Gilmartin (1969b) indicated little relative difference between the phenetic distance values computed with standardized and non-standardized data but the other families may show less consistency in this regard and this will be tested.

Rogers and Fleming's (1964) negative logarithmic transformation of the matching coefficient ($D = \log_2 S_{rt}$, where S_{rt} = the sum of the matches in the character-states divided by the sum of the number of comparisons) will estimate phenetic distance between OTUs (operational taxonomic units). Table 2 reproduces the Computer Program in Fortran IV for the IBM 7040 (easily modifiable to the IBM 360) which computes the phenetic distance between each member of an array of OTUs. This program has been run repeatedly, successfully.

The characters will be scored by the Principal Investigator and the IBM data cards punched by computer center personnel. Samples for intra-population phenetic distances will be generated by scoring OTUs randomly selected from given populations and submitting these data to the program for phenetic distance to obtain the basic distance matrix. The m.p.d. and its variance will be computed from the respective distance matrices. These two parameters, the m.p.d. and its variance estimate respectively the similarity between individuals of a population and the density or cohesiveness of the members.

9. Summary

The hypothesis will be tested that the alpha taxonomist's subjective judgments about variability are quantifiable and that phenetic estimates of variability using a few individuals and many characters with EDP provide estimates of variability comparable to statistical ones using many individuals and few characters. The methods which have been devised by the Principal Investigator for estimating the average similarity and its variance within taxonomic groups will be employed with populations. In addition, the effects of missing data, both in kind and number will be tested.

The results of my previous studies in numerical phenetics and taxonomic circumscriptions in the family Bromeliaceae and the proposed population analyses of species in this family and two others will offer valuable comparisons. If the proposed method proves satisfactory it will be valuable for studies of variability in many tropical woody species.

Variation within taxa and populations is one of the prime concerns for all systematic work including alpha taxonomy. As pointed out by Eplin and Catlin (1950), taxonomic methods which abstract characters while overlooking particular differences shed little light on the process of evolution. They can only describe the results. Description of the results of evolution is the domain of alpha taxonomy but systematics also aims to interpret and explain. It is precisely the particular differences in the pattern of variability of populations which are summarized by the m.p.d. and its variance. In the proposed study the degrees of variation within naturally occurring populations of tropical plants, temperate plants, mesophytes and xerophytes will be quantified using 60 or more characters. In this way nebulous concepts such as "great intraspecific variability of tropical over temperate species" may be quantified and thus tested.

Comparisons between m.p.d. values and variance will be made from comparable systems. For example, comparisons will be made between xeric and mesic tropical bromels and between mesic populations of the woody, perennial Cassia sp. and xeric ones of the same species. The herbaceous Potentilla species will afford a comparison with a temperate, annual and will provide a comparison between two quantitative estimates of variability: statistical and phenetic.

Comparisons probably are not warranted between populations not confamilial although they may nevertheless be of interest. Corner (1963) for example, has suggested that the concept of species as used by tropical botanists may be somewhat different from the temperate species concept. Michener (1969) has pointed out the inadequacy of present subjective methods for deciding on the relative differences between groups which do not share the same characters. Perhaps we may eventually accept quantified comparisons of phenetic difference and variability which utilize different dimensions or characters.

Since the monographer of the family, Lyman B. Smith, is currently revising and perfecting the Monograph of the Bromeliaceae, the information to be gathered concerning populations within several bromeliad species and the estimates of variability will be particularly helpful to Smith at this time. In addition, the general plant collections in the Cape District, Baja California which the Principal Investigator will make as part of the proposed study will be useful in the preparation of a modern flora of the Cape District (estimated 800 species) now in the check-list stage and under the direction of John Thomas.

Finally, the Principal Investigator has been appointed to the Flora North American Project as a Research Associate at the Smithsonian Institution 1969-70 as part of a team effort to produce a small scale, model taxonomic treatment for the FNA Project and will be charged with deciding what types of taxonomic data should be included in a computerized flora and with helping to format the data. This work will be highly complementary to the proposed study and will offer many opportunities for testing taxonomic theories and floristic applications.

10. Budget

	<u>Year 1</u>	<u>Year 2</u>
Personnel Compensation:		
Principal Investigator (50% time)	\$6,750	\$7,586
Personnel Benefits	675	759
Travel and Transportation of Persons:		
Round trip economy class air fare, Washington, D. C. - Phoenix Arizona and La Paz, Baja California	\$355	
Per diem 5 days @ \$16	80	
Living expenses (field work - 20 days)	200	
Miscellaneous local travel	<u>200</u>	
Total Travel & Transportation of Persons	835	900
Equipment Rental:		
Data processing (IBM 360 - 2 hrs.)	\$600	
Pack animals (field work -20 days)	<u>80</u>	
Total Equipment Rental	680	700

	<u>Year 1</u>	<u>Year 2</u>
Other Services:		
Technical Assistant Guide (for field work 20 days)	\$200	
Programming consultant 10 hours	100	
Keypunching	210	
Typing services	<u>100</u>	
Total Other Services	\$ 610	\$ 700
Supplies and Materials:		
E.D.P. materials, data sheets, punch cards, tapes for storing data	\$200	
Stationery	70	
Chemical reagents and vials	80	
Plant collecting equipment, press, driers, blotters	40	
Laboratory supplies	<u>150</u>	
Total Supplies and Materials	540	600
Equipment:		
Dissecting microscope	<u>350</u>	<u> </u>
Total Direct Cost	\$10,440	\$11,245
Indirect Expenses - 27%	<u>2,819</u>	<u>3,036</u>
Total Cost	<u>\$13,259</u>	<u>\$14,281</u>
Total Estimated Cost - two years:	\$27,540	

	(genus)	(species)	No. _____	
	(determined by)	(det. date)	(coll. date)	
HABIT	_____ x _____ cm.			
SUBSTRATUM	_____			
SCAPE	_____ x _____ cm.	(color)	(erect, curved, etc.)	
SCAPE BRACTS	_____ x _____ cm.	(color)	(erect, reflexed, etc.)	
LEAVES	_____ x _____ cm.	(color)	(flat, involute, etc.)	
	(surface)	(erect, spreading, reflexed, etc.)		
SHEATH	_____ x _____ cm.	(color)	(surface)	
	(shape)	(conspicuous, indistinct, etc.)		
RACHIS	(cross section)	(color)	(exposed - covered)	

	(genus)	(species)	No. _____				
		length	width	altitude	color	shape	surface
INFLORESCENCE							
primary branches							
secondary branches							
tertiary branches							
primary bracts							
secondary bracts							
floral bracts							
sepals free							
connate							
petals free							
joined							
flowers, c. petals							
sin petals							
ovary							
stamens	(length)	exserted	includ.	connate to petals	(yes, no, partly)		
FRUIT		seeds	distance between flowers				
LOCALITY	_____						
COMMENTS	_____						

Table 2. Print-out of Program for Computing Phenetic Distance (D_{rf})GILKINSON TAXONOMY, DISTA
ISN SOURCE STATEMENT

FORTRAN SOURCE LIST

```

0  SIBETC
1  DIMENSION IA(100,110),NAT(100), D(100,100),T(100),IK(100)
   C MM(110), NN(110),IEXCAR(110), A (13, 13)
2  READ (5,201) IEXC
4  201 FORMAT (13)
5  IF (IEXC) 21, 21, 22
6  22 READ (5,200) (IEXCAR (I),I=1,IEXC)
13  200 FORMAT (26I3)
14  21 READ(5,1111)L1,L2
17  1111 FORMAT(2I3)
20  DO 1112 I=1,L1
21  IF (L2-25) 10, 10, 12
22  10 READ(5,100)NAT(I),(IA(I,ICAR),ICAR=1,L2,
27  100 FORMAT(14,1X,25I3)
30  GO TO 1112
31  12 READ(5,100) NAT (I),(IA(I,J),J=1,25)
32  READ (5,110) (IA(I,J),J=26,L2)
43  110 FORMAT (5X,25I3)
44  1112 CONTINUE

C
C  DISTANCE FUNCTION FOR EACH POSSIBLE DIFFERENT OTU PAIR
C
   SUM = 0
   SUMSQ = 0
4  DO 54 KATT=1,L2
   MM(KATT)=0
54  NN(KATT)=0
   LL=L1-1
500  FORMAT (1H1)
50  301 FORMAT (41H OTU PAIRS WITH LESS THAN 100 COMPARISONS//
C34H OTU1 OTU2 NUMBER OF COMPARISONS)
57  ZN = 0
60  LINES=0
61  DO 4I=1,LL
62  K=I+1
63  DO 4J=K,L1
64  M=1
65  N=1
6  ZN = ZN +1.
7  DO 3 KATT=1,L2
C
C  CALL FOR THE SUBROUTINE TO EXCLUDE DESIGNATED CHARACTERS
C
70  CALL XCLUDE (KATT,IEXC, IEXCAR,INDIC)
71  IF (INDIC) 23, 23, 3
72  23 IF (IA(I,KATT).EQ.0.OR. IA(J,KATT).EQ.0) GO TO 3
75  IF (IA(I,KATT) - IA(J,KATT)) 1,2,1
76  2 X=M+1
77  N = N+1
100  MM(KATT) =MM(KATT)+ 1
101  GO TO 3
102  1 N = N+2
103  NN(KATT)=NN(KATT) +1
104  3 CONTINUE
106  XM=M

```

Table 2 (continued)

GILMARTIN TAXONOMY, DISTA	FORTRAN SOURCE LIST
ISN	SOURCE STATEMENT
107	XN=N
110	D(I,J) = -(ALOG(XM/XN)/ALOG(2.0))
111	SUM = SUM + D(I,J)
112	SUMSQ = SUMSQ + D(I,J)**2
C	
C	NUMBER OF COMPARISONS FOR OTU PAIRS, PRINTOUT WHEN
C	LESS THAN 100
113	NUMCOM=(N+M)/2 - 1
114	IF (NUMCOM.GT. 99) GO TO 4
117	LINES=LINES+1
120	IF(MOD(LINES-1, 50)) 703,704,703
121	704 PRINT 500
122	PRINT 301
123	703 PRINT 300, NAT(I), NAT(J), NUMCOM
124	300 FORMAT (15, 1X, 15, 1 10)
125	4 D(J,I) = D(I,J)
C	
C	ESTIMATION OF PARAMETERS OF DISTANCE FUNCTION & ARRAY
C	
130	XBAR = SUM/ZN
131	SS = SUMSQ - SUM**2/ZN
132	STDEV = SQRT (SS/(ZN-1.))
133	NJ=L1+5
134	NJ=1
135	DO 52 JJ=5,NJ,5
136	IF (JJ.GT.L1) JJ=L1
141	LINES=0
142	DO 51 I=1,L1
143	LINES=LINES+1
144	IF(MOD(LINES-1, 25)) 51,702,51
145	702 WRITE (6,101) (NAT(J),J=MJ,JJ)
152	101 FORMAT (1H1,5X,17HDISTANCE FUNCTION//5115//)
153	51 WRITE (6,102) NAT(I), (D(I,J), J=MJ,JJ)
161	102 FORMAT(3X,14, F12.8, 4F15.8, /)
162	IF (JJ-L1) 52,53,52
163	52 MJ=MJ+5
C	
C	R ROUTINE
C	
165	53 DO 6 I=1,L1
166	DO 6 J=1,L1
167	IF (I.EQ.J) GO TO 6
172	DO 70 K=1,L2
173	CALL XCLUDE (KATT,IE,C, IEXCAR,INDIC)
174	IF (INDIC) 24, 24, 70
175	24 IF (IA(I,K).NE.IA(J,K)) GO TO 70
200	IR(I)=IR(I)+1
201	GO TO 6
202	70 CONTINUE
204	6 CONTINUE
C	
C	H-T-ROUTINE
C	
207	DO 80 I=1,L1
210	SIH=0.0



Figure i. SOUTHERN CAPE DISTRICT, BAJA CALIFORNIA.

Literature Cited

- Anon. 1968. Flora and Fauna del Territorio de B.C. Sur. Agencia Gral. de Agric. La Paz.
- Brandegee, T.S. 1892a. Flora of the Cape region of Baja California. Proc. Calif. Acad. Sci. series 2, III:108-182.
- , 1892b. Flora of the Cape region. Zoe III:223-408.
- Cook, S.A. and M. Johnson. 1968. Adaptation to heterogeneous environments. I. Variation in heterophylly in Ranunculus flammula L. Evol. 22(3):496-516.
- Corner, E.J.H. 1963. The Evolution of Tropical Forest, pp. 46-59. In J. Huxley, A.C. Hardy, and E.B. Ford, eds., Evolution as a Process. Collier Books, New York.
- Crovello, T.J. 1968a. Quantitative taxonomic studies in the genus Salix Section Sitchensis. Publications in Botany, Univ. California. Vol. 44. Univ. Calif. Press, Berkeley.
- , 1968b. The effect of missing data and of two sources of character values on a phenetic study of the willows of California. Madroño 19(8):301-315.
- , 1969. Effects of change of characters and of number of characters in numerical taxonomy. Amer. Midland Naturalist 81(1):68-86.
- Davis, P.H. and V.H. Heywood. 1963. Principles of Angiosperm Taxonomy. Oliver and Boyd, Edinburgh and London.
- De Wet, J.M.J. 1968. Biosystematics of the Bothriochloa barbinodis complex (Gramineae). Amer. J. Bot. 55(10):1246-1250.
- Epling, Carl and W. Catlin. 1950. The relation of taxonomic method to an explanation of organic evolution. Heredity 4:313-325.
- Gilmartin, A.J. 1966. Las Bromeliaceas de Honduras. Ceiba (Honduras) 11(2):1-81.
- , 1967. Variance of phenetic affinities with different randomly selected character-sets in an alpha-numerical taxonomic study of the Bromeliaceae. Amer. J. Bot. 54(5):646 (abstr.).
- , 1968a. Taxonomic notes on some Bromeliaceae of Ecuador. Phytologia 12(2):153-167.
- , 1968b. The Bromeliaceae of Ecuador, an alpha-numerical study. Ph.D. thesis, Univ. Hawaii. Univ. Microfilms Ann Arbor. Dissert. Abstr. 30(1).
- , 1969a. The quantification of some plant-taxa circumscriptions. Amer. J. Bot. 56(6):in press.

- , 1969b. Numerical phenetic samples of taxonomic circumscription in the Bromeliaceae. *Taxon* 18:378-392.
- , 1970. Bromeliaceae of Ecuador. Monog. Phaner. J. Cramer Lehre, Germany. In press.
- Goodman, M.M. 1968. A measure of 'overall variability' in populations. *Biometrics* 24(1):189-192.
- Kannenberg, L.W. and R. Allard. 1967. Population studies in predominantly self-pollinated species. VIII. Genetic variability in the Festuca microstachys complex. *Evol.* 21(2):227-240.
- Levin, Donald A. 1967. Variation in Phlox divaricata. *Evol.* 21(1):92-108.
- Lindsay, Delbert W. and R.K. Vickery, Jr. 1967. Comparative evolution in Mimulus guttatus of the Bonneville Basin. *Evol.* 21(3):439-456.
- Long, R.W. 1966. Biosystematics of the Helianthus muttallii Complex (Compositae). *Brittonia* 18(1): 64-79.
- Marshall, D.R. and S.K. Jain. 1968. Phenotypic plasticity of Avena fatua and A. barbata. *Amer. Naturalist* 102(927):457-467.
- Michener, C.D. 1969. Principles of systematic zoology. *Syst. Zoo.* 18(2):232-234.
- Read, R.W. 1969. Variation and taxonomy of Pitcairnia (Bromeliaceae) in Jamaica. *Brittonia* 21(1):83-90.
- Rogers, D.J. and H. Fleming. 1964. A computer program for classifying plants. II. A numerical handling of non-numerical data. *BioSci.* 14(9):15-28.
- Seeligmann, P. and R.E. Alston. 1967. Complex chemical variation and the taxonomy of Hymenoxys scaposa. *Brittonia* 19(3):205-211.

CURRICULUM VITAE

Amy Joan Gilmartin

Biographical summary:

- | | |
|---------|---|
| 1932 | Born at Red Bluff, California. |
| 1941-50 | Early schooling in Hawaii, graduated from Hilo High School. |
| 1950-54 | Pomona College, California, B. A. in botany. |
| 1954 | Married |
| 1954-56 | University of Hawaii, M.Sc. in botany. |
| 1956-58 | University of British Columbia, additional graduate course work, mathematics. |
| 1957-61 | Gave birth to three children. |
| 1961-64 | Residence in Latin America. |
| 1965-68 | University of Hawaii, Ph.D. in botany (June, 1968). |

Teaching experience:

- | | |
|---------|---|
| 1954 | Teaching Assistant; general botany, during senior year as undergraduate. |
| 1956-59 | Teaching Assistant; general botany, general biology; University of British Columbia. |
| 1957-59 | Instructor; plant morphology; University of British Columbia. |
| 1962-64 | Profesora Contratada; upper division courses in plant taxonomy, comparative morphology; Universidad de Guayaquil, Ecuador; taught in Spanish. |
| 1965-66 | Teaching Assistant; general botany; University of Hawaii. |

Research interests and experience:

- | | |
|---------|--|
| 1954-56 | Fertilization, compatibility and seed formation in <u>Passiflora edulis</u> ; Research Assistant (1/2); Department of Physiology, Hawaii Agriculture Experiment Station. |
|---------|--|

- 1957 Variation in *Viola* and *Saxifraga*; Research Assistant (summer); Department of Botany and Biology, University of British Columbia.
- 1957-58 Flora of the Pacific North West; Assistant Herbarium Curator (1/2); Herbarium, University of British Columbia.
- 1959-60 Inheritance of seedling characters in the Douglas Fir; Research Assistant (1/2); Department of Forestry, University of British Columbia.
- 1961-64 Descriptions of xerophytic communities in coastal Ecuador, and the effects of man; Seasonal changes in a semi-deciduous community in Ecuador; Recipient of two Grants-in-Aid of Research from the Society of the Sigma Xi.
- 1964 Flora of the Galapagos; Member, Calif. Acad. Science, International Expedition to the Galapagos.
- 1964-65 Bromeliaceae of Honduras; Visiting Researcher (summer, 1964), Escuela Agricola Pan-Americana in Honduras; Visiting Researcher, San Diego Natural History Museum.
- 1965-68 Bromeliaceae of Ecuador, an alpha and numerical taxonomic study; PhD thesis research; 1965(summer) McInerney Foundation Grant for field work in Ecuador and upper Amazon basin; 1966 (summer) Training Grant*, U. S. National Herbarium, Washington, D.C., for work on their Bromeliaceae collections; 1967-68 Predoctoral AAUW (American Association of University Women) Fellow, PhD research.
- 1967-68 Flowering time and differentiation in a tropical species; 1967(summer) field work, southern Sierra Madre Occidental of Mexico; further field work 1968(May); living bromeliads maintained and studied in Monterey, California.

*Awarded a 1 year Predoctoral Research Fellowship which I was unable to accept because of familial obligations; a 2 1/2 month summer training grant was thereby arranged as an alternative.

BIBLIOGRAPHY

Amy Jean Gilmartin

Minor Publications:

- GILMARTIN, A. J. 1964. Angiospermas colectadas en el Cerro Azul. Inst. Cien. Nat. Symp. Vol. (Ecuador).
- GILMARTIN, A. J. 1964. Ecuador - Bromeliad Country. Bromel. Soc. Bull., 14(1):4-7.
- GILMARTIN, A. J. 1964. La expedición internacional a las Galápagos. Ciencia y Naturaleza (Ecuador), 7(1):3-4.
- GILMARTIN, A. J. 1965. List of botany texts and general references written in Spanish. Bull. Assoc. Trop. Biol., 4:27-29.
- GILMARTIN, A. J. 1966. Bromels in Hawaii. Hawaiian Bot. Soc. Newsletter., 5(3):21-24.
- GILMARTIN, A. J. 1967. A plea to Bromel collectors. Bromel. Soc. Bull., 17(5):101-102.
- GILMARTIN, A. J. 1967. The international conference on systematic biology. Hawaiian Bot. Soc. Newsletter., 7(3):20-25.
- GILMARTIN, A. J. 1969. The sex life of bromeliads. Bromeliana. / accepted (New York)

Scientific Papers and Abstracts:

- GILMARTIN, A. J. 1958. Post-fertilization seed and ovary development in Passiflora edulis Sims. Jour. Imp. Col. Trop. Agric., 35(1):41-58.
- GILMARTIN, A. J. 1964. Algunas especies leñosas xerofilas en el litoral del Ecuador, especialmente en La Península de Santa Elena. Adial (Ecuador), 1:14-21.
- GILMARTIN, A. J. 1965. Las Bromeliaceas de Honduras. Ceiba (Honduras), 11(2):1-81.
- GILMARTIN, A. J. 1967. Numerical taxonomy - an eclectic viewpoint. Taxon, 19(2):8-12.
- GILMARTIN, A. J. 1967. Variance of phenetic affinities with different randomly selected character sets in an alpha-numerical taxonomic study of the Bromeliaceae. Amer. J. Bot., 54(5):646. abstract.

- GILMARTIN, A. J. 1968. Taxonomic notes on some Bromeliaceae of Ecuador. *Phytologia*. 16(2):153-167.
- GILMARTIN, A. J. 1968. Baker's Law and dioecism in the Hawaiian flora. *Pac. Sci.*, 22(3):285-292.
- GILMARTIN, A. J. 1968. The Bromeliaceae of Ecuador, an alpha-numerical study. PhD. Thesis, Univ. Hawaii Univ. Microfilms, Ann Arbor. Dissert. Abstr. 30(1)
- GILMARTIN, A. J. 1969. Numerical phenetic samples of taxonomic circumscriptions in the Bromeliaceae. *Taxon*. 18:378-392.
- GILMARTIN, A. J. 1969. The quantification of some plant taxa circumscriptions. *Amer. J. Bot.* 56 (6) part 1.
- GILMARTIN, A. J. 1970. Family Bromeliaceae. in Wiggins, Ira (ed.) *Flora of the Galapagos*. accepted
- GILMARTIN, A. J. Quantitative and qualitative changes in character-sets and estimates of phenetic distance. draft typescript in preparation.
- GILMARTIN, A. J. Variation of population samples of two species in the moist forest of San Vito, Costa Rica. in preparation
- Books:
- GILMARTIN, A. J. 1970. The Bromeliaceae of Ecuador. Monog. Phaner. J. Cramer, Lehre, Germany. in press

NATIONAL SCIENCE FOUNDATION
PROPOSAL RATING SHEET

Reviewer: Professor David Rogers
Department of Biology
University of Colorado
Boulder, Colorado 80302

Proposal No.: JO 00791
Investigator: Belzer
Institution: U of Pittsburgh
Please return to: ER&T
if possible by: 9/18/69

Comments (Continue on additional sheet if necessary)

1. From page 2: "This concept has been partially implemented for the NASA Regional Dissemination Center" If NASA has been interested, let NASA continue to support it.

2. The "scramble" technique is not well-documented as to methods of development, but I think that it essentially duplicates other such techniques in other applications. Check, for example, with members of OSIS in NSF for several software packages they have supported.

3. The procedures do not seem to be user-oriented, with respect to input/output.

4. There is no demonstrated need for this particular application.

5. Purely heuristic methods seem to be the basis for the design and testing, at a stage when such methods have already shown their lack of productivity.

OVERALL RATING

- ☐ EXCELLENT
☐ VERY GOOD
☐ GOOD
☐ FAIR
☒ POOR

Signature of Reviewer:

David G. Rogers Date: *Sept. 9, 1969*

Other suggested reviewers (optional):

6 August 1969

Mrs. Rosemary Smolker
Quarterly Review of Biology
Department of Biological Sciences
State University of New York
Stony Brook, New York 11790

Dear Mrs. Smolker:

I have received Rektory's Survey of Applicable Mathematics. I would be pleased to further receive a reprint of your book review section, in order that I might familiarize myself with it.

Thank you,

George F. Estabrook

GFE:gm

BRITTONIA

Publication of

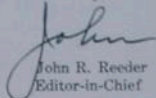
THE AMERICAN SOCIETY OF PLANT TAXONOMISTS

Dear Dave:

Thank you for your careful review of the paper by: Grady Webster
"A revision of Phyllanthus . . . "

Your comments will be extremely helpful in deciding on the final disposition of this manuscript.

Sincerely,



John R. Reeder
Editor-in-Chief

Date August 12, 1969

Department of Botany, University of Wyoming, Laramie, Wyo. 82070

NATIONAL SCIENCE FOUNDATION
PROPOSAL RATING SHEET

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

Proposal No.: B017214
Investigator: R. H. Flake
Institution: Univ. of Texas
Please return to: Systematic Biology
if possible by: August 25, 1969

Comments (Continue on additional sheet if necessary)

This seems to be a sound proposal, and one within the capabilities of the coprincipal investigators. I have no hesitancy in recommending that this proposal be approved.

OVERALL RATING

- ☒ EXCELLENT
☐ VERY GOOD
☐ GOOD
☐ FAIR
☐ POOR

Signature of Reviewer:

Date 30 July 1969

Other suggested reviewers (optional):

8017214-23JL69

Proposal for Support of Research Entitled

COMPUTER-AIDED SYSTEMATIC STUDIES

Submitted to

National Science Foundation

By

The University of Texas at Austin

Austin, Texas 78712

July 1, 1969

Proposal to the
National Science Foundation
from the University of Texas at Austin
for Support of Research Entitled:
Computer-Aided Systematic Studies

Abstract

New computer-aided methods have been developed to aid in biological classification problems, particularly at the infraspecific level or lower. The main goals of the proposed research are to apply these new techniques to a number of selected systematic problems using both chemical and morphological data, to develop new numerical methods of populational data characterization, and to develop statistical techniques for estimating the predictive quality of classifications produced by this new approach.

Previous Investigations

Numerous techniques have been developed to utilize the data-processing capabilities of the modern high-speed computer to aid in the classification problems associated with the areas of biological taxonomy and information retrieval (Sneath, 1957; Rogers & Tanimoto, 1960; Needham, 1961; Parker-Rhodes, 1961; Sokal & Sneath, 1963; Wirth, Estabrook & Rogers, 1966; Estabrook, 1966; Rubin, 1967; Lance & Williams, 1967; Flake & Turner, 1968; Crovello, 1968; Jardine & Sibson, 1968; Flake, von Rudloff & Turner, 1969).

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 where, as a research associate in the computation center, he was involved in developing computer programs to assist a graduate student in Botany. The student was doing a systematic study of genera of the subtribe Oncidiinae of the Orchidaceae for his doctoral research. The clustering methods proposed by Rogers and Tanimoto (Rogers & Tanimoto, 1960) were employed in that study. The graduate student subsequently 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 in data processing seemed essential.

During the past three years the co-principal investigators, in collaboration, have developed new computer-aided methods for studying infra-specific populations that the investigators believe are highly significant contributions to the field. (Flake & Turner, 1968; Flake, von Rudloff & Turner, 1968). A number of computer programs embodying these techniques have been developed for processing populational data.

This new approach is primarily designed to assist the biological worker concerned with systematic problems at the infraspecific level, 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 the application of this approach to a study of the clinal variation in Juniperus virginiana populations is contained in the preprint in the appendix (Flake, von Rudloff & Turner, 1969).

Proposed Research

The main goals of the proposed research are to: i) apply these new techniques to a number of systematic problems; ii) develop new methods of characterizing populational data, both morphological and chemical, at the infraspecific and specific levels, and iii) develop statistical methods for estimating the predictive quality of the ensuing hierarchical structures produced by the numerical classification computer algorithms.

Systematic Applications

Juniperus - The species of this genus are unusually adapted for combined chemosystematic and numerical study. They are widespread evergreen trees, easy to locate during the winter, and they possess a wide array of terpenoids that are relatively easy to survey for, especially using gas liquid chromatography (von Rudloff, 1968; Vinutha & von Rudloff, 1969).

Because of this, several of the taxa have been selected for intensive chemosystematic study, at the infraspecific level. For example, von Rudloff, Irving and Turner (1967; 1969) have shown that the allopatric species I. virginiana and I. ashei, previously thought to be highly introgressed as a result of hybridization were, in fact, not forming hybrid swarms nor could introgression in either of the species be detected using chemical data. The morphological evidence for introgression was also found, upon re-examination, to be dubious.

More recently, I. virginiana has been subject to intensive populational study and it appears that there is a chemical cline in this species from the vicinity of Washington, D. C. to eastern Texas (Flake, von Rudloff and Turner, 1969; cf. appendix). This study was based upon samples taken along a linear transect. We envision extending the study to include regional samples in the form of a grid so that contour mapping of the characters (Adams, 1969) and populational clustering throughout the range of the species might be made.

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 mountain 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 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 each site and the chemical data will be taken from the same plants from which the morphological data are assembled. For each populational sample, a statistic will be available which can be utilized in the computer-aided classification study. It will be interesting to compare the numerical results with the taxonomic treatment accorded these two genera on morphological criteria, for the character used most frequently to distinguish between the genera has been that of pod shape and it appears that, unless that character were given considerable weight, the distinctions between the genera might be more a matter of convenience than a reflection of phylogeny.

Baptisia - The programs developed will also be applied to a number of systematic problems involving the genus Baptisia. This taxon has been studied intensively by comparative phytochemists at the Bio-chemical Systematic Laboratories in The University of Texas 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 100 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 of chemical constituents from over a broad geographical region have been assembled. Preliminary treatment of some of these data from B. nuttalliana (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.

Numerical Techniques

Characterization. The quality of a classification depends on how well the data reflect or characterize the pertinent biological information. One approach envisioned here for investigating the relative value of data characterization will involve the calculation of the implied relative "weight" of the coded characters in a set of taxonomic data in order to better inform the biological worker about the character weighting resulting from a variety of methods of coding. This will make it less difficult to implement a coding process that reflects the biological worker's estimate of the proper a priori weight that should be assigned to each character. This is a relatively well-defined aspect of the characterization problem, and the previous work of Estabrook (1967) 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 posterioria weight to the taxonomic

characters, for reprocessing 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.

The third aspect of the characterization problem is the determination of the relative weight to be assigned to characters describing populations. An approach has been developed for infraspecific studies which essentially gives more weight to characters that have relatively small variations within natural populations and de-emphasizes characters that have large (seemingly random) populational variances (cf. appendix). It is proposed to extend these techniques to interspecific studies in which the character set may not be uniform for all taxa.

Prediction - Prediction with reference to a classification refers to two sets of 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 statistical methods for estimating the predictive value of a classification. A technique has been developed for estimating the statistically significant portions of a hierarchical structure obtained from applying the numerical classification algorithm to populational data. This new development is reported in the Juniperus study (cf. appendix) and represents an initial contribution to a solution of the problem of estimating predictive quality.

Collaborative Efforts - The computer-aided techniques developed in this research program will also be available to other workers who might wish assistance in evaluating their own data. For example, Dr. J. Hanover, Associate Professor of Forestry, Michigan State University, has an enormous amount of quantitative data for the terpenes in populations of the genus Picea, and he has been collaborating with Dr. Flake in the analysis of these data. In a previous such instance, Dr. Flake worked with Dr. Donald E. Stone of Duke University who had quantitative data for the fatty acids of the genus Carya. This study will soon be published (Stone, D.E., Adrouny, G.A. & Flake, R.H., 1969).

These numerical techniques will also be available to a number of graduate students working with a wide range of organisms and characters, both morphological and 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 computer-aided methods.

References

- Adams, R. P. (1969). "Chemosystematic and Numerical Studies in Natural Populations of Juniperus." Doctoral Dissertation. The University of Texas at Austin, Austin, Texas.
- Brehm, B. 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.
- Crovello, T. J. (1968). "Key Communality Cluster Analysis As a Taxonomic Tool". Taxon 17, 241.
- 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.
- Flake, R. H. and Turner, B. L. (1968). "Numerical Classification for Taxonomic Problems," J. Theoret. Biol. 20, 260.
- Flake, R. H., von Rudloff, E. and Turner, B. L. (1969). "Quantitative Study of Clinal Variation in Juniperus virginiana using Terpenoid Data," Proc. Natl. Acad. Sci. (communicated May 15, 1969).
- Horne, D. B. (1965). "The Distribution of Flavonoids in Baptisia nuttalliana and B. lanceolata and their Taxonomic Implications." Doctoral Dissertation, The University of Texas, Austin, Texas.
- Jardine, N. and Sibson, R. (1968). "The Construction of Hierarchic and Non-Hierarchic Classifications." Computer J. 11, 177.
- Lance, G. H. and Williams, W. T. (1967). "A General Theory of Classificatory Sorting Strategies." Computer J. 9, 373.
- Mayr, Ernst. (1965). "Numerical Phenetics and Taxonomic Theory." Syst. Zool. 14, 73.
- 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. 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 Sneath, P. H. A. (1965). "Principles of Numerical Taxonomy," San Francisco: W. H. Freeman.
- Stone, D. E. Adrouny, G. A., and Flake, R. H. (1969). "New World Juglandaceae. II. Hickory Nut-Oils, Phenetic Similarities, and Evolutionary Implications in the Genus Carya." Am. J. Bot. (in press).
- Turner, B. L. (1967). "Plant Chemosystematics and Phylogeny." J. Pure & Appl. Chem. **14**, 189.
- Vinutha, A. R. and von Rudloff, E. (1969). "Gas-Liquid Chromatography of Terpenes Part XVII. The volatile oils of the leaves of J. virginiana." Canad. J. Chem. **46**(3), 3743.
- Von Rudloff, E., 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**, 600 (Abstract).
- Von Rudloff, E. (1968). "Gas-Liquid Chromatography of Terpenes Part XVI. The volatile oil of the leaves of Juniperus ashei. Buchholz. Canad. J. Chem. **46**, 679.
- Von Rudloff, E., Irving, R. and Turner, B. L. "Reinvestigation of the Case for Allopatric Introgression Between Juniperus ashei and J. virginiana." (In prep.)
- 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.

Equipment

There are two major items of equipment necessary to expedite the proposed study. The first is a large general purpose digital computer required to process the systematic data, and the second is a gas liquid chromatographic apparatus for the terpene analysis.

Digital 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.

Gas Liquid Chromatograph - It has been found that the acquisition of the precise data needed for the computer-aided analysis requires a high-resolution dual-detection analytical instrument fitted with capillary

columns and with a digital read-out integrator of the type found on the Varian 1520 C. instrument. The second year we plan to add a second recorder and integrator to the instrument to double the data output rate.

Period of Performance

It is recommended that the proposed grant, if awarded, be made effective for a period of two years, beginning January 1, 1970, or some other mutually 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.

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: Computer-aided classification, control theory.

Publications: R. H. Flake, "Volterra Series Representation of Nonlinear Systems," IEEE Trans. Part II, 1963, p. 330.

R. H. Flake, "Volterra Series Representation of Time-Varying Nonlinear Systems," Proc. 1963 IFAC (Basel), 2, p. 91

J. Zaborszky & R. H. Flake, "On Fitting with Functional Polynomials," ASME Trans., 89, Series D, 1967, p. 399.

R. H. Flake & J. R. Cox, "Filter Design for the Average Response Computer," 19th SWIEEEO Conf. Record, (Dallas), 1967.

R. H. Flake & J. G. Hartwell, "A Reference Orbit Useful in Integrating Between Activity Spheres," AIAA J., 5, 1967, p. 599.

R. H. Flake & J. H. Satterfield, "A Computer Classification of Psychiatric Patients," 6th S. E. ACM Conf. (Chapel Hill), 1 1967.

R. H. Flake, "Numerical Classification of Biological Populations," (Abstract) Proc. Conf. on Engr. in Med. and Biol. (Houston), 1968.

R. H. Flake & B. L. Turner, "Numerical Classification for Taxonomic Problems," J. Theoret. Biol. 20, p. 260.

D. E. Stone, G. A. Adrouny & R. H. Flake, "New World Juglandaceae. II. Hickory Nut-Oils, Phenetic Similarities, and Evolutionary Implications in the Genus *Carya*," Am. J. Bot., 1969 (in press).

R. H. Flake, E. von Rudloff & B. L. Turner, "Quantitative Study of Clinal Variation in *Juniperus virginiana* using Terpenoid Data," Proc. Natl. Acad. Sci. (communicated May 15, 1969.)

- 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.

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: (last five years)

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)

Chromosome numbers in the Compositae. VIII. Mexican and Central American species. Southwest 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 Southwest 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 *Stenocarpa* (Compositae-Heliantheae-Galinsoginae). The Southwest 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-Heliantheae). Madrono 18: 203-206.

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

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

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 canavine 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. Cuatrecasas)

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.

1968

Miller, H. E., T. J. Mabry and B. L. Turner, 1968, Intraspecific variation of sesquiterpine lactones in Ambrosia psilostachya (Compositae). *Am. J. Bot.* 55: 316-324.

Flake, R. H. and B. L. Turner, 1968. Numerical classification for taxonomic problems. *J. Theoret. Biol.* 20: 260-270.

1969

Turner, B. L., 1969. Problems of Hybrid Nomenclature. Proc. First Int. Citrus Symposium 1: 451-453.

Skvarla, J. J. and B. L. Turner, 1969. Fine structure of Petrobinaeae (Compositae-Heliantheae) pollen walls. *Am. J. Bot.* 56: 418-419.

Turner, B. L., 1969. Chemosystematics: recent developments. *Taxon* 18: 134-151.

Estimated Budget for First Twelve Months
(January 1, 1970 through December 31, 1970)

		<u>NSF Grant</u>	<u>Grantee Share</u>
A.	Salaries and Wages		
1.	Senior Personnel		
a.	2 Co-Principal Investigators 3½ Summer Mos.	\$ 7,111	
	(co-principal investigator, R. H. Flake, two summer mos. @ 9 mo. academic rate of \$14,000)	\$3,111	
	(co-principal investigator, B.L. Turner, 1½ summer mos. @ 9 mo. academic rate of \$24,000)	\$4,000	
2.	Other Personnel		
d.	2 Graduate Students 9 Academic Mos.	8,505	
	6 Summer Mos.		
	1 Res. Sci. Asst. I (½ time 9 mos., full-time 3 summer mos. @ full-time rate of \$514 per mo.)		
	1 Res. Sci. Asst. III (½ time 9 mos., full-time 3 summer mos. @ full-time rate of \$620/mo.)		
f.	1 Pre-Baccalaureate Student 3 Calendar Mos.	1,300	
	(Approx. 650 hrs. @ 2.00/hr for data coding and key punching)		
g.	1 Senior Secretary(15% time 12 mo. @ full-time rate of \$450/mo)	810	
	Total Salaries and Wages	17,726	
B.	Fringe Benefits	560	
	Total of A & B	18,286	
C.	Permanent Equipment (See Budget Notes)	16,100	
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 (3 hours @ \$200 per hr.)	600	
H.	<u>Total Direct Costs</u>	37,236	

I.	<u>Indirect Costs</u>		20
	1. On Campus 32.3% of Salaries and Wages	5,725	
	Less Difference between 18% of TDC exclusive of Permanent Equipment (\$16,100) and computer charges (\$600) and the audited indirect cost rate of 32.3% of S & W	2,029	2,029
	Total Indirect Costs	<u>3,696</u>	<u> </u>
J.	<u>Total Estimated Budget for First Twelve Months</u>	40,932	2,029

Estimated Budget for Second Twelve Months
(January 1, 1971 through December 31, 1971)

21

		<u>NSF Grant</u>	<u>Grantee Share</u>
A. Salaries and Wages			
1. Senior Personnel			
a. 2 Co-Principal Investigators	3½ Summer Mos.	\$ 7,666	
(co-principal investigator, R.H. Flake, two summer mos. @ 9 mo. academic rate of \$15,000)	\$3,333		
(co-principal investigator, B.L. Turner, 1½ summer mos. @ 9 mo. academic rate of \$26,000)	4,333		
2. Other Personnel			
d. 2 Graduate Students	9 Academic Mos.		
	6 Summer Mos.	8,918	
1 Res. Sci. Asst. I (½ time 9 mos., full-time 3 summer mos. @ full-time rate of \$539/mo.)			
1 Res. Sci. Asst. III (½ time 9 mos., full-time 3 summer mos. @ full-time rate of \$650/mo.)			
f. 1 Pre-Baccalaureate Student (approx. 650 hrs. @ 2.00/hr. for data coding and key punching)	3 Calendar Mos.	1,300	
g. 1 Senior Secretary (15% time, total 12 mo. @ full-time rate of \$462/mo.)		832	
	Total Salaries and Wages	18,716	
B. Fringe Benefits		625	
	Total of A & B	19,341	
C. Permanent Equipment (See Budget Notes)		8,100	
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 (3 hrs. @ \$200/hr.)		600	
H. <u>Total Direct Costs</u>		30,291	

I.	<u>Indirect Costs</u>		
1.	On Campus 32.3% of Salaries and Wages	6,045	
	Less Difference between 18% of TDC exclusive of permanent equipment (8,100) and computer charges (600) and the audited indirect cost rate of 32.3% of S & W	2,159	2,519
	Total Indirect Costs	<u>3,886</u>	<u> </u>
J.	<u>Total Estimated Budget for Second Twelve Months</u>	34,177	2,159
	Total Estimated Budget for Twenty-Four Months	75,109	4,188

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 (4.8% of 1st \$7800 of each salary; 5.2% after January 1, 1971); 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.

C. Permanent Equipment

First year - 1 gas chromatograph type Varian 1520 C dual ionization (\$7000), 1 Digital Integrator Type Varian 480 (\$7100), 1 Recorder Type Varian 20 (\$1000), Capillary Columns (\$1000)

Second year - 1 Digital Integrator Type Varian 480 (\$7100), 1 Recorder Type Varian 20 (\$1000)

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.

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 Electronics Research Center

and

Original Signed by

B. L. Turner
Chairman
Department of Botany

APPROVED:

Original Signed by

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

APPENDIX
QUANTITATIVE STUDY OF CLINAL VARIATION IN JUNIPERUS VIRGINIANA
USING TERPENOID DATA*

By Robert H. Flake, [†] Ernst von Rudloff ^{††} and B. L. Turner ^{†††}

^{†††} [†] Department of Electrical Engineering and
Department of Botany and Cell Research Institute
University of Texas at Austin

^{††} Prairie Regional Research Laboratory,
N.R.C., Saskatoon, Canada

Summary

Nine populations of Juniperus virginiana were sampled at approximately 150 mile intervals along a 1500-mile transect from northeastern Texas to Washington, D. C. Individual plants were examined for terpenoids by gas liquid chromatography and the resulting data analyzed by numerical classification methods utilizing characters weighted according to their estimated variance in the natural populations.

The results of the analysis show that these populations of J. virginiana cluster clinally from northeast to southwest, the more homogeneous populations occurring in the Appalachian region of North America; the more divergent populations found in progressively more distant regions, as measured along the transect from the northeast toward the southwest. No biochemical evidence could be found to support the hypothesis that hybridization with J. ashei might be causing this variability, as had been widely supposed previously.

INTRODUCTION

Several North American species of Juniperus have been studied in considerable detail, especially concerning their participation in situations involving hybridization^{1,2}. Hall² analyzed the morphological variation found in six characters of the allopatric species J. ashei and J. virginiana, concluding that the former has influenced the latter by hybridization and subsequent introgression " . . . throughout the Ozark Plateau and probably as far east as the Tennessee River in the vicinity of the 36th parallel." More recently Hall and Carr³, after studying additional putative hybrids and introgressed

individuals in Oklahoma, have stated that genes of J. ashei "...appear in red cedar (J. virginiana) populations of the granite outcrops along the fall line of the Piedmont upland in Georgia."

Clearly what is implied in these studies is introgression or gene-flow on a sub-continental scale. Indeed, the study of hybridization between J. ashei and J. virginiana has been repeatedly referred to as one of the most detailed, best documented examples of allopatric introgression in the literature⁴⁻⁶.

Recent work,⁷⁻⁹ however, using both chemical and morphological data, has failed to substantiate the case for hybridization and introgression between J. ashei and J. virginiana. In fact, at localities where the species occur side by side, including some of the putative hybridized populations sampled by Hall, hybrids and/or introgressed individuals were not found, either in areas of sympatry or in populations removed from regions of contact. Barber and Jackson¹⁰ have been the only other workers to question allopatric introgression as an explanation for the variation found in J. virginiana. They suggested rather that the variation found in the latter species was due to differential selection along an ecological gradient. The so-called introgressed populations of J. virginiana (consisting of trees with wider aprons, shorter terminal shoots, etc.) were believed to be due to selective influences working on a variable gene pool in the western part of its range so as to produce populations superficially like J. ashei. In short, these authors suggest that the morphological variation found in J. virginiana is clinal, i.e. the species has formed or is in the process of forming regional races as a result of adaptational mechanisms arising out of its own gene pool, this being unrelated to the possible influx of genes from the largely allopatric J. ashei.

We believe that the absence of evidence for either sympatric or allopatric hybridization⁷ disproves the introgression model proposed by Hall and others. The present study then has been designed to test the clinal variation model proposed by Barber and Jackson.

As implied in the above discussion, because of the considerable subjectivity inherent in the evaluation of morphological characters, particularly those involving habitual features of the plant, terpenoid characters were selected for intensive study using gas chromatographic techniques. Gas chromatographic techniques are unusually suited for the study of infraspecific variability, especially where the data processing capabilities of modern computers are used in the evaluation of populational data. Modern chromatographic units with automatic print-out devices, accurate to the second decimal place, enable the chemosystematist to procure large sets of quantitized data of an unbiased nature in a matter of minutes. These data need not be subjectively coded for computer analysis since, as indicated below, programs have been devised to handle read-out figures directly.

Using this approach then we ascertained the quantitative variation in terpenoid compounds for individuals both within and between populations; these data were then used in combination to test for clinal variation within the I. virginiana complex.

Confirmation of a clinal structure in the populations comprising I. virginiana was made possible through a clump-searching numerical classification algorithm recently developed¹¹ to assist in problems of biological classification, particularly at the infraspecific level. This numerical procedure yields a hierarchy of collections of mutually similar objects, called aggregations. A brief discussion of the numerical classification method is contained in the Appendix.

The clump-searching algorithm utilizes a matrix of similarity measures containing an element for each pair of objects. The measure of object-pair similarity or, alternatively, a measure of distance (metric) is chosen to reflect the character-by-character similarity of each pair of objects. The selection of the characters used to describe the objects and the selection of a numerical procedure for coding the characters, inherently involves assigning a weight to each character.

Two philosophies of character weighting are considered in this paper. A weighting procedure believed to be consistent with the subjective weight that most taxonomists use in systematic studies, which essentially gives more weight to characters that have relatively small variations within natural populations and deemphasizes characters that have large (seemingly random) populational variations, was utilized in obtaining the results reported here as evidence of a northeast-to-southwest clinal variation in J. virginiana. In addition, results obtained using an Adansonian approach, where equal weight is assigned to each character, are included for comparison.

Finally, an estimate is made of the sensitivity of the hierarchy of aggregations to statistical variations in the population data. The aggregations that, under the assumptions of the statistical analysis, are statistically significant and thus have a reasonably high probability of retaining their relative position in the hierarchy upon resampling, are identified. The resulting hierarchy of statistically significant aggregations which were obtained by weighting the characters according to variance, as opposed to Adansonian weighting, is considered to be the principal result of the analysis.

Population Data

The *J. virginiana* data were collected in January, 1968, in the form of samples of about ten plants from nine population sites (Table 1) at approximately 150-mile intervals on a transect from near Texarkana, Texas, to Washington, D. C. Two additional population samples of ten plants each were collected on the 8th and 9th of December, 1967, near Durham, N. C., and South Hill, Va., respectively; a third population of ten plants was obtained from Ontario, Canada, on the 12th of December, 1967.

Chemical Analysis

A gas chromatographic analysis of the terpinoids in *J. virginiana* yielded percentage distributions of 37 distinct compounds that were detected in sufficient quantities to be well above instrumental error as well as having the appearance of being present in reasonably consistent amounts within the individual populations.

The chemical analysis was described in detail previously⁸⁻⁹. The individual components 1 to 37 used in the analysis were α -pinene and thujene (peaks 1a and 1b), sabinene (peak 3), 3-carene (peak 5), myrcene (peak 6), limonene (peak 7), β -phellandrene (peak 8), γ -terpinene (peak 9), terpinolene (peak 11), unidentified (?thujone, peak 16), unidentified (?isothujone, peak 17), linalool (peak 20), methyl citronellate (peak 21), unidentified (peak 22), 4-terpinenol (peak 23), unidentified (peak 25), estragole (peak 26), unidentified (?borneol, peak 27), methyl vinyl anisole (peak 28), δ -cadinene (peak 29), unidentified (peak 30), citronellol (peak 31), unidentified (peak 32), nerol (peak 33), unidentified (peak 34), geraniol (peak 35), unidentified aliphatic alcohol (peak 37), unidentified (peak 39), methyl eugenol (peak 40), unidentified (peaks 41 and 42), elemol, (peak 44), elemyl acetate (peak 45),

unidentified (peak 46), γ - eudesmol (peak 48), unidentified aromatic compound (peak 49), α - eudesmol and elemicin (peaks 51a and 51b), β - eudesmol (peak 52), unidentified (peak 53), and acetate II (peak 55).

The separation of certain double peaks (e.g. α - eudesmol and elemicin) is difficult to achieve on a single gas-chromatographic column and the error in determining the percentage of such components is considerably higher than the 0.1 percent, which is possible with well resolved peaks. Also, several components (peaks 22, 43, 44, 45) tend to decompose slightly during the gas chromatographic analysis, adding to the errors of the technique.

Data Characterization

The terpenoids had different ranges of variation, and clearly could not vary independently, since only relative percentage values were available. Therefore, before any character weighting was attempted, a set of uncorrelated characters was estimated utilizing a transformation common in principal-axis factor analysis studies constructed from the eigenvectors of the pooled sample correlation matrix. The data for each plant were transformed to yield uncorrelated characters with pooled sample variances of unity and zero means. Data transformed in this manner are usually referred to as factor scores in the terminology of factor analysis.¹²

The Adansonian characterization of each population was obtained by averaging the transformed data over the individual population samples. This was considered to be an Adansonian characterization, because normalizing the variation of each of the uncorrelated characters and making all the weighting coefficients in the taxonomic distance expression equal, tends to equalize each character's influence in the calculation of taxonomic distance or similarity.

In this study the taxonomic distance for a pair of populations indexed j and k was defined as

$$d(j, k) = \left[\sum_{n=1}^N w_n^2 (X_{nj} - X_{nk})^2 \right]^{\frac{1}{2}}$$

where N is the number of uncorrelated characters considered, and X_{nj} is the average value of the n^{th} uncorrelated character from the sample of the j^{th} population. The weighting coefficients, w_n , $n = 1, 2, \dots, N$, satisfy the constraint $\sum_{n=1}^N w_n^2 = 1$.

The measure of similarity for populations j and k was defined as

$$S(j, k) = 1 - (0.1) d(j, k),$$

where the factor of 0.1 was conveniently chosen to ensure that $0 \leq S(j, k) \leq 1$, for all object pairs. (In this study a measure of similarity was introduced only because the computer software required a matrix of similarity values rather than a matrix of taxonomic distances.)

The number of uncorrelated characters produced by the transformation of eigenvectors is limited by the rank of the sample correlation matrix. The rank of the correlation matrix for this pooled *J. virginiana* data was 36. Eigenvectors with small eigenvalues are often omitted in the transformation reducing the number of transformed characters on the basis that they may be due to "random error".¹³ The question of exactly how many characters to use in the calculation of taxonomic distance is especially important in an Adansonian analysis, since each character is equally weighted in the calculation. However, this type of analysis is not considered further because the results from the Adansonian characterization are somewhat secondary to the discussion, and are presented here only for gross comparison with the principal results based on characters weighted according to variance. Weighting characters according to variance effectively

accomplishes the task of eliminating characters that are due to "random error" by de-emphasizing the characters that are relatively variable.

This method of character weighting, which is considered to be more consistent with present systematic methods than the Adansonian approach¹⁴, is based on using weighting coefficients in the taxonomic distance expression determined from the average of the individual population sample standard deviation estimates of each uncorrelated character and the estimated pooled sample standard deviations of each character.

The weighting coefficient for character n was defined as

$$w_n(s_{pn}, s_n) = \begin{cases} (1 - s_{pn}/s_n), & \text{if } s_{pn} \leq s_n \\ 0, & \text{if } s_{pn} > s_n \end{cases}$$

where s_{pn} is the average of the population sample estimates of the standard deviations for character n , and s_n is the pooled sample standard deviation estimate for the n^{th} character.

The significance of these weighting coefficients is that they give relatively large fractional weight to those characters that tend to have small variations on the average within each population sample, and simultaneously have different population means which result in relatively large pooled sample variations.

The specific functional form of the weighting coefficients is clearly somewhat arbitrary, just as the functional form for the taxonomic distance is somewhat arbitrary: systematic principles yield only guidelines here that are too general to determine any unique expressions for taxonomic distance or weighting factors.

Results and Interpretation

Adansonian Characterization - The hierarchy of aggregations illustrated in the form of a contour map for the twelve J. virginiana populations using the Adansonian data

characterization is shown in Figure 1. The various aggregation levels are noted on the contours in a manner similar to elevation levels shown on a surface contour map. The aggregation with the highest level contains populations 6 and 7, which appear approximately in the middle of the range of the transect from near Texarkana, Texas to Washington, D.C. The aggregations at successive lower levels tend, grossly speaking, to include more and more populations on both sides of the middle until the level of the last aggregation is reached where the Canadian population is finally included with the other eleven populations. However, an inspection of some of the finer details of the hierarchy reveals a few apparent discrepancies with this general trend. For example, populations 10 and 11, which are near Durham, N. C., do not appear in the hierarchy of aggregations at the level in which their near neighbor population 8 is included, but instead become included, only at a much lower level. This is an indication that the quantitative distribution of the terpenoids in populations 10 and 11 was not particularly similar, in a relative sense, with the distribution found in their nearest neighbor. However, this was not entirely unexpected because, as indicated in the previous section, populations 10 and 11 were sampled approximately five weeks earlier (before winter temperatures had firmly set in) than the others on the transect, and there is reason to believe that terpenoids vary quantitatively not only from season to season, but often during different months of the season, especially in foliage samples of the type collected in this study.¹⁵⁻¹⁸ The fact that populations 10 and 11 clustered according to the season collected (late fall) and not according to the geographical proximity of nearest neighbor populations speaks for the sensitivity of the combined approach of chromatographic data analyzed by the numerical classification techniques

used here, and lends credulity to the northeast to southwest clinal structure discussed in the next section.

The other notable discrepancy in this general trend in Figure 1 is that population 3 is included in the hierarchy of aggregations at a level higher than the levels at which populations 4 and 5 are enclosed. However, an inspection of the levels involved makes it apparent that, in contrast with the previous case relating to populations 10 and 11, the levels in this instance are very nearly equal. This suggests that perhaps this particular order of aggregations in the hierarchy may not be expected to be preserved under the purely statistical fluctuations of repeated resampling even in the dormant winter period when the terpenoids are relatively stable.

Weighted Characters A contour map showing the hierarchy of aggregations for the twelve J. virginiana populations, using characters weighted according to variance, is shown in Figure 2. The aggregation at the highest level contains populations 8 and 9, located near Durham, North Carolina and Washington, D. C. At lower levels, the aggregations tend to enclose successive populations along the transect from Washington, D. C., to the vicinity of Texarkana, Texas, indicating a northeast to southwest clinal variation in J. virginiana, differing in this respect from the results shown in Figure 1 using an Adansonian characterization. However, there are, as in the previous Adansonian results, discrepancies with the general trend. Populations 10 and 11, near Durham, N. C., which were sampled five weeks earlier than the others on the transect, again appear to be relatively distinct from their neighbor population 8. Similarly, Population 3 appears in the hierarchy of aggregations at a slightly higher level than populations 4 and 5. However, the aggregation levels involved in this latter instance are again very nearly equal and relatively small variations in sample means for these populations would alter the order of these aggregations in the hierarchy. Clearly, an analysis yielding an

estimate of the statistical significance of the ordered aggregations in a hierarchy is desirable. Fortunately an estimate of the significance of the order of aggregations in a hierarchical arrangement may be obtained for this particular numerical classification procedure, since the aggregations are composed of intersecting clusters with cluster similarities above a certain level, where a cluster similarity is defined as the average of the measures of similarity of all the distinct pairs of objects it contains. The question, therefore, can be resolved by testing the significance of the difference in means of similarity measures using a statistical test such as the Duncan's New Multiple Range Test¹⁷.

Consider, for example, the apparent discrepancy in the general trend of a northeast to southwest clinal variation when Population 3 appeared in the hierarchy of aggregations (Fig. 2) at a slightly higher level than the level at which Population 4 is included. The statistical significance of such apparent discrepancies may be tested by applying a multiple range test to the cluster similarity levels of the clusters responsible for forming the ordered hierarchy of aggregations. If the difference between the mean of the similarity measures for the cluster (or clusters) containing Population 3 that is included in the hierarchy of aggregations at a level above the first cluster that appears in the hierarchy containing Population 4, and the mean of similarity measures of the latter cluster is significant in at least one instance with a sufficiently high probability, then the discrepancy is accepted; otherwise the two populations are considered to be included in the hierarchy at the same level, and the aggregation contours are merged.

A test of the statistical significance of the aggregations containing the populations along the transect was performed utilizing the New Multiple Range Test, at the

5% level. This test assumes that the variances of the Populations being sampled are all equal. The estimated variance associated with Population 5 was notably larger than those associated with the other populations on the transect. Variance ratio tests involving Population 5 with its neighboring populations were significant at the 1% level and, as a result, Population 5 was not considered in the calculations of the New Multiple Range Test.

The results of the statistical analysis showing the aggregations with statistically significant level differences may be summarized by listing the population identification numbers in the order in which they appear in the hierarchy of aggregations in Figure 2 as follows:

1 10 11 2 4 3 6 7 8 9

These population numbers may be used to specify the various aggregations by associating each aggregation with the population that entered the hierarchy at the aggregation's level. Those population numbers that are underscored by the same line identify aggregations with insignificant level differences. Those population numbers not underscored by the same line identify aggregations with significantly different levels. These results are illustrated by the aggregation contours in Figure 3. The dotted contour sections in the vicinity of Populations 6 and 10 signify the corresponding overlap of the significant level ranges.

Since results from both unequally-weighted data and an Adansonian characterization are presented in this paper, it seems pertinent to compare the hierarchical structures produced by these approaches with phytogeographical theory. Both weighting methods reveal a clinal structure in the populations: the Adansonian characterization results

in hierarchical clustering from the center of the cline outward toward both the north-east and southwest (Fig. 1); that using unequally-weighted characters results in hierarchical clustering from the northeast to southwest (Fig. 2). The latter approach shows the Appalachian populations (6, 7, 8, and 9) as clustering homogeneously at the highest significant level of similarity with the more geographically distant populations, as measured along the transect from the northeast toward the southwest, clustering at successively lower levels of similarity.

The hierarchical structure resulting from the unequally-weighted characterization is more consistent with what is believed to have been the phyletic history of the extant populations that comprise the species; i.e. that J. virginiana was originally a species belonging to an essentially northeastern flora, having occupied the ancient land mass of Appalachia in remote times but subsequently extending its range to peripheral western areas as previously submerged land became newly exposed for occupancy.⁴

* * *

*Supported by research grants GM-11111-04A1 from the National Institutes of Health, National Science Foundation grant numbers NSF-GB-5448X and the Department of Defense's Joint Services Electronics Program under grant AF-AFOSR-766-67.

We wish to thank Dr. W. Morsink, Faculty of Forestry, University of Toronto for collecting the juniper samples from Ontario, and Mr. M. Granat for excellent technical assistance in the chemical analyses.

References

1. Fassett, N. C., Bull. Torr. Bot. Club 72, 480-482 (1945).
2. Hall, M. T. Ann. Mo. Bot. Gard. 39, 1-64 (1952).
3. Hall, M. T. and C. J. Carr. Butler Univ. Bot. Stud. 14, 21-40 (1964).
4. Anderson, E. Biol. Rev. 28, 280-307 (1953).
5. Grant, V. Cold Spring Symp. Quant. Biol. 23, 337-363 (1958).
6. Davis, P. H. and V. H. Heywood. Principles of Angiosperm Taxonomy. Oliver and Boyd (1963).
7. von Rudloff, E., R. Irving and B. L. Turner. Amer. J. Bot. 54, 660 (1967); unpublished data.
8. von Rudloff, E. Can. J. Chem. 46, 679-683 (1968)
9. Vinutha, A. R. and E. von Rudloff. Can. J. Chem. 46, 3743-3750 (1968)
10. Barber, H. N. and W. D. Jackson. Nature 179, 1267-1269 (1957).
11. Flake, R. H. and B. L. Turner. J. Theoret. Biol. 20, 260-270 (1968).
12. Cooley, W. W. and P. R. Lohnes. Multivariate Procedures for the Behavioral Sciences. Wiley (1962).
13. Lawley, D. N. Biometrika 43, 128-136 (1956).
14. Farris, J. S. Evolution, Lancaster, Pa. 20, 587-591 (1966).
15. Adams, R. P. Chemosystematic and Numerical Studies in Natural Populations of Juniperus. Doctoral Thesis Univ. of Texas (1969).
16. von Rudloff, E. Can. J. Bot. 45, 891-901 (1967).
17. Steel, G. D. and J. H. Torrie. Principles and Procedures of Statistics. McGraw-Hill (1960).

Table 1. Collection data for the populations indicated in Figures 1, 2 and 3.*

pop. no.	locality	date collected
1.	Texas: Panola Co.	19 Jan 1968
2.	Louisiana: Claiborne Par.	19 Jan 1968
3.	Mississippi: Leflore Co.	20 Jan 1968
4.	Mississippi: Monroe Co.	20 Jan 1968
5.	Alabama: Lawrence Co.	20 Jan 1968
6.	Georgia: Dade Co.	21 Jan 1968
7.	North Carolina: Catawba Co.	22 Jan 1968
8.	North Carolina: Alamance Co.	22 Jan 1968
9.	Maryland: Prince George Co.	25 Jan 1968
10.	North Carolina: Durham Co.	8 Dec 1967
11.	Virginia: Mecklenburg Co.	9 Dec 1967
12.	Canada: Ontario, St. Williams and Rondeau Park Area	12 Dec 1967

*Voucher material deposited at the University of Texas Herbarium.

Appendix: Numerical Classification Concepts

The clump-searching numerical classification procedure utilized in this study is based on a quantitative notion of a "group of similar objects" referred to as a clump or cluster, and consists of the construction of a hierarchical arrangement of such clusters¹¹. This concept of a cluster has been defined in terms of "fuzzy" sets.

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 .

Definition of Cluster Let X be the space of objects under consideration, and define a suitable measure of similarity for each pair of objects in X . Then let $\mathcal{O}$ be the class of subsets $\mathcal{o}$ of elements of X such that each set $\mathcal{o}$ contains two or more elements of X . Let C be the fuzzy set (class) of sets $\mathcal{o}$ in $\mathcal{O}$ with membership function f_C , where $f_C(\mathcal{o})$ is the average of the measures of similarity of all distinct pairs of elements in $\mathcal{o}$. Each set $\mathcal{o}$ is called a cluster and the numerical value $f_C(\mathcal{o})$ is a measure of cluster similarity of the cluster $\mathcal{o}$, and C is a fuzzy class of clusters.

The clusters of $\mathcal{O}$ are ordered according to their cluster similarity in this numerical classification method, and a hierarchy of partitions is constructed by forming particular collections of the ordered clusters called "aggregations".

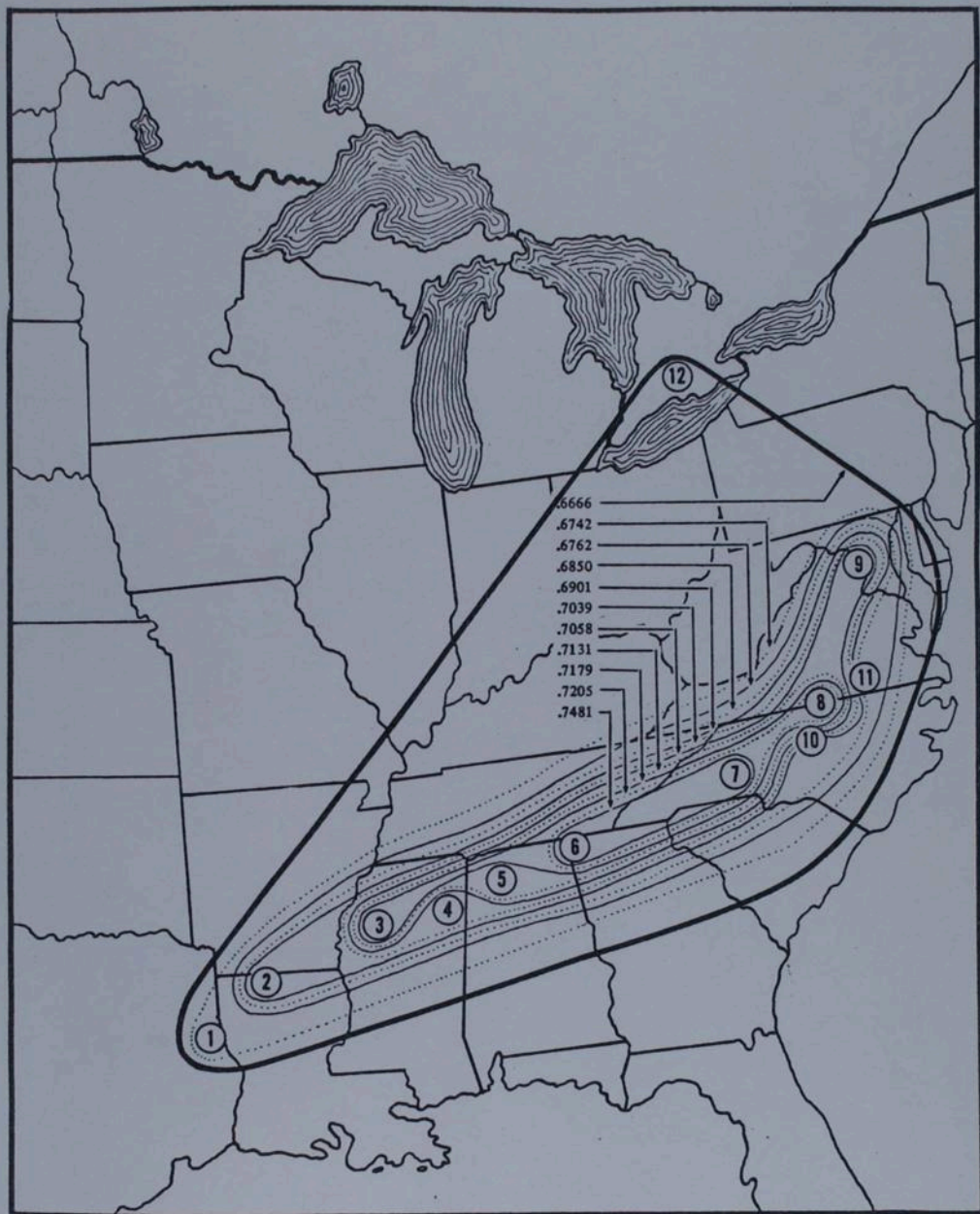
Definition of Aggregation Let $\mathcal{O}_s = \{ \mathcal{o}_i : \mathcal{o}_i \in \mathcal{O}, i \in I_s, f_C(\mathcal{o}_i) \geq s \}$ be the subclass $\mathcal{O}_s$ of $\mathcal{O}$ indexed by a set I_s , such that a cluster $\mathcal{o}_i$ is a member of $\mathcal{O}_s$ if the cluster

similarity $f_C(o_i)$ is at least as high as the level s , where s is some real number in $[0, 1]$. Now let $I_{A_s} \subseteq I_s$ be a subset of the indexing set for 0_s which identifies the sets contained in a particular subclass A_s of 0_s having the following restrictive properties: 1) the union of the sets indexed by I_{A_s} is disjoint from all the other sets in 0_s , and 2) the union of any nonempty proper subcollection of the sets indexed by I_{A_s} shares at least one element with at least one of the other sets indexed by I_{A_s} . Then the class $A_s = \{o_i : i \in I_{A_s}\}$ is called an aggregation in 0_s at the level s . The collection of all aggregations at level s forms a partition of 0_s , and a hierarchy of partitions is constructed by ordering aggregations according to their level s .

The gross nature of the hierarchy is easily illustrated as indicated by the contour maps in this paper, by simply identifying the distinct objects each aggregation contains rather than attempting to enumerate the individual intersecting clusters contained in the aggregations.

FIGURE CAPTIONS

- Fig. 1. I. virginiana Aggregation Contours - Adansonian Characterization
- Fig. 2. I. virginiana Aggregation Contours - Weighted Characters
- Fig. 3. Statistically Significant I. virginiana Aggregation Contours -
Weighted Characters



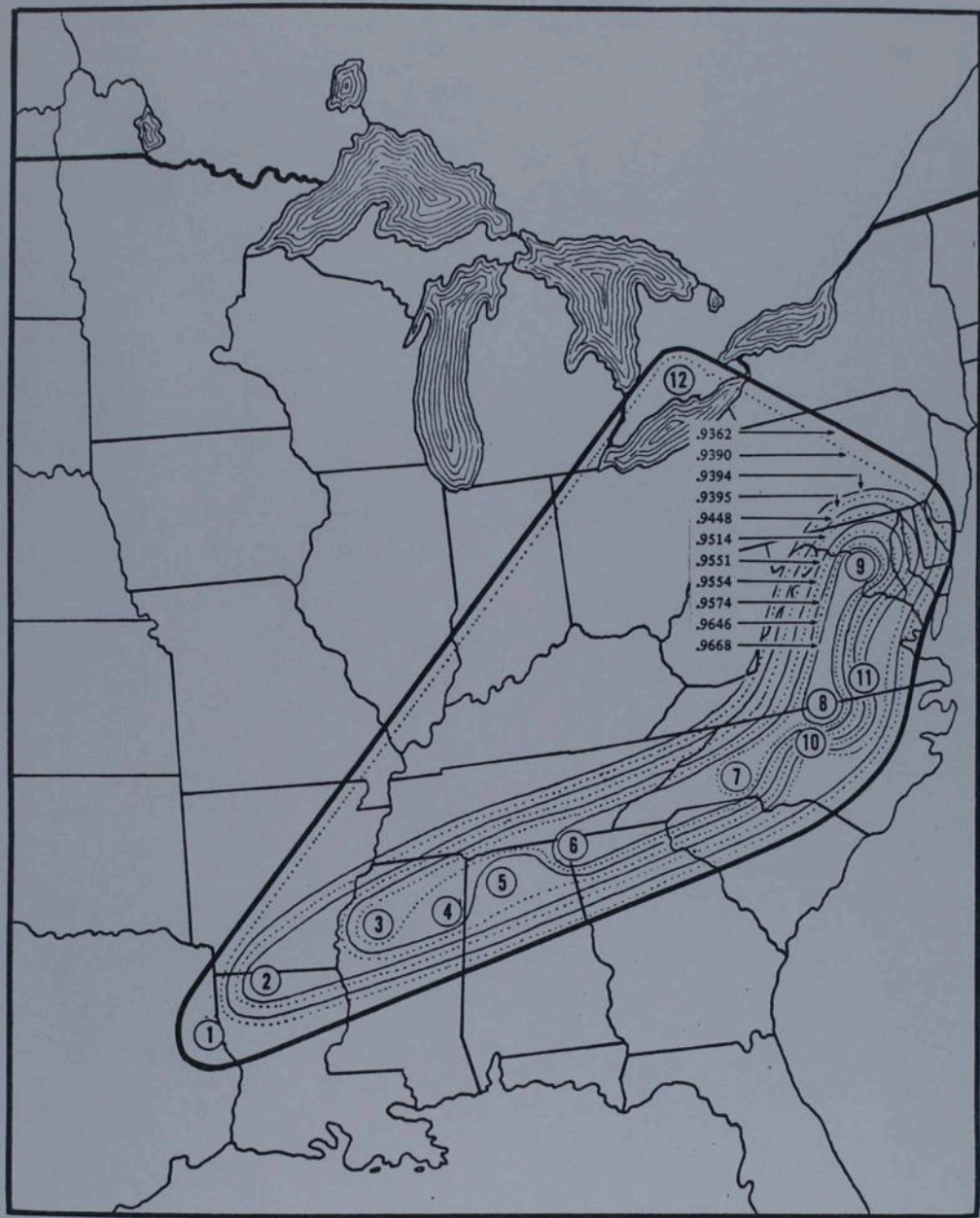


Fig. 2

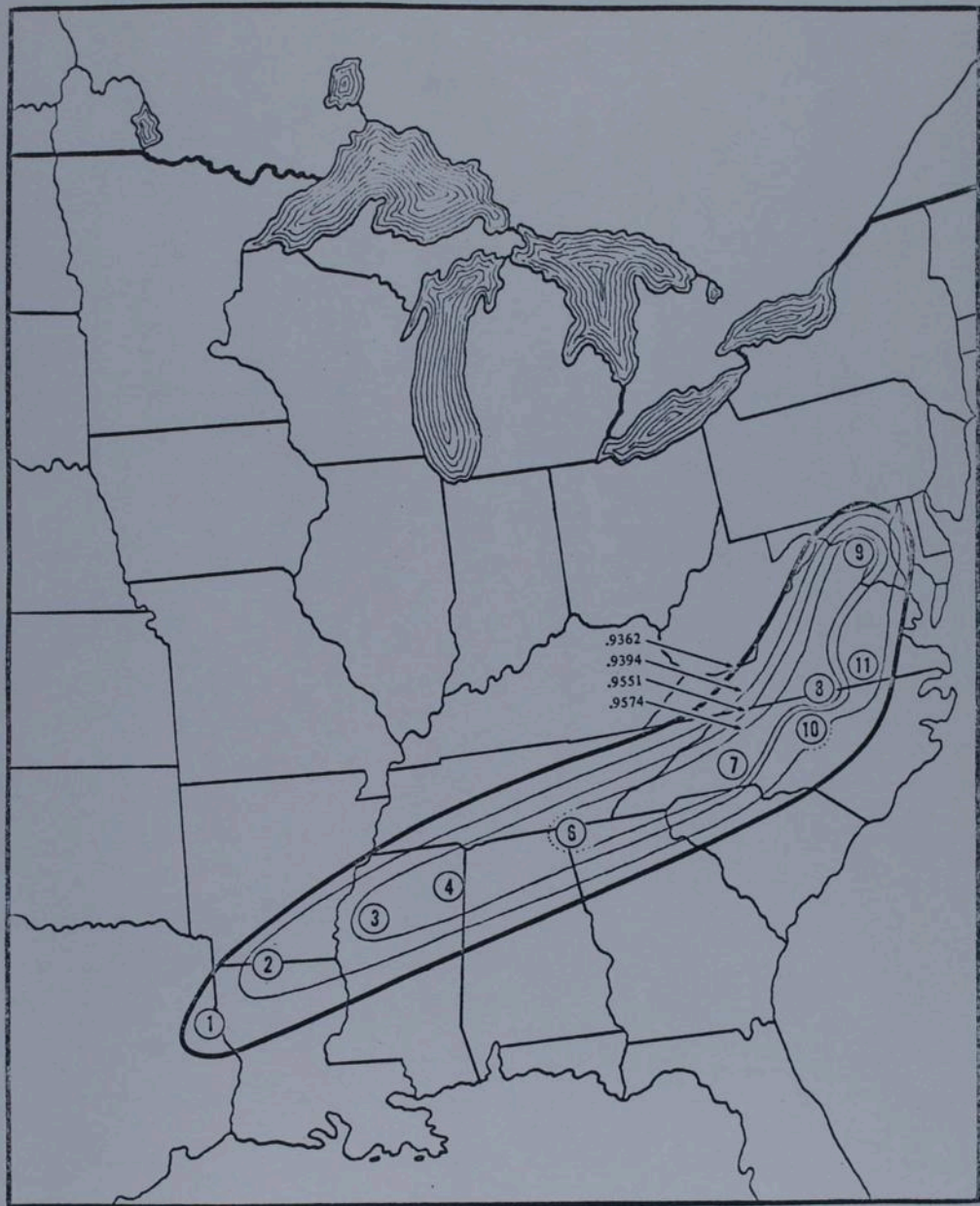


Fig. 3
Digitized by the Hunt Institute for Botanical Documentation

25 July 1969

Dr. John R. Reeder
Editor-in-Chief, BRITTONIA
Department of Botany
University of Wyoming
Laramie, Wyoming 82070

Dear John:

At long last I've gotten to Grady Webster's paper on
Phyllanthus. I think the comments speak for themselves.
He has done quite a job.

Sincerely,

David J. Rogers
Professor of Biology

DJR:GM

for Britton
July 24, 1969

file

Review of "A Revision of *Phyllanthus* (Euphorbiaceae) in the Continental United States" by G. L. Webster

Webster's work is very valuable - anyone willing and able to work in this difficult family deserves publication, and Grady has done a service in sorting out the *Phyllanthus* taxa of the U. S.

Generally the format of presentation is good, and the following comments are intended as suggestions which you can accept or reject.

There are one or two inconsistencies noted in underlining. Case in point: Usually the collector of the type specimen of a species is underlined; for species #7, p. 34, this has not been done. There may be others but I did not see them.

The section on geographic distribution seems to be almost a separate paper on biogeography of the southern U.S. and Mexico, with interpretations more significant to biogeography than is needed in this paper.

However, the morphology section says little about some of the characters used in the key - for example in #3 in the contrasts between leaves "scarcely pungently tipped" vs. "pungently tipped" - does pungent refer to sharp and spiny, or does it refer to smell? I know he refers us to another paper on more extensive details of morphology, but my philosophy here is that this paper should stand by itself, and not require the user to have another reference available. Therefore, a reiteration, or summary, of the characters used in the keys of this paper must be added.

With respect to descriptions of taxa, I suppose there are many discussions about proper styles of presentation, but I personally cannot tolerate negative statements, which tell only what the characters aren't, but not what they are. Example, p. 14, Subg. *Isocladus*, "... ultimate axes not regularly deciduous, leaves on main axes not reduced to scales ..." etc.

With respect to illustrations, I am a strong believer in habit photos. If Webster has photos of each of the taxa (and I should think he would) these should be added.

Please note the problem of captions in Fig. 2. Caption indicates cb and cs as indicated in photo, but photo does not point to these features.

NATIONAL SCIENCE FOUNDATION
PROPOSAL RATING SHEET

Reviewer George F. Estabrook
Taximetrics Laboratory
Armory 101
University of Colorado
Boulder, Colo. 80302

Proposal No B016154
Investigator Rohlf and Sokal
Institution: SUNY/Stony Brook
Please return to: Systematic Biology
if possible by: 7 JUL 1969

Comments (Continue on additional sheet if necessary)

There are two main efforts to be undertaken in this proposed project:

- 1) (NT-SYS) the implementation (or more properly, the completion of the implementation) of a collection of computer programs for processing classificatory information, and
- 2) (Taxocrit Experiment) an alleged investigation into the thought processes of biological classifiers using drawings and models of artifacts completely artificial biologically.

Section 3 of the proposal, Report of Work Done from Inception of Project to Date, with the exception of A6 (effort 2 as described above) and G17 (effort 1 as described above) does not describe what work this proposal will explicitly cover but rather provides contextual history and precedent justification for effort 1, above.

I intend to address my comments to the specific activities proposed (1 and 2 above) and defer comment on the rest of Section 3 except generally as below.

Effort 1: There is very little question that these investigators have functioned in the recent past (Univ. of Kansas) as a service team for much of the biological community. In essence, this proposal is asking for support to enable them to continue to do so at SUNY, Stony Brook. In spite of the evidence that there is some question arising from some quarters of the biological community as to whether this service is indeed not a disservice (and I myself question the validity of some aspects of their work), it is my resolute opinion that they should be financed to continue this controversial adventure into theoretical biology. The stimulation which this group, as questionable as some of their activities may be, has given biological thought in the past few years has itself constituted a major contribution to biological research. It would be a severe loss

OVERALL RATING

- ☒ EXCELLENT Effort 1
☐ VERY GOOD
☐ GOOD
☐ FAIR
☒ POOR

Signature of Reviewer: *George F. Estabrook*

Date:

Other suggested reviewers (optional):

Effort 2 which, if deleted, will result in a very worthy proposal

to science if this group were not allowed to continue to provoke us to a less traditional examination of our basic thought processes in Biology.

Effort 2: With respect to the "Taxocrit Experiment" I feel that it is truly unfortunate that what is, in essence, a very worthy proposal for support should be embarrassed with this insult to Biology. The proposal would be strengthened considerably if the investigators did not ask for the overt endorsement of this "psychological experiment" by the Systematic Panel. This grant request should be for the single worthy project of mounting the NT SYS at SUNY, Stony Brook. I am reasonably certain that the budget requested is less sufficient for that single purpose as it is. It would be to everyone's advantage to keep separate such unrelated (on the basis of scientific supportability) activities as these two. If the investigators are intent on pursuing their study of the psychological reactions of biologists (and others) to someone's drawings of imaginary creatures, the project should be written into its own proposal so that it can be evaluated on its own merits. It should not be allowed to compromise an otherwise worthy request.

NSF CODE NO.

042 1460

RESEARCH PROPOSAL
(Transfer of present Grant GB-4927)

B016154-23MY69

Submitted to:

NATIONAL SCIENCE FOUNDATION
Biological and Medical Science Division
Systematic Biology Branch
Washington, D. C. 20550

Submitted by:

The Research Foundation of the State University
of New York
P. O. Box 7126
Albany, New York 12224
for and in conjunction withState University of New York
CampusState University of New York at Stony Brook
Stony Brook, Long Island, New York 11790

Research site:

Department of Biological Sciences
SUNY at Stony Brook
Stony Brook, New York 11790
(516) 246-5030

Co-Principal Investigators:

Dr. F. James Rohlf
Associate Professor of Biological SciencesDr. Robert R. Sokal
Professor of Biological Sciences

Address all correspondence to Dr. Sokal.

Title of research:

The Development and Testing of Methods of
Numerical Taxonomy

Desired starting date:

1 July 1969

Proposed duration of grant:

6 months

Amount requested:

\$21,371

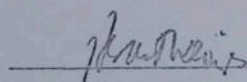
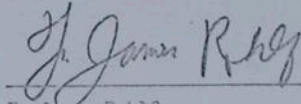
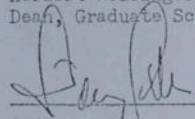
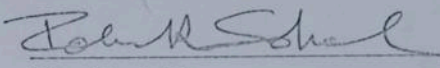
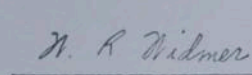
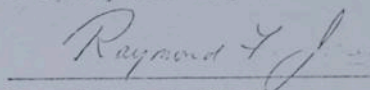

Herbert Weisinger
Dean, Graduate School
F. James Rohlf
Co-Principal Investigator
Sidney Gelber
Acting Vice President for Liberal Studies
Robert R. Sokal
Co-Principal Investigator
Research Foundation
State University of New York
Raymond F. Jones, Chairman
Department of Biological Sciences

TABLE OF CONTENTS

Information relating to administration of this grant	ii
I. Abstracts of reports and proposals	1
(1) Proposed research for 6-month period 1 July 1969 - 31 December 1969	1
(2) Results to date	1
II. Detailed proposals and reports	2
(1) Proposed research during the 6-month period 1 July 1969 - 31 December 1969	2
(2) The intellectual milieu for the proposed research.....	2
(3) Report of work done from inception of project to date	3
(4) Bibliography of publications resulting from this project since its inception	18
(5) Facilities	21
(6) Personnel	21
(7) Proposed budget	26
(8) Explanation of budget	27
(9) Current research support of co-principal investigators	28
(10) Applications for related research to other agencies	29

Information Relating to the Transfer
of this Grant

Drs. Rohlf and Sokal, two of the co-principal investigators, were appointed to the staff of the State University of New York at Stony Brook beginning 1 September 1968. They will move to Stony Brook early in June of 1969. Dr. J. H. Camin, one of the current co-principal investigators at The University of Kansas, does not plan to continue in this field of research. The funds remaining in this grant should carry us through 31 December 1969 as detailed in the budget section.

A renewal application (042-1460) jointly with Dr. J. S. Farris, who is already at SUNY Stony Brook, has been recommended for funding to start 1 January 1970 for two years.

I. Abstracts of reports and proposals.

(1) Proposed research for 6-month period, 1 July 1969 - 31 December 1969.

The main undertaking will be the implementation and further development of the NT-SIS system of computer programs for numerical taxonomy at the Computation Center of the State University of New York at Stony Brook to enable future intensive work in this field to be carried out efficiently by the co-principal investigators and workers at other institutions. These programs will be distributed widely to workers at many institutions.

The applicants also propose to carry out the analysis and interpretation of the Taxocrit study which is an investigation into the nature of taxonomic judgements.

(2) Results to date.

Work has been carried out on 9 separate approaches to the development and application of numerical taxonomy and related subjects. These approaches range from the general theory of systematics, over the principles of taxonomic methodology, to numerical classifications of various groups of organisms. In aggregate they have led to 23 separate research projects, some completed and published, others still underway. An extensive programming effort has been made for numerical taxonomy and this is benefitting not only our work but that of colleagues at other institutions. A total of 18 publications are in various stages of preparation from work which has already been completed. Parallel and interrelated studies in numerical taxonomy financed by the National Institutes of Health are briefly described. These have led to another 11 publications in the field.

II. Detailed reports and proposals.

(1) Proposed research during the 6-month period - 1 July 1969 - 31 December 1969.

Our main effort will be directed at implementation of the NT-SYS system of computer programs (see item 17 in the report of work done, below) at the SUNY Stony Brook Computation Center. The success of our proposed research during the following 2 years will depend materially on the facility with which computational work can be carried out. Not only will this be necessary for our own work but implementation of NT-SYS will also facilitate the work of numerous other scientists at Stony Brook, ranging from geologists to psychologists and anthropologists.

Also we are constantly requested by colleagues at other institutions to run sets of data for them and we consider such service (within reasonable limits) to be an obligation which we willingly meet in return for the support that our program development has received from Federal Granting Agencies. For individuals with limited and infrequent data sets, it is generally less expensive and more efficient to have their data processed by us. We expect to develop at Stony Brook a program library and expertise, similar to that established over the years at The University of Kansas.

In the six months our remaining funds (and time) will not permit much research beyond the implementation of our programs. However, we hope to complete one major research study, the taxocrit experiment, described in some detail below as item 6 in the report of work done. The data for this study have been collected but all the analysis and interpretation remain.

Our proposed research for the years 1970-1971 has already been described in our previous renewal application.

(2) The intellectual milieu for the proposed research.

The research will be carried out in the Section on Ecology and Evolution being organized as a subdivision of the Department of Biological Sciences at The State University of New York at Stony Brook. This section will eventually assume departmental status. It is being organized under the chairmanship of Professor Lawrence B. Slobodkin, formerly of The University of Michigan and currently consists of sixteen staff members. Three further staff members have been committed for the 1969-1970 academic year. The establishment of a strong mathematically and statistically oriented unit in ecology and evolution is one of the primary objectives of the biological sciences program at SUNY Stony Brook, and the early establishment of strength in the field will greatly ease the recruitment of further staff members for the general field of population biology, which is currently undergoing rapid and extensive development. The interaction of a research group in numerical taxonomy with others in the general field of population biology should lead to mutually beneficial results.

The main emphasis in the past work of the co-principal investigators has been on the development of a general theory and methodology for systematics rather than on the detailed classification of specific groups of organisms, although all three of us have done this as by-products of our general and methodological researches. For this reason we expect the results of the research carried out under this grant to have broad applicability in the field of systematics. In order to apply our new ideas and techniques to the field of systematics, interaction with various practicing systematists will be sought. We plan to collaborate with workers in the Department of Biological Sciences, the newly formed Marine Sciences Center, and the Department of Earth and Space Sciences at SUNY Stony Brook, and The American Museum of Natural History. We propose to make arrangements whereby our students may serve research apprenticeships in The American Museum of Natural History, using its collections for numerical taxonomic theses.

(3) Report of work done from inception of project to date.

This report is subdivided into eleven separate approaches (labeled A through J), differing considerably in extent, amount of work done, success of the study, and importance of the work. Under each approach there are one or more studies, representing separate research projects. These projects are numbered continuously from 1 to 24 over all the approaches to permit easy cross-reference. For background information regarding our work for the present reviewing panel we also include brief accounts of projects carried out prior to the current grant period.

Although it is in the nature of this work that studies under most of the approaches are continued as new techniques and insights develop, several of the studies have reached states of temporary completion, and publications or manuscripts in press have resulted from them.

A. THE GENERAL THEORY OF SYSTEMATICS

Numerical taxonomy is undergoing continuous change and elaboration; the results of this work profoundly affect the research carried out under the various approaches listed below.

1. A review of numerical taxonomy.

This paper aimed at a botanical audience and prepared at the invitation of the American Society of Plant Taxonomists was subsidized by this grant and has been published (Sokal, 1963).

2. Nomenclature.

C. D. Michener has been interested in integrating numerical with orthodox taxonomy. He has advocated the application of a uninominal nomenclature as well as of an appropriate numbering system for organisms to accompany the advent of electronic data processing in taxonomy. These views are stated in a general article on "Future Developments in Taxonomy" (Michener, 1963) and an illustrative example of his nomenclatural technique is featured in a subsequent article (Michener, 1964).

3. A re-examination of the principles of numerical taxonomy.

During the past year Sokal has been engaged in an extensive review of the theory of systematics from the point of view of the numerical taxonomist. This work is in connection with the preparation of the second edition of the "Principles of Numerical Taxonomy" in collaboration with P. H. A. Sneath. This will be an entirely rewritten volume. Development of the theory will range over the entire field of systematics. Much of the theoretical work has been completed and a summary of the present status of the field has been given in two extensive review papers published in Russia at the invitation of the editor of the Soviet Journal of General Biology (Sokal 1967, 1968).

4. Relations between phenetics and cladistics.

A paper by Camin entitled "The significance of incongruence between phenetic and cladistic dendrograms," has been completed. The paper explores the biological meaning of phenetic-cladistic congruence and incongruence and discusses the possible application of comparative studies of phenetics and cladistics for the same group and of organisms and characters in providing some insight into relative rates of evolutionary divergence in different sections of such groups.

5. A critical evaluation of the biological species concept.

This study by Sokal and Crovello (not anticipated in the original application) grew out of the extension of the methods and philosophy of numerical taxonomy to studies at the species level. It appeared to these authors that the biological species was not an operational concept (as had also been claimed by others, Ehrlich and Holm, 1962). It therefore seemed worthwhile to develop a model in which we would go step by step through a procedure which would enable us to ascertain whether a given sample of organisms comprised one or more biological species. Even under optimal conditions when full knowledge about biological operations and relations is available, it became apparent that phenetic decisions must be made at certain stages in the procedure, or no progress can be made. Consequently, the biological species concept is clearly not operational under ordinary circumstances, when knowledge and time are limited.

These authors then examine whether the biological species concept has heuristic value or is necessary for biological theory. They find that neither is the case and recommend that attempts to describe biological species be superceded by the description of phenetic and reproductive relationships among organisms without any preconceived hierarchic arrangement. This manuscript has just been completed and will be submitted for publication soon.

B. PRINCIPLES OF TAXONOMIC METHODOLOGY

We have been much concerned with the analysis of the principles by which conventional taxonomists operate when classifying organisms and with the principles which would seem suitable for such classificatory operations whether done by men or by computer. A significant step in this direction has been our study of the "intelligent ignoramus" which is discussed in Section J below.

6. The Taxocrit study.

This is a related study employing the Caminalcules, hypothetical imaginary animals used earlier for cladistic research. In this study numerous students, colleagues, and even laymen were asked to classify the 29 "species" or Caminalcules into natural groups by whatever criteria they wished to employ. These classifications will be quantified and compared with a detailed phenetic numerical taxonomic study of this group in which a large number of characters had been scored. The method employed is borrowed from a personnel evaluation study by Ward and Christal and will indicate to us which aspects of the phenetics of the Caminalcules different taxonomists emphasize in making their classification. It will be interesting to see what groupings exist among taxonomists in terms of their classificatory procedures and how these are related to their perception of form differences. The experimental part of this study has been completed and only the analysis remains.

7. The robustness of numerical taxonomic methods with respect to random errors in homology of characters.

Fisher and Rohlf (1969) investigated the effect of errors in determining the homology of characters on the results of a numerical taxonomic study. It was found that the distance coefficient was much less sensitive to these random errors in the data matrices. These analyses also suggested that in a study based on 74 characters the classification was not degraded by confusion about homologies for six characters, but was adversely affected by a larger number of miscoded characters.

8. Test of hierarchic structure in taxonomic data.

Rohlf and Fisher (1968) established approximate critical bounds for testing for the presence of hierarchic structure in taxonomic data constructed using random samples from multivariate normal and uniform distributions. Sokal's taxonomic distance coefficient was found to be much more sensitive to differences in multivariate structure of the sampled distributions than was the correlation coefficient.

C. REPRESENTATION OF TAXONOMIC STRUCTURE

An important part of the work of the taxonomist, is to represent relationships in some coherent taxonomic system. We have engaged in the following studies relating to this problem.

9. The use of stereograms in numerical taxonomy.

Rohlf (1968) described a technique for summarizing the results of a numerical taxonomic study by a principal components analysis (with projection of the OTU's onto the first three principal components). The resulting three-way scatter diagram can be constructed as a solid model, as perspective drawings, or as stereo pairs. Computer programs were written to compute the coordinates of the OTU's so that a perspective stereo drawing could be made. Programs were also written to prepare tapes so that the stereograms could be automatically drawn by a Benson-Learner incremental plotter. The data for the *Hoplitis* complex (Michener and Sokal, 1957) were used as an illustrative example in this paper. The principal advantage of these techniques over conventional cluster analyses is that they enable one to ascertain visually which OTU's are intermediate between clusters, and to distinguish these from OTU's distant from all clusters in the study. It is sometimes difficult to distinguish these two types of OTU's when only a phenogram is available. This method also enables one to observe relations in hyperspace among clusters (phenograms contain no information about this).

10. The effect of correlated characters on numerical taxonomic data.

Rohlf (1967) studied the effect of using a set of rather highly intercorrelated characters in a numerical taxonomic study (74 external morphological characters of 45 species of North American mosquito pupae). It was found that the main effect of the character correlations was to cause the clusters to be elongated into hyper-ellipsoids. Standard methods of cluster analysis have a tendency to make the OTU's at the ends of these clusters appear to be much more isolated than they perhaps "should be". Methods for correcting for this bias were discussed, and the use of multidimensional scaling techniques such as principal components analysis were suggested for such data. As is typical of data with a large general factor (due to the overall correlation of characters), cluster analysis of the correlation rather than the distance matrix gave a much clearer result.

11. Analysis of hyper-ellipsoidal clusters.

Rohlf (1965, 1967, and 1968) has been interested in the effects of different types of data structures on the results of cluster analyses. The majority of present-day cluster analyses work most efficiently on clusters which are hyperspheroidal in shape, i.e., OTU's clustered around a point, the centroid of the cluster. But other shapes of clusters are possible and have been found to exist in taxonomic data. These are typically handled very poorly by conventional cluster analysis. Rohlf (1966, unpublished paper presented at the Brookhaven meetings of the Biometric Society; 1969 in preparation) has developed a clustering algorithm which allows for a wide variety of basic cluster shapes. The clustering

algorithm is basically a sequential agglomerative clustering procedure. A modified generalized distance function is used to compute the distance between a cluster and the OTU's which are candidates for addition to the cluster in the next clustering cycle. The use of this distance function rather than an ordinary average distance or a centroid distance function allows the clusters to become elongated into hyper-ellipsoids, since the generalized distance function considers distances in certain directions to be more important than others. OTU's at the ends of an ellipsoidal cluster are considered to be closer than OTU's around the side of the cluster. The generalized variance of a cluster (its P-volume) is used as an index of homogeneity of OTU's within a cluster. By a simple technique of adding dummy variables to the original variables included in the study one can compute distances in a way such that OTU's which are consistent with the trend of the shape of the cluster are considered close to the cluster, and OTU's inconsistent with the trend are considered further away. For example, in a ring-shaped cluster the addition of the dummy variables X^2 and Y^2 will cause the generalized distance function to represent distances of OTU's from the ring rather than the centroid of the cluster. Consequently OTU's near the centroid would also be considered farther away from the cluster than OTU's around the outer edge of the ring. The rate at which distances increase as one moves away from the cluster is a function of the closeness with which the OTU's in the cluster follow a definite curvilinear relationship. Clusters of any shape may be handled by this technique if the appropriate dummy variables are added to the matrix. As a by-product of the computations, the cluster analysis indicates the apparent shape of the cluster at each step in the clustering so that one is aware of the fact that, for example, a parabolic shaped cluster has been isolated. The principal limitation of this method is that one must decide a priori what class of dummy variables should be added to the matrix, and this in turn determines what classes of cluster shapes can be recognized by the program.

12. CAROD cluster analysis.

Drs. Rohlf and F. Möller (Istituto di Statistica, Università Bocconi, Milano, Italy) have worked on developing a form of cluster analysis which has the following properties: 1. It is based on the notion of the information content of a character. Characters are evaluated and used on the amount of information they contain about other characters in the study. 2. The method clusters by subdividing OTU's into groups (rather than by adding OTU's onto previously found clusters), i.e., it is a divisive type of clustering. It was hoped by this approach to give greater reliability to the important higher level clusters. 3. It attempts to reduce the effect of errors in sampling of OTU's, i.e., to allow for the fact that one may have selected a large number of OTU's from one genus, and a few from another. The procedure also allows for the simultaneous construction of an identification key for the clusters it finds. The programming of this method has been completed, but

it was found that it did not give satisfactory results when used in actual practice. The principal difficulty seems to be the fact that it uses a divisive clustering algorithm, and the fact that the clustering program divides OTU's into groups on the basis of a monothetic criterion. The principal result of the consultation with Dr. Möller during his stay in Lawrence in the summer of 1966 was to abandon this type of clustering.

D. THE EFFECT OF CONVERGENCE ON PHENETIC RELATIONSHIPS

13. A study of rodents.

In this study by A. J. Boyce of Oxford University, a research assistant with us for one year, the behavior of numerical taxonomic methods were investigated when applied to forms some of which are known to be convergent.

A jumping mode of progression has supposedly arisen independently in several rodent groups. Usually such groups also contain members which do not show this specialization and forms were chosen so that both types were represented. The material was taken mainly from the North American families which contain the kangaroo rats and mice, the spiny and pocket mice, the gophers and the jumping rats and mice. Representatives of the Old-World gerbils and spring-hares and the generalized murine forms, the wood and black rat, were also studied. Each of 17 genera was represented by a number of male and female specimens (64 in all). For each OTU 244 linear skeletal measurements were taken (122 cranial, and 122 post-cranial). Various similarity coefficients were computed. No coefficient produced a pattern of relationships in complete agreement with the existing taxonomic arrangement. The correlation matrix indicated relationship in functional adaptation and the distance matrix seemed indicative of overall size. The effects of size may be reduced by ignoring the first principal component of the variance-covariance matrix. The pattern of relationships with respect to the second and third principal components is in fairly good agreement with that produced by a cluster analysis of correlations between the forms.

The distribution of the families in the space defined by the second and third latent vectors clearly shows the differences in diversity within and between the families. Similarities between the members of different families adapted to a jumping mode of progression are not sufficient to produce any overlap of the families in this space.

The study was incorporated as part of Mr. Boyce's thesis for the doctorate at Oxford University (Boyce, 1965). It is to be published soon.

E. THE NONSPECIFICITY HYPOTHESIS

There is continuing interest in measuring the amount of congruence between classifications based on different suites of characters.

14. A new test of the hypothesis of non-specificity on the Hoplitis complex.

This is a study of the degree to which phenetic similarities among species of bees of the Hoplitis complex, are the same in the two sexes and in different body regions. A study was carried out on 97 species, partitioning the 122 recorded characters into 69 found in females and in both sexes versus 53 found only in males. In a companion study of the same species using the same 122 characters, 60 head characters were compared with 62 nonhead characters. Phenograms were computed showing the classifications of these bees based on all characters and on the subsets of characters described above. The cophenetic correlation based on correlation matrices are 0.71 for sexes and 0.61 for head versus nonhead characters. Our study discusses the partial failure of the nonspecificity hypothesis on the basis of the present findings and compares these with the results of other recent authors. This study has been published by Michener and Sokal (1966).

F. NUMERICAL CLADISTICS

There have been several recent attempts to develop methods for estimating cladistic relationships (branching sequences of evolutionary lineages). Our work at The University of Kansas on this subject originated from the investigation on the principles of taxonomic methodology reported in Section B.

15. The original study.

This study was based on the "Caminalcules", a group of imaginary animals generated by J. H. Camin according to principles intentionally known so far only to him but which he believes to be consistent with what is generally known of transspecific evolution. Genetic continuity was established by tracing the drawings of the animals from sheet to sheet, permitting the preservation of all characters except for such modifications as were desired.

This study has led to an empirical method based on the principle of parsimony of evolutionary steps which appears capable of estimating probable cladistics from the characters of existing organisms. This method has been successfully applied to known phylogenies of hypothetical animals and to assumed phylogenies of well known groups of organisms such as fossil horses. A first publication on this work has appeared (Camin and Sokal, 1965).

Several studies on other organisms have been completed and are awaiting publication. These include the *Fusulinidae* (with R. Kaesler), the *Mecoptera* (by G. W. Byers), the *Hoplitis* complex of bees (R. R. Sokal and C. D. Michener), the vipers (H. Marx and J. Felsenstein), the genus *Arctostaphylos* in plants (P. V. Wells), chromosomal inversions in blackflies (R. Hansell and D. Chant), and tiger beetles (H. Willis). A computer program has been prepared by Mr. Ron Bartscher to carry out the procedures of numerical cladistics and present the results in easily intelligible fashion. This program is generally available (see Section G).

16. Some recent improvements.

In a doctoral dissertation, Hendrickson (1967) examined the pattern cladogram as defined by Camin and Sokal and found that certain anomalies could be traced to the presence of an empty internode in some places in this pattern cladogram. The pattern cladogram was therefore redefined by him as "the included subset pattern cladogram". A procedure for determining the maximum number of extra steps a given character can take in being fitted to another pattern is given, and the possible use of this information in a weighting function has been discussed. Since the compatibility matrix constructed from the redefined pattern cladogram is symmetrical with respect to compatibilities, it is possible to make a number of inferences about the structure of the cladogram from this compatibility matrix. This led to the development of the restricted extra step partitioning procedure.

The character space is presented in terms of graph theory, and the question of parsimony is examined as a problem in this field. The most parsimonious cladogram is a tree of minimum total length, constrained to a directed city-block lattice. This is discussed in detail in Hendrickson (1968).

Some suggestions have been made for the use of techniques of "data analysis" in locating cases in which certain of the assumptions of character coding should be relaxed. In particular, cases of probable convergence and reversals have been considered.

Several sets of data have been analyzed using these methods, and these have pointed out some of the shortcomings of this method.

In general, the conclusions that can be reached are that many more characters need to be considered than are normally considered for cladistic analyses, and that in view of this conclusion, it is paradoxically more likely that satisfactory cladistic reconstructions can be made on the basis of studies of recent organisms than from fossil organisms.

Since the reconstruction of the cladistic sequence is subject to considerable bias and possible distortion in most sets of data, the use of cladistics as a basis for classification seems unwarranted. The most effective use of numerical cladistics seems to be as a test of the internal consistency of evolutionary speculation.

This work is being prepared for publication.

G. THE DEVELOPMENT OF COMPUTER PROGRAMS FOR NUMERICAL TAXONOMY

Since the inception of the project, numerous computer programs for numerical taxonomy have been prepared. These programs have been organized into two comprehensive systems of programs. The NT-SYS programs (developed by Dr. F. J. Rohlf in collaboration with Mr. John Kishpaugh, Mr. Ron Bartcher, and Mr. Jerry Crow) are for numerical phenetics. The CLADON programs perform computations for numerical cladistics (developed by Mr. Ron Bartcher).

17. The NT-SYS programs.

These have been organized into an integrated system of programs with provisions for simplified free format parameter cards which call up the various subprograms and specify the data to be acted on. Provisions are also made for storage of data matrices (and matrices which result from the computations) on a master "stack" tape along with symbolic identification codes. The present version (October 1968) of NT-SYS contains over thirty basic entry points which allow the calling up of programs which constitute over 70 links.

This package of programs (prepared under the current grant) is one of the most sophisticated multivariate analysis packages available today. It was used extensively on the IBM 7040 and the present GE 625 computing system at The University of Kansas. Requests for this program have come from numerous institutions in this country and overseas. At present plans are being made to publish the system in the Computer Contributions of the Kansas State Geological Survey. Work has just started on developing a version of NT-SYS for the IBM 360/65 computing system at the State University of New York, Stony Brook.

The following is a brief listing of the types of computations allowed for within NT-SYS.

a. Matrix algebra. Input to this program consists of statements such as $X = (A + B) * C^{-1}$ which would result in matrices A and B being added together, their sum being multiplied by matrix C, and the inverse of the result stored as matrix X.

b. The coefficient of association program allows the computation of various association coefficients.

c. The CORDST program is used for the computation of correlation, distance, variance - covariance, and cosine θ data matrices.

d. An EDIT program is included which permits manipulation of the master stack tape containing the matrices.

e. A general FACTOR extraction program is included with provisions for estimation and iteration of communality estimates in factor analysis. Options are provided for using Thurstone's complete centroid method of factor extraction, the power method, the Jacobi method, and the Wilkinson-Householder methods for computation of eigenvalues and eigenvectors. Plans are being made for the inclusion of options for alpha, image, canonical, and maximum likelihood factor analyses.

f. The GTAXON program is a generalized sequential agglomerative cluster analysis program which takes into consideration within-group variation of OTU's. It also allows for elongated and curvilinear clusters.

g. The INPUT program is used for the inputting of matrices and storing the information on the master stack tape.

h. The MAXTAX program performs iterative manipulation of a phenogram to attempt to maximize the cophenetic correlation.

i. The MIAM program performs an oblique rotation of a factor matrix to simple structure using Sokal's mass modification of Thurstone's analytical method for simple structure.

j. The MIXCNG program allows for the deletion and reordering of rows and columns of matrices.

k. The MULDIS program is a multi-group generalized discriminant function analysis program.

l. The MXCOMP program performs a comparison of two symmetric matrices for congruence and computes the cophenetic correlation.

m. The MXPLOT program is used to plot rows or columns of a matrix against one another on a two-way scatter diagram on the line printer.

n. The OUTPUT program is used for the printing and punching out matrices.

o. The PHNGN1 and PHNGN2 programs are special phenogram output programs.

p. The PROJCT program performs a projection of OTU's onto a given set of axes.

q. The SORTIT program prints a sorted listing of correlation or distance values for each variable in a correlation or a distance matrix.

r. The STAND program standardizes the characters of a data matrix.

s. The TAXON cluster analysis program has options for single and complete linkage, weighted and unweighted average linkage analysis, as well as Spearman's sums of variables method and centroid clustering.

t. The TEXT program is a general text editing routine which is used for formatting and outputting writeups of the NT-SYS system.

u. The TRANFM program performs various transformations of elements of matrices. Transformations such as arc sin, arc cos, Fisher's z , log, $\log(Y + \frac{1}{2})$, and others are allowed for.

v. The VARMAX performs the varimax orthogonal rotation of a factor matrix to simple structure.

w. The WILKIN program computes the Wilk's statistic.

x. The GOWER program is used to transform association and distance matrices in a way such that the eigenvectors computed from the transformed matrix corresponds to the projections of the OTU's onto the principal components which would result if one were to analyze the matrix of correlations among the characters. This program is used whenever there are many fewer OTU's than characters.

A number of other programs are currently in preparation.

18. The CLADON programs.

The CLADON programs for numerical cladistics consist of three separate programs for computing a compatibility matrix, obtaining an initial solution by the monothetic method, and an iterative method for finding the minimum length tree. The cladogram is close to the form in which taxonomists are used to looking at data of this sort, i.e., cladograms with characters marked across branches of a cladogram, and there is a total count of the number of required evolutionary steps at the bottom of a tree. This program has also been requested by a number of institutions. Among them are the University of Chicago, The Chicago Natural History Museum, Kansas State University, The University of Michigan, etc. It has been published by its developer (Bartcher, 1966).

H. NUMERICAL TAXONOMIC STUDIES OF VARIOUS TAXONOMIC GROUPS

19. A numerical classification of the orders of insects.

This research project was started from "seeding funds" by the General Research Fund of The University of Kansas. It has been financed from the current grant for two years. The aim is to see whether methods for numerical taxonomy are applicable to higher categories (using the exemplar method) and what can be learned about the methodology in such a case. A second aim is to see what light methods of numerical taxonomy can throw upon the phenetic similarities among representatives of most of the orders of insects. Thirdly we wish to apply the methods of numerical cladistics to this study as well. The study is based on 39 species representing the major insect groups. These data have been processed through the computer, and are awaiting analysis and preparation of a manuscript.

20. A numerical taxonomic study of the mite superfamily Celaenopsoidea.

A Ph.D. dissertation by Richard C. Funk, entitled "Revision of the family Euzerconidae and its relationships within the superfamily Celaenopsoidea (Acarina: Mesostigmata), aided by the techniques of numerical taxonomy," has been completed. It is a numerical taxonomic analysis of a relatively unknown group of mites, including 30 species of Euzerconidae and 16 additional species of non-euzerconid mites in the superfamily Celaenopsoidea. The study was conducted in two parts. The first part was a study of adult female specimens of 25 species of Euzerconidae in which individual specimens served as OTU's. The study was based on 76 OTU's and 116 characters. The second part was an exemplar study, using both males and females, of 46 species of Celaenopsoidea, the largest superfamily of Trignyaspid Mesostigmata, and included the Euzerconidae. This study was based on 179 characters, 80 of which had been utilized in the first study. In addition to yielding a useful classification of the family Euzerconidae and a tentative classification of the entire superfamily Celaenopsoidea, the study resulted in the synonymy of two and the reclassification of four previously described species, plus the description of 20 new species and seven new genera of Euzerconidae, four new species, three new genera and two new families of non-euzerconid Celaenopsoidea, and the raising of one previously described subfamily of Diplogyninae to family rank. The thesis (Funk, 1968) is being prepared for publication.

21. A numerical taxonomic analysis of the chigger mite genus Buschoengastia Ewing (Acarina: Trombiculidae).

This study, not anticipated in the original application, has been undertaken by William J. Wrenn. The classical taxonomy of chiggers is based almost exclusively on the characters of the parasitic larval instar and in most cases it has been possible

to match the predaceous post-larval instars with the larvae only by the tedious process of laboratory rearing of these mites through the entire life cycle. This study includes the larvae of 54 taxa and the reared and correlated nymphs of 30 of these. Data on 360 characters have been tabulated for most of the larvae and 238 nymphal characters have been selected for analysis. In addition to analyzing the taxonomy of the genus Euschoengastia by the techniques of numerical taxonomy, we intend to assess the value of numerical taxonomy in matching the post-larval instars of chiggers with their respective larval instars and the possible application in matching the various stages of other organisms with complex life histories, such as parasitic trematodes, hydroscoan coelenterates, and holometabolous insects.

22. Numerical classification of mosquito pupae.

Rohlf (1967) reported on an analysis of 45 species of North American mosquito pupae using 74 external morphological characters. The resulting classifications agreed surprisingly well with the standard classification of the family considering the fact that the pupal stage is not typically taken into account when determining distinctions between genera and subfamilies. In addition, the characters used in this study were not ones ordinarily considered reliable by conventional taxonomists.

I. STATISTICAL METHODS IN SYSTEMATICS

23. A review.

Another activity under this project not anticipated at the time of the original grant application, concerned an extensive and intensive review of the application of statistical methods to systematics in general, including not only numerical taxonomy but the far wider application of statistics to the entire field of systematics. The occasion for such a review arose when R. R. Sokal spent a sabbatical year abroad as a Fulbright professor during 1963-64. This review (Sokal, 1965) is the only such review in the literature and should be of considerable value to diverse systematists contemplating the application of statistical methods to their problems. International interest in this paper, the publication of which was supported by the present grant, has been considerable. A supply of over 500 copies of the reprint was exhausted in a few months.

J. THE PARALLEL RESEARCH IN NUMERICAL TAXONOMY CARRIED OUT UNDER THE SPONSORSHIP OF THE NATIONAL INSTITUTES OF HEALTH

There is close interaction between the activities in numerical taxonomy on the two projects. The development of computer programs for instance is sponsored and partially funded by NSF and by NIH. However, in terms of actual research projects, the work supported by NSF has been

kept separate from that supported by NIH. The latter projects can be divided into the following four general activities: (1) the application of numerical taxonomy to new taxonomic groups, (2) work on the principles and theory of taxonomy, (3) new approaches in numerical taxonomy and the application of numerical taxonomy to new fields, (4) experiments with the techniques of numerical taxonomy.

(1) Among the applications of the techniques to new groups is a study of the mosquito genus Psorophora undertaken and published by Hendrickson and Sokal (1967). In this study the phenetic relationships among 29 species of culicine mosquitoes were based on 158 adult characters. The species included 15 in the genus Psorophora and (for purposes of comparison) 14 in Aedes. The findings were reported as phenograms and as graphs representing the position of the species in a three-dimensional principal components space. Although the various subgenera appear distinct, there seems little justification for uniting three of the established subgenera in the genus Psorophora as distinct from Aedes. From the analysis it is obvious that the classification of these 29 species in the conventional, hierarchic, mutually exclusive manner is unsatisfactory because it does not represent the phenetic relationships revealed by this study.

A second study just completed by Huber (1968) is a revision of 30 genera in 10 families of cockroaches (Blattariae). The purpose of this study (which included first instar and last instar nymphs as well as adults) was to study the changes in phenetic relationships within this group during ontogeny. It was found that classifications based on last instar nymphs are much closer to those of adults than of first instar nymphs. Huber's study included a very thorough reanalysis of a recent revision of the roaches by McKittrick and is perhaps the first extensive reanalysis of a conventional taxonomic study by numerical taxonomy.

(2) Important recent theoretical developments have involved a re-examination of the principles of classification. This was prompted by the development (at Kansas and elsewhere) of numerical methods for analysis of phylogenetic sequences (cladistic analysis). In an extensive review of the principles of classification, Sokal and Camin (1965) examined and contrasted conventional and operational taxonomy and subdivided relationships into those of phenetics and cladistics. They demonstrated that phenetic methods are to be preferred for general classificatory work but that phenetics and cladistics are necessary for a complete understanding of the systematics of a group of organisms. In another paper Sokal et al. (1965) replied in detail to criticism of the phenetic method in numerical taxonomy by Ernst Mayr. The advent of numerical taxonomy has had considerable repercussions on general taxonomic practices. For museum procedures considerable automation has been suggested and numerous projects in this regard are underway. A detailed discussion of efficiency in taxonomy and the application of computers to taxonomic work was prepared by Sokal and Sneath (1966) under this grant. A special analysis of the problems of using electronic

data processing techniques in biological collections was presented by Crovello (1967). Among the projects still underway is that of the "intelligent ignoramus" in which two untrained persons were asked to reclassify a group of specimens of bees that had been previously classified by orthodox taxonomy by C. D. Michener and by numerical taxonomy by Michener and Sokal. The purpose of this experiment was to investigate whether untrained personnel could define characters sufficiently accurately and rapidly so that numerical taxonomy could proceed and establish useful classifications from such data. The results were generally satisfactory, the error between the two "ignoramuses" and between Michener's classification being no greater than that between numerical taxonomic classifications based on different sets of characters. This study by Sokal, Rohlf and Michener has been completed and will be written up and published in the near future.

(3) An application of certain techniques of numerical taxonomy and of trend surface analysis to the classification of the terrestrial vertebrate fauna of Kansas was undertaken by Fisher (1968). By principal component analysis five factors could be extracted from a matrix of correlations among the localities based on the species. Factor scores were mapped by trend surface analysis. These were shown to be of interest where they yielded overlapping faunal zones.

By applying numerical taxonomy to thirteen biotic parameters of eight weevil strains, Fujii (1969) in our laboratory was able to show that taxonomic units clustered on the basis of these characters yielded a satisfactory and consistent classification of these strains and that one could use numerical taxonomy in this way to define similarities between ecological properties of these strains. These properties define imperfect images of ecological niches, whose similarities can then be measured. This application of numerical taxonomy to ecology has been extended and applied to a series of protozoan populations collected in fresh waters by Cairns. Study and analysis of these data by Sokal, Cairns, and Fujii are underway.

Among the new techniques employing numerical taxonomy was a first attempt at simulating the scanning of biological images randomly as well as systematically by means of perforated punch cards superimposed on the images. These studies by Sokal and Rohlf (1966) and Rohlf and Sokal (1968) were applied to 29 species of hypothetical animals (the Caminalcules) and pupae of 32 species of culicine mosquitoes. Black lines appearing through a hole were scored "1" and empty holes were scored "0". Similarity matrices were constructed on the basis of matching scores for corresponding masks and holes. The resulting phenograms agreed well with previous numerical taxonomic studies obtained by deliberately looking for and scoring conventional characters. Complete scanning of the entire image also gave satisfactory results but classifications were not necessarily better than by the random scanning methods. Increasing the size of the apertures was also attempted. It slightly degraded the classifications obtained from the finer scan.

(4) Among studies investigating the techniques of numerical taxonomy, Crovello (1968) examined the effects of standardization versus condensation and of using only an estimate of central tendency as well as that plus an estimate of variation in a basic data matrix. Changes in these techniques appear to result in different relationships. In another analysis of character correlations of the Hoplitis complex of bees first worked on by Michener and Sokal (1957), Rohlf (1968) found that when a three-dimensional model was constructed of these groups the operational taxonomic units were clustered into five clusters corresponding well with the previous classification of the group both by numerical taxonomy and conventional taxonomy. The analysis of these correlations is essentially completed and should be written up in the near future. This is a joint study by Rohlf, Michener, and Sokal.

A review of various methods of cluster analyses was undertaken at the invitation of Drs. Maurice Lorr and Samuel B. Lyerly and published as Sokal (1967). It permitted the interchange of considerable information between social scientists and biological numerical taxonomists on common problems and approaches.

(4) Bibliography of publications resulting from this project since its inception.

1. Bartcher, R. L. 1966. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040. Computer Contr. 6. State Geological Survey, The University of Kansas. 54 pp.
2. Basford, N. L., J. E. Butler, C. A. Leone and F. J. Rohlf. 1968. Immunologic comparisons of selected coleoptera with analysis of relationships using numerical taxonomic methods. Systematic Zool. 17: (in press).
3. Boyce, A. J. 1965. The methods of quantitative taxonomy with special reference to functional analysis. Ph.D. Thesis. Oxford University. 190 pp.
4. Carin, J. H. and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. Evolution 19: 311-326.
5. Fisher, D. R. and F. J. Rohlf. 1969. Robustness of numerical taxonomic methods to errors in homology. Systematic Zool. 18: 33-36.
6. Funk, R. C. 1968. Revision of the family Euzerconidae and its relationships within the superfamily Celaenopsoidea (Acarina: Mesostigmata) aided by the techniques of numerical taxonomy. Ph.D. Thesis. The University of Kansas. 258 pp. + illustrations.
7. Hendrickson, J. A., Jr. 1967. A methodological analysis of numerical cladistics. Ph.D. Thesis. The University of Kansas. 88 pp. + appendices.
- + 8. Hendrickson, J. A., Jr. 1968. Clustering in numerical cladistics: a minimum-length directed tree problem. Mathematical Biosciences (in press).

9. Michener, C. D. 1963. Some future developments in taxonomy. *Systematic Zool.* 12: 151-172.
 10. Michener, C. D. 1964. The possible use of uninomial nomenclature to increase the stability of names in biology. *Systematic Zool.* 13: 182-190.
 11. Michener, C. D. and R. R. Sokal. 1966. Two tests of the hypothesis of nonspecificity in the Hoplitis complex (Hymenoptera: Megachilidae).
 12. Rohlf, F. J. 1967. Correlated characters in numerical taxonomy. *Systematic Zool.* 16: 109-126.
 13. Rohlf, F. J. 1968. Stereograms in numerical taxonomy. *Systematic Zool.* 17: 246-255.
 14. Rohlf, F. J. 1969. Adaptive hierarchical clustering schemes. *Systematic Zool.* (in press).
 15. Rohlf, F. J. and D. R. Fisher. 1968. Tests for hierarchical structure in random data sets. *Systematic Zool.* 17: 407-412.
 16. Sokal, R. R. 1963. The principles and practice of numerical taxonomy. *Taxon* 12: 190-199.
 17. Sokal, R. R. 1965. Statistical methods in systematics. *Biol. Rev. (Cambridge)* 40: 337-391.
 18. Sokal, R. R. 1967. Recent ideas on the theory of systematics [in Russian]. *Zhurnal Obshchye Biologii* 28: 658-674.
 19. Sokal, R. R. 1968. Numerical taxonomy: methodology and recent developments [in Russian]. *Zhurnal Obshchye Biologii* 29: 297-315.
- Biography of publications in numerical taxonomy supported by the parallel NIH grant.
- Crovello, T. J. 1967. Problems in the use of electronic data processing in biological collections. *Taxon* 16: 481-494.
- Crovello, T. J. 1968. The effect of alteration of technique at two stages in a numerical taxonomic study. *Univ. Kans. Sci. Bull.* 47: 761-786.
- Fisher, D. R. 1968. A study of faunal resemblance using numerical taxonomy and factor analysis. *Syst. Zool.* 17: 48-63.
- Fujii, K. 1969. Numerical taxonomy of ecological characteristics and the niche concept. *Systematic Zool.* (in press).
- Hendrickson, J. A., Jr. and R. R. Sokal. 1967. A numerical taxonomic study of the genus Psorophora (Diptera: Culicidae). *Ann. Entomol. Soc. Amer.*, 61: 385-392.
- Rohlf, F. J. and R. R. Sokal. 1967. Taxonomic structure from randomly and systematically scanned biological images. *Systematic Zool.* 16: 246-260.

- Sokal, R. R. 1967. A review of clustering methods in biological taxonomy. In Lorr, M. A. (ed.), *Proceeding Conference on Cluster Analysis of Multivariate Data*. Bur. Soc. Sci. Res., Inc. pp. 3.01-3.10.
- Sokal, R. R. and J. H. Camin. 1965. The two taxonomies: Areas of agreement and conflict. *Systematic Zool.* 14: 176-195.
- Sokal R. R., J. H. Camin, F. J. Rohlf and P. H. A. Sneath. 1965. Numerical taxonomy: some points of view. *Systematic Zool.* 14: 237-243.
- Sokal, R. R. and F. J. Rohlf. 1966. Random scanning of taxonomic characters. *Nature*, 210: 461-462.
- Sokal, R. R. and P. H. A. Sneath. 1966. Efficiency in taxonomy. *Taxon* 15: 1-21.

C. SIGNIFICANCE OF THIS RESEARCH

One of the dramatic developments in biology in this decade has been the rapid emergence of the newly integrated field of population biology. Just as advances in microbiology are largely of a chemical nature, so in the field of population biology many of the new concepts are formulated mathematically. It is also characteristic that the so-called "new mathematics" is largely involved in this work. The application of various methods of biostatistics, cybernetics, information theory, linear programming, the theory of games, and other methods, many of them intimately tied up with the development of computers, is currently revolutionizing many of our previously held views on the nature of the interactions among organisms in population dynamics and evolution. Closely connected with this development are the technical advances being made in automatic data recording, processing, and storing, so that procedures and conventions in fields, such as taxonomy, which date back to the scientific capabilities of the 19th century are becoming obsolete. A new taxonomy will play an important role in the new population biology which is currently being founded and for this reason alone the research above is of importance. However, it is self-evident that biologists of all kinds will continue to need a systematic foundation on which to base their researches. Modern population biology needs a taxonomy which will fit in with its general concepts; and this present-day taxonomy does not do. The current re-evaluation of taxonomy is valuable also in subjecting present-day systematic theory to a thorough-going review. Even if the answers proposed by numerical taxonomists are not always the correct ones, the challenge they have put forward to the established views has caused orthodox taxonomists to re-examine their own principles and this should lead to an improvement in systematics.

Since the original application for this grant was written in 1961, numerical taxonomy has progressed from a programmatic proposal for taxonomy to a widely accepted and practiced aspect of systematic work, whose impact is being felt in ever increasing measure by systematists in all fields. Admittedly, some of the early hopes for numerical taxonomy as the infallible panacea for taxonomic work have not been

realized. New problems have been uncovered, leading in time to new theoretical formulations and to new methodologies. But to an ever increasing extent the measurement of phenetic similarity and diversity has become a valuable approach in evolutionary studies, and numerical cladistic approaches promise to follow suit. The successes of numerical taxonomy in ordinary taxonomic work such as monographs can be easily documented and the literature has become so voluminous that Sokal and Sneath, authors of the "Principles", have had great difficulty in keeping up with it for the second edition of their book.

Beyond biological systematics in the strict sense, numerical taxonomy has contributed to the methodology of numerous fields: ecology - classification of habitats, leading to an understanding of common factors determining the composition of plant and animal communities; medicine - disease classification, providing objective measures for common symptomatology and leading to efficient keys for automatic diagnosis; psychology - classification of personality types and mental disorders; pedology - soil classifications; and applications to machine parts, geological strata, legislatures, political systems, archeological artifacts and many more.

(5) Facilities.

For numerical taxonomic research, general office and laboratory space is available in the Biological Sciences Department. A special laboratory (about 700 square feet) is being set up to house desk calculators, remote computer terminals, IBM card punch and IBM card file cabinets.

The facilities of the University's computation center are available and essential for this project. The computer is an IBM360/67 (but running as a 65), with 512 K core memory, 6 tape drives, an IBM 2314 disc, and 2 IBM 1100 lpm printers. From our previous experience we expect these facilities to be adequate for our computing needs for the next several years. The center has plans for short and long range improvements. Sometime in 1970 it intends to have 30 IBM 2740 remote terminals. Of these at least one will be available to this project to be used for the development of computer programs for numerical taxonomy and for the calling up of programs for executing sets of data.

A Culler-Fried remote terminal useful for graphic display of mathematical functions will also be available.

(6) Personnel.

Co-Principal Investigator: F. James Rohlf

1. Academic Biography: Born Blythe, California, 1936.
A. B. (Zoology) San Diego State College, 1958.
Ph. D. (Entomology) The University of Kansas, 1962.

2. Positions.

Research Associate in Entomology, The University of Kansas,
summers 1962, 1965.
Assistant Professor of Biology, University of California, Santa Barbara,
1962-1966.
Visiting Assistant Professor of Entomology, The University of Kansas,
1965.
Associate Professor of Statistical Biology, The University of Kansas,
1966-1969.
Associate Professor of Biological Sciences, State University of
New York at Stony Brook, 1968-

3. Publications.

F. James Rohlf has published 30 papers in several fields of biological research, 18 of which are relevant to the present project and are listed in the joint bibliography below. A textbook on biometry and a set of statistical tables, in co-authorship with Robert R. Sokal, will be published in the spring of 1969.

Co-Principal Investigator: Robert R. Sokal

1. Academic Biography: Born Vienna, Austria, 1926 (U.S. Citizen).
B. S. (Biology) St. John's University, Shanghai, China, 1947.
Ph. D. (Zoology) University of Chicago, 1952.

2. Positions.

Assistant in Zoology, University of Chicago, 1950-1951.
Instructor in Entomology, The University of Kansas, 1951-1953.
Assistant Professor of Entomology, The University of Kansas,
1953-1958.
Associate Professor of Entomology, The University of Kansas,
1958-1961.
Professor of Statistical Biology, The University of Kansas, 1961-1969.
Professor of Biological Sciences, State University of New York at
Stony Brook, 1968-
Senior Postdoctoral Fellow, NSF, University College, London, 1959-1960.
Fulbright Visiting Professor, Hebrew and Tel Aviv Universities, Israel,
1963-1964.
NIH Career Investigator, 1964-1969.

3. Publications.

Robert R. Sokal has published 64 papers in several fields of biological research, as well as one book on the principles of numerical taxonomy. Twenty-eight of these publications are directly relevant to the present project and are listed in the joint bibliography below. A textbook of biometry and a set of statistical tables in co-authorship with F. James Rohlf will be published in the spring of 1969. A revision of "Principles of Numerical Taxonomy" with P. H. A. Sneath is well underway.

Chief Programmer: John R. L. Kishpaugh

1. Academic Biography: Born Independence, Kansas, 1942.
B. S. (Electrical Engineering) University of Kansas, 1965.
Graduate work in Population Biology, University of Kansas, 1966-

2. Positions.

Systems and Applications Programmer, Department of Entomology, The University of Kansas, 1966-

Systems Programmer: David Kirk.

1. Academic Biography: Born Pittsburg, Kansas, 1947.
University of Kansas, 1965-1968.
Electrical Engineering

2. Positions.

Programmer, Kansas Highway Commission, 1967.

Research Assistant, CRES, 1967.

Programmer, Kansas Water Resources Board, 1968.

Consultant, CRES, 1968.

Systems Programmer, Department of Entomology, University of Kansas, 1969-

Joint Bibliography of the Co-Principal Investigators Pertinent to this Application.

Basford, N. L., J. E. Butler, C. A. Leone and F. J. Rohlf. 1968. Immunologic comparisons of selected coleoptera with analysis of relationships using numerical taxonomic methods. *Systematic Zool.* 17:388-406.

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

Fisher, D. R. and F. J. Rohlf. 1969. Robustness of numerical taxonomic methods to errors in homology. *Systematic Zool.* 18:33-36.

Michener, C. D. and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11:130-160.

Michener, C. D. and R. R. Sokal. 1966. Two tests of the hypothesis of nonspecificity in the *Hoplitis* complex (Hymenoptera: Megachilidae). *Ann. Entomol. Soc. Amer.* 59:1211-1217.

Rohlf, F. J. 1961. Numerical taxonomic study of the genus *Aedes* of America north of Mexico (Abstract). *Proc. N. Cent. Branch E.S.A.* 16:12-13.

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

- Rohlf, F. J. 1963. Classification of *Aedes* by numerical taxonomic methods (Diptera: Culicidae). *Ann. Entom. Soc. Amer.* 56: 798-804.
- Rohlf, F. J. 1964. Methods for checking the results of numerical taxonomic studies. *Systematic Zool.* 13: 102-104.
- Rohlf, F. J. 1965. Character Correlations in numerical taxonomy. *Proc. 12th Inter. Cong. Ent.* pp. 109-110.
- Rohlf, F. J. 1965. Multivariate methods in taxonomy. In *Proceedings of the IBM Scientific Computing Symposium on Statistics*, October 21-23, 1963. IBM, White Plains, New York. pp. 3-14.
- Rohlf, F. J. 1965. A randomization test of the nonspecificity hypothesis in numerical taxonomy. *Taxon* 14: 262-267.
- Rohlf, F. J. 1967. Correlated characters in numerical taxonomy. *Systematic Zool.* 16: 109-126.
- Rohlf, F. J. 1968. Stereograms in numerical taxonomy. *Systematic Zool.* 17: 246-255.
- Rohlf, F. J. 1969. Adaptive hierarchical clustering schemes. *Systematic Zool.* (in press).
- Rohlf, F. J. and D. R. Fisher. 1968. Tests for hierarchical structure in random data sets. *Systematic Zool.* 17: 407-412.
- Rohlf, F. J. and R. R. Sokal. 1962. The description of taxonomic relationships by factor analysis. *Systematic Zool.* 11: 1-16.
- Rohlf, F. J. and R. R. Sokal. 1965. Coefficients of correlation and distance in numerical taxonomy. *Univ. Kans. Sci. Bull.* 45: 3-27.
- Rohlf, F. J. and R. R. Sokal. 1967. Taxonomic structure from randomly and systematically scanned biological images. *Systematic Zool.* 16: 246-260.
- Sneath, P. H. A. and R. R. Sokal. 1962. Numerical taxonomy. *Nature* 193: 855-860.
- Sokal, R. R. 1958. Quantification of systematic relationships and of phylogenetic trends. *Proc. 10th Inter. Cong. Entom.* 1: 409-415.
- Sokal, R. R. 1959. Comments on quantitative systematics. *Evolution* 11: 420-423.
- Sokal, R. R. 1959. Die Grundlagen der numerischen Taxonomie. *Proc. 11th Inter. Cong. Entom.* 1: 7-12.
- Sokal, R. R. 1961. Distance as a measure of taxonomic similarity. *Systematic Zool.* 10: 70-79.
- Sokal, R. R. 1962. Typology and empiricism in taxonomy. *J. Theor. Biol.* 3: 230-267.

- Sokal, R. R. 1963. The principles and practice of numerical taxonomy. *Taxon* 12: 190-199.
- Sokal, R. R. 1964. The future systematics. In C. A. Leone [ed.], *Taxonomic Biochemistry and Serology*, Ronald Press, New York. pp. 33-48.
- Sokal, R. R. 1964. Numerical taxonomy and disease classification. In J. A. Jacquoz [ed.], *The Diagnostic Process*. U. Mich. pp. 51-68.
- Sokal, R. R. 1965. Current progress in numerical taxonomy. *Proc. 12th Inter. Cong. Ent. Entomol.* p. 108.
- Sokal, R. R. 1965. Statistical methods in systematics. *Biol. Rev. (Cambridge)* 40: 337-391.
- Sokal, R. R. 1966. Numerical Taxonomy. *Scientific American* 215: 106-116.
- Sokal, R. R. 1967. A review of clustering methods in biological taxonomy. In M. Lehr and S. B. Iyerly [eds.], *Some Critical Issues and Problems in Cluster Analysis*. Bur. Soc. Sci. Res., Inc. pp. 3.01-3.10.
- Sokal, R. R. 1967. Recent ideas on the theory of systematics [in Russian]. *Zhurnal Obshchye Biologii* 28: 658-674.
- Sokal, R. R. 1968. Numerical taxonomy: methodology and recent developments [in Russian]. *Zhurnal Obshchye Biologii* 29: 297-315.
- Sokal, R. R. and J. H. Camin. 1965. The two taxonomies: areas of agreement and conflict. *Systematic Zool.* 14: 176-195.
- Sokal, R. R., J. H. Camin, F. J. Rohlf, and P. H. A. Sneath. 1965. Numerical taxonomy: some points of view. *Systematic Zool.* 14: 237-243.
- Sokal, R. R. and C. D. Michener. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* 38: 1409-1438.
- Sokal, R. R. and C. D. Michener. 1967. The effects of different numerical techniques on the phenetic classification of bees of the *Hoplitis* complex (Megachilidae). *Proc. Linn. Soc., Lond.* 178: 59-74.
- Sokal, R. R. and F. J. Rohlf. 1962. The comparison of dendrograms by objective methods. *Taxon* 11: 33-40.
- Sokal, R. R. and F. J. Rohlf. 1966. Random scanning of taxonomic characters. *Nature* 210: 461-462.
- Sokal, R. R. and P. H. A. Sneath. 1963. The principles of numerical taxonomy. W. H. Freeman, San Francisco. 359 pp.
- Sokal, R. R. and P. H. A. Sneath. 1966. Efficiency in taxonomy. *Taxon* 15: 1-24.

(7) Proposed Budget

1 July 1969 - 31 December 1969

Requested from NSF SUNY Contribution Total Cost

PERSONNEL

Co-principal investigators:

Robert R. Sokal, Professor

None

None

F. J. Rohlf, Associate Professor
(100% of time, 1 month summer salary,
1/9 academic year salary)

\$ 2,056

\$ 2,056

John L. Kishpaugh, Chief Programmer
(100% of time, 4 months)

3,167

3,167

David Kirk, Systems Programmer
(100% of time, 4 months)

2,400

2,400

Secretary
(100% of time, 4 months)

1,833

1,833

2 Graduate Research Assistants
(50% of academic year for 4 months)

2,448

2,448

SALARY SUBTOTAL

\$11,904

\$11,904

Fringe Benefits13.8% of salaries and wages for
Research Foundation Employees

1,305

1,305

PERSONNEL TOTAL

\$13,209

\$13,209

CONSUMABLE SUPPLIES (data sheets,
data processing binders, etc.)

500

500

PUBLICATION COSTS

1,250

1,250

TRAVEL to scientific meetings

350

350

COMPUTER CHARGES

2,500

2,500

SUBTOTAL DIRECT COSTS

\$17,809

\$17,809

INDIRECT COSTS (20% of total direct costs)

3,562

4,437*

7,999*

TOTAL BUDGET

\$21,371

\$4,437

\$25,808

* University contribution consists of difference between requested overhead of 20% of total direct cost and the A-21 audited overhead rate of 67.2% of salaries and wages.

(8) Explanation of budget.

The budget outline in the previous pages is mostly self-explanatory. Some comments on special aspects of the budget follow.

A summer salary is provided for one month for co-principal investigator F. J. Rohlf. R. R. Sokal will spend half time on this project during the summer but is paid from an NIH career development award. Both principal investigators will spend 15-20% of their time each month on this project during the rest of the academic year.

Our chief programmer, Mr. John R. L. Kishpaugh has been with our research projects for the past four years and has proven invaluable in program development and in applications work. It is our hope that his salary will eventually be paid from a continuation grant by NIH, which has already been scientifically approved but not yet funded. We had to include his salary in this budget to assure him a minimum of financial security to enable him to move to New York with us. If the NIH grant is approved we shall use the \$3,000 saved to slightly bolster our severely cut grant budget for the years 1970 and 1971.

The employment of a full-time systems programmer is most important to the success of the present proposal. The progress that has been made during the past period is largely due to the fact that we have had the assistance of two able systems programmers (one furnished by the present project and one by NIH) to write the many complex computer programs involved in this research. See Section (3)G for details of the NT-SYS system of computer programs. Mr. David Kirk, our present programmer paid from NIH funds, plans to move to SUNY Stony Brook with us and will start implementing NT-SYS and CIADON at Stony Brook starting in May 1969. His work time in the past and as planned under this grant is divided as follows: approximately one fourth of his time is taken up processing research data in numerical taxonomy generated by the co-principal investigators and research assistants. The second quarter of the programmer's time is taken up with activities which can best be described as public service. An increasing number of persons at The University of Kansas and all over the country are employing numerical taxonomic programs and are in constant need of advice and help in running these programs. Our present programmers spend a considerable amount of time answering replies from other institutions, sending out tapes or card decks, printouts of programs, and aiding researchers and students from various departments of the university who have difficulties with their programming. Our research group has effectively become the center of numerical taxonomic activity in this country and such activity will steadily increase. We are certain that the demand for such services will continue at our new location. By devoting a portion of the programmer's time to such activities, the granting agency is in effect subsidizing a wide range of taxonomic endeavors by scientists as a whole. We could easily document this with 20 to 30 research studies in biological systematics, ecology, sociology, psychology, political science, history, and even the humanities, which have directly benefitted from the developmental work undertaken under this grant.

The remaining half time of the programmer's work is devoted to the writing of new types of numerical taxonomic programs as these are developed by statistically minded investigators. By concentrating the programming investment at one locality, writing these programs in FORTRAN and then widely distributing them to the community of taxonomists, considerable saving of funds and conservation of effort can be achieved. During the six months of this grant most of Messrs. Kishpaugh and Kirk's efforts will be devoted to implementing NT-SYS on the SUNY-Stony Brook computer.

Our present systems programmer is Mr. David Kirk who has considerable programming experience especially with an IBM 360 system. He is highly qualified for this position.

The two graduate students would be employed at half time for the four months.

Secretary. We currently swap one secretary with our manuscripts and research correspondence. The departmental secretary would not be able to keep up with this work.

Computer costs. The computer charges of \$2,500 are only a portion of the estimated computer costs of our research. We estimate the total computer charges for our research under this grant will be in excess of \$10,000 for the six-month period. The additional computer time will be furnished by the SUNY Stony Brook Computation Center as an added University contribution to this research.

(9) Current research support of co-principal investigators.

F. J. Rohlf	Remaining balance
NSF On-line computation for statistical analysis (equipment only, no salaries)	
Expires November 1970.	\$ 13,000

R. R. Sokal	
NSF Natural selection during population growth of <i>Tribolium</i> . 1 June 1964 - 31 May 1970.	Out of funds

NIH Research in statistical biology.	Remaining balance
Terminates 31 August 1969.	\$ 30,000

Robert R. Sokal is currently supported by a Career Development Award of the NIH which will terminate 31 August 1969.

The following applications are pending.

R. R. Sokal and F. J. Rohlf	
NIH Research in statistical biology	Recommended at \$79,000
To start 1 September 1969.	but not yet funded

R. R. Sokal	
NSF An experimental and simulation study of ecogenetic processes in laboratory insects populations.	

J. S. Farris, E. J. Rohlf and R. R. Sokal
Renewal of this grant
To start 1 January 1970.

Recommended at
\$ 70,000
for 2 years
but not yet funded

(10) Application for related research to other agencies.

No research proposal for the above projects has been made to any other agency.

Some work in numerical taxonomy will be supported by NIH if the grant entitled "Research in Statistical Biology" will be renewed. Only a limited portion of that grant is devoted to numerical taxonomy, the other topics of investigation being geographic variation analysis, factor analysis, and mathematical modelling. Individual projects in numerical taxonomy will be assigned to one or the other of the grants, so that no overlap will result.

NATIONAL SCIENCE FOUNDATION

WASHINGTON, D. C. 20550

RESEARCH GRANT BUDGET SUMMARY

INSTITUTION National Science Foundation		PRINCIPAL INVESTIGATOR(S) Prof. Sokal and Rohlf			PROGRAM NAME			DURATION	GRANT NUMBER
		NSF Funded Man Months			Grantee Man Months			NSF GRANT	GRANTEE SHARE
		Cal	Acad	Sum	Cal	Acad	Sum		
A. SALARIES AND WAGES									
1. Senior Personnel									
a. (Co) Principal Investigator(s)				1				\$	\$
b. Faculty Associates									
Sub-Total									
2. Other Personnel (Non-Faculty)									
a. Res. Assoc. - Postdoc.									
b. Non-Fac. Professionals - Doc.									
c. Non-Fac. Professionals - Other									
d. Grad. Students (Res. Asst.)				8					
e. Professional School Students									
f. Pre-Baccalaureate Students									
g. Secretarial-Clerical									
h. Technical									
i.									
3. Special Allowances								\$	\$
Total Salaries and Wages								\$	\$
B. FRINGE BENEFITS									
Total Salaries, Wages and Fringe Benefits (A+B)								\$	\$
C. PERMANENT EQUIPMENT									
1.								\$	\$
2.								\$	\$
Total Permanent Equipment								\$	\$
D. EXPENDABLE SUPPLIES AND EQUIPMENT									
E. TRAVEL									
1. Domestic								\$	\$
2. International								\$	\$
Total Travel								\$	\$
F. PUBLICATION COSTS									
G. OTHER COSTS									
1.								\$	\$
2.								\$	\$
3.								\$	\$
Total Other Costs								\$	\$
H. TOTAL DIRECT COSTS (A through G)									
I. INDIRECT COSTS									
1. On Campus % of								\$	\$
2. Off Campus % of								\$	\$
Total Indirect Costs								\$	\$
J. TOTAL COSTS (H and I)									
K. AMOUNT OF THIS AWARD (ROUNDED)									

NATIONAL SCIENCE FOUNDATION

WASHINGTON, D.C. 20550

APR 1 1969

In reply refer to:
9/639-5021
American Institute of
Biological Sciences

Dr. G. F. Estabrook
Taximetrics Laboratory
University of Colorado
Boulder, Colorado 80302

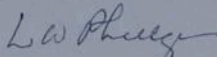
Dear Dr. Estabrook:

We are writing to ask your advice on the accompanying proposal, which has been received by the Foundation for consideration under the Special Projects Program. The objectives of this program are to encourage and assist scientists and science educators to carry out projects for the improvement of science education at the undergraduate level. Primary consideration is given to projects which show promise of being widely applicable, but this does not preclude proposals for more local orientation which contain innovative ideas. With these objectives in mind, we should appreciate having your views on the merit of the proposed activity, the feasibility and effectiveness of the procedures, the qualifications of the project personnel, the reasonableness of the budget, and any other factors you feel are pertinent. Your specific recommendation as to the desirability of Foundation support will be appreciated.

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

A reply at the earliest time that is convenient for you will help us in our effort to respond quickly to proposals. We should like to be able to make a decision on this proposal by April 23, 1969. Please return the proposal and your reply in the enclosed envelope to Professional Associate, Undergraduate Education in Science.

Sincerely yours,



Lyle W. Phillips
Division Director
Undergraduate Education in Science

Enclosures

4 April 1969

Mr. Lyle W. Phillips, Division Director
Undergraduate Education in Science
National Science Foundation
Washington, D. C. 20550

9/639-5021 A.I.B.S.

Dear Mr. Phillips:

We are returning Dr. Olive's above-cited proposal, because Mr. Estabrook is in Europe for the month, and unable to meet your deadline.

We regret causing this inconvenience.

Sincerely yours,

(Mrs.) Genevra Metcalf
Secretary

m
Enc.

NATIONAL SCIENCE FOUNDATION

WASHINGTON, D.C. 20550

January 9, 1969

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

Dear David:

Ref: B012587

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.

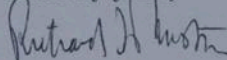
Your critical review of the merits of the project will be of great value. Aspects which may enter into such an evaluation include the relative importance, feasibility, and conceptual framework of the project. You may also wish to consider the capacity of the investigator to conduct successfully the contemplated 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.

Sincerely yours,



Richard F. Johnston
Program Director for
Systematic Biology

Enclosures

NATIONAL SCIENCE FOUNDATION
PROPOSAL RATING SHEET

Reviewer

Proposal No. B012587
Investigator: Dr. Sokal
Institution: SUNY/Stony Brook
Please return to: Systematic
if possible by: 24 JAN 1969

Comments (Continue on additional sheet if necessary)

The application cannot be evaluated in the normal sense of a research proposal inasmuch as the application solicits funds to continue the work of a school rather than a particular research endeavor. We have to assume that we can anticipate continued productivity from the very productive school which Dr. Sokal has led for a number of years. The addition of Dr. James Farris to the group adds considerable luster to the application. He has demonstrated already considerable ability as a designer of models in systematic biology and his collaboration with Drs. Sokal and Rohlf seems to add a new dimension. I therefore recommend this proposal as being meritorious.

The budget involved seems to be generally well constructed, but with two points where changes would be reasonable. Under personnel, the systems programmer, whose salary is indicated as 9000 a year, all from this grant, seems to be low. This individual will be of prime importance as the group moves from Kansas to New York. His function, amongst others, will be the conversion of programs running on the computers in

OVERALL RATING

- ☐ EXCELLENT
☐ VERY GOOD
☐ GOOD
☐ FAIR
☐ POOR

Signature of Reviewer

Date:

Other suggested reviewers (optional):

Kansas to run on the machine in New York. A systems programmer in the New York area makes a considerably larger salary than in Kansas and the individual will find himself torn between his work for Dr. Sokal and the desire to achieve larger income. The other point in the budget that needs comment is the computer charges, which I consider to be high. The computer system in New York is probably much faster than the systems used by the group in Kansas and can probably execute programs at a much faster rate. While I do not know the hourly charge for the Central Processing Unit on the new machine, I would guess it will be very favorable to in-house users, and, therefore, it would be my anticipation that unless very large production runs are planned (which are not indicated) a much smaller budget would be necessary for the computer use. However, this item should be checked with Dr. Sokal, and better justification of the item given.

PN-9654

RESEARCH PROPOSAL
(Renewal of Grant GE-4927)

NSF CODE NO.

042 1460

Submitted to:

B012587-30DE68

NATIONAL SCIENCE FOUNDATION
Biological and Medical Science Division
Systematic Biology Branch
Washington, D.C. 20550

Submitted by:

The Research Foundation of State University
of New York
P. O. Box 7126
Albany, New York 12224
for and in conjunction withState University of New York
Campus:State University of New York at Stony Brook
Stony Brook, Long Island, New York 11790

Research site:

Department of Biological Sciences
SUNY at Stony Brook
Stony Brook, New York 11790

Co-Principal Investigators:

Dr. James S. Farris
Assistant Professor of Biological Sciences
Dr. F. James Rohlf
Associate Professor of Biological Sciences
Dr. Robert R. Sokal
Professor of Biological SciencesAddress all correspondence to Dr. Sokal.
Until Summer 1969 at
Department of Entomology
The University of Kansas
Lawrence, Kansas 66044
(913) 864-3706Title of proposed research:
(Renewal)The Development and Testing of Methods of
Numerical Taxonomy

Desired starting date:

1 January 1970

Proposed duration of renewal:

2 years

Amount requested:

\$203,125

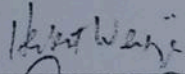
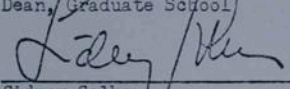
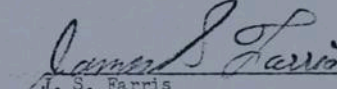
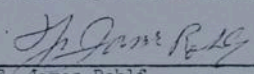
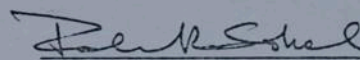
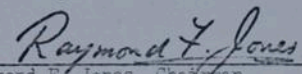

Herbert Weisinger
Dean, Graduate School
Sidney Gelber
Acting Vice President for Liberal Studies
J. S. Farris
Co-Principal Investigator
F. James Rohlf
Co-Principal Investigator
Robert R. Sokal
Co-Principal Investigator
Raymond F. Jones, Chairman
Department of Biological SciencesThe Research Foundation of
State University of New York
Robert G. Parr
Treasurer

TABLE OF CONTENTS

Information relating to administration of this grant.....	ii
I. Abstracts of reports and proposals.....	1
(1) Results to date.....	1
(2) Proposed research if support is continued.....	1
II. Detailed reports and proposals.....	2
(1) Reasons for joint proposal.....	2
(2) The intellectual milieu for the proposal renewal.....	2
(3) Report of work done from inception of project to date.....	3
(4) Bibliography of publications resulting from this project since its inception.....	13
(5) Proposed research if support is continued.....	20
(6) Facilities.....	27
(7) Personnel.....	27
(8) Proposed budget for continued support.....	33
(9) Explanation of budget.....	35
(10) Current research support of co-principal investigators.....	37
(11) Applications for related research to other agencies.....	38

Information Relating to
Administration of this Grant

Drs. Rohlf and Sokal, two of the co-principal investigators, will join the staff of the State University of New York at Stony Brook beginning 1 September 1969. Dr. J. H. Camin, one of the current co-principal investigators at The University of Kansas, does not plan to continue in this field of research. However, Dr. J. S. Farris, (newly appointed to the staff of the Department of Biological Sciences at SUNY Stony Brook) has been very active in numerical taxonomy. He wishes to become co-principal investigator with the remaining two members for the renewal of this grant.

On or before 1 September 1969 we expect to transfer the current grant (GB 4927) to the State University of New York at Stony Brook. The funds remaining in this grant should carry us through 31 December 1969 and the present application is being made to assure continuance of our work beyond that date.

I. Abstracts of reports and proposals.

(1) Results to date.

Work has been carried out on 9 separate approaches to the development and application of numerical taxonomy and related subjects. These approaches range from the general theory of systematics, over the principles of taxonomic methodology, to numerical classifications of various groups of organisms. In aggregate they have led to 23 separate research projects, some completed and published, others still underway. An extensive programming effort has been made for numerical taxonomy and this is benefitting not only our work but that of colleagues at other institutions. A total of 18 publications are in various stages of preparation from work which has already been completed. Parallel and interrelated studies in numerical taxonomy financed by the National Institutes of Health are briefly described. These have led to another 11 publications in the field.

(2) Proposed research if support is continued.

The applicants propose continuation and completion of several of the projects whose progress is discussed in detail under previous research [Section (3) below].

A variety of new investigations is planned which can be subsumed under the following four headings: 1) the further development of general systematic theory; 2) improvement of the techniques of numerical taxonomy; 3) tests of the various hypotheses underlying the theory of numerical taxonomy; and 4) the application of numerical taxonomy to various exemplary groups of organisms. The various projects discussed in detail in Section (5) of this proposal will be undertaken singly by one of the three co-principal investigators or by various combinations of these, their associates and graduate students. Accompanying these specific research projects there will be strong emphasis on the further development of a powerful library of computer programs for taxonomy and considerable effort will be made to distribute these programs widely to workers at other institutions.

II. Detailed reports and proposals.

(1) Reasons for submitting a joint proposal.

We have decided to continue to submit a joint research proposal for two reasons: administrative efficiency and intellectual and pedagogical interaction. It is simpler for the home institution to administer one grant rather than three separate ones dealing with a single research topic. We suspect that this holds true for the administrative work at the granting agency as well.

Our main reason for the joint application, however, is the nature of our past and intended research efforts. Our research in general systematic theory and numerical taxonomy at The University of Kansas has been characterized by considerable interaction and collaboration among staff members and students. We intend to continue in this manner within the newly established Ecology and Evolution Section of the Department of Biological Sciences at the State University of New York at Stony Brook. We shall continue our system of weekly seminars in which new ideas and new techniques are constantly exposed to our colleagues for mutual criticism. The co-principal investigators maintain rather different philosophical positions in their attitudes to numerical taxonomy. It is hoped that this diversity of approaches will lead to a profitable re-examination of our points of view and to fruitful interaction on a variety of subjects.

The three co-principal investigators will remain free to pursue their own lines of investigation, yet at the same time will interact in various as yet unpredictable ways to collaborate on certain aspects of their research. We intend to share research funds, facilities, and occasionally even graduate students. It is expected that this approach will be as productive at Stony Brook as it has during our tenure at The University of Kansas.

(2) The intellectual milieu for the proposed research.

The research will be carried out in the Section on Ecology and Evolution being organized as a subdivision of the Department of Biological Sciences at the State University of New York at Stony Brook. This section will eventually assume departmental status. It is being organized under the chairmanship of Professor Lawrence B. Slobodkin, formerly of The University of Michigan and currently consists of eleven staff members. Four further staff members are expected to be added for the 1969-1970 academic year. The establishment of a strong mathematically and statistically oriented unit in ecology and evolution is one of the primary objectives of the biological sciences program at SUNY Stony Brook, and the early establishment of strength in the field will greatly ease the recruitment of further staff members for the general field of population biology, which is currently undergoing rapid and extensive development. The interaction of a research group in numerical taxonomy with others in the general field of population biology should lead to mutually beneficial results, some of which are suggested in the body of the application.

The main emphasis in the past work of the co-principal investigators has been on the development of a general theory and methodology for systematics rather than on the detailed classification of specific groups of organisms, although all three of us have done this as by-products of our general and methodological researches. For this reason we expect the results of the research carried out under this grant to have broad applicability in the field of systematics. In order to apply our new ideas and techniques to the field of systematics, interaction with various practicing systematists will be sought. We plan to collaborate with workers in the Department of Biological Sciences, the newly formed Marine Sciences Center, and the Department of Earth and Space Sciences at SUNY Stony Brook, and The American Museum of Natural History. We propose to make arrangements whereby our students may serve research apprenticeships in The American Museum of Natural History, using its collections for numerical taxonomic theses.

(3) Report of work done from inception of project to date.

This report is subdivided into eleven separate approaches (labeled A through J), differing considerably in extent, amount of work done, success of the study, and importance of the work. Under each approach there are one or more studies, representing separate research projects. These projects are numbered continuously from 1 to 24 over all the approaches to permit easy cross-reference. For background information regarding our work for the present reviewing panel we also include brief accounts of projects carried out prior to the current grant period.

Although it is in the nature of this work that studies under most of the approaches are continued as new techniques and insights develop, several of the studies have reached states of temporary completion, and publications or manuscripts in press have resulted from them.

A. THE GENERAL THEORY OF SYSTEMATICS

Numerical taxonomy is undergoing continuous change and elaboration; the results of this work profoundly affect the research carried out under the various approaches listed below.

1. A review of numerical taxonomy.

This paper aimed at a botanical audience and prepared at the invitation of the American Society of Plant Taxonomists was subsidized by this grant and has been published (Sokal, 1963).

2. Nomenclature.

C. D. Michener has been interested in integrating numerical with orthodox taxonomy. He has advocated the application of a unimomial nomenclature as well as of an appropriate numbering system for organisms to accompany the advent of electronic data processing in taxonomy. These views are stated in a general article on "Future Developments in Taxonomy" (Michener, 1963) and an illustrative example of his nomenclatural technique is featured in a subsequent article (Michener, 1964).

3. A re-examination of the principles of numerical taxonomy.

During the past year Sokal has been engaged in an extensive review of the theory of systematics from the point of view of the numerical taxonomist. This work is in connection with the preparation of the second edition of the "Principles of Numerical Taxonomy" in collaboration with P. H. A. Sneath. This will be an entirely rewritten volume. Development of the theory will range over the entire field of systematics. Much of the theoretical work has been completed and a summary of the present status of the field has been given in two extensive review papers published in Russia at the invitation of the editor of the Soviet Journal of General Biology (Sokal 1967, 1968).

4. Relations between phenetics and cladistics.

A paper by Camin entitled "The significance of incongruence between phenetic and cladistic dendrograms," has been completed. The paper explores the biological meaning of phenetic-cladistic congruence and incongruence and discusses the possible application of comparative studies of phenetics and cladistics for the same group and of organisms and characters in providing some insight into relative rates of evolutionary divergence in different sections of such groups.

5. A critical evaluation of the biological species concept.

This study by Sokal and Crovello (not anticipated in the original application) grew out of the extension of the methods and philosophy of numerical taxonomy to studies at the species level. It appeared to these authors that the biological species was not an operational concept (as had also been claimed by others, Ehrlich and Holm, 1962). It therefore seemed worthwhile to develop a model in which we would go step by step through a procedure which would enable us to ascertain whether a given sample of organisms comprised one or more biological species. Even under optimal conditions when full knowledge about biological operations and relations is available, it became apparent that phenetic decisions must be made at certain stages in the procedure, or no progress can be made. Consequently, the biological species concept is clearly not operational under ordinary circumstances, when knowledge and time are limited.

These authors then examine whether the biological species concept has heuristic value or is necessary for biological theory. They find that neither is the case and recommend that attempts to describe biological species be superseded by the description of phenetic and reproductive relationships among organisms without any preconceived hierarchic arrangement. This manuscript has just been completed and will be submitted for publication soon.

B. PRINCIPLES OF TAXONOMIC METHODOLOGY

We have been much concerned with the analysis of the principles by which conventional taxonomists operate when classifying organisms and with the principles which would seem suitable for such classificatory operations whether done by men or by computer. A significant step in this direction has been our study of the "intelligent ignoramus" which is discussed in Section J below.

6. The Taxocrit study.

This is a related study employing the Caminalcules, hypothetical imaginary animals used earlier for cladistic research. In this study numerous students, colleagues, and even laymen were asked to classify the 29 "species" or Caminalcules into natural groups by whatever criteria they wished to employ. These classifications will be quantified and compared with a detailed phenetic numerical taxonomic study of this group in which a large number of characters had been scored. The method employed is borrowed from a personnel evaluation study by Ward and Christal and will indicate to us which aspects of the phenetics of the Caminalcules different taxonomists emphasize in making their classification. It will be interesting to see what groupings exist among taxonomists in terms of their classificatory procedures and how these are related to their perception of form differences. The experimental part of this study has been completed and only the analysis remains.

7. The robustness of numerical taxonomic methods with respect to random errors in homology of characters.

Fisher and Rohlf (1969) investigated the effect of errors in determining the homology of characters on the results of a numerical taxonomic study. It was found that the distance coefficient was much less sensitive to these random errors in the data matrices. These analyses also suggested that in a study based on 74 characters the classification was not degraded by confusion about homologies for six characters, but was adversely affected by a larger number of miscoded characters.

8. Test of hierarchic structure in taxonomic data.

Rohlf and Fisher (1968) established approximate critical bounds for testing for the presence of hierarchic structure in taxonomic data constructed using random samples from multivariate normal and uniform distributions. Sokal's taxonomic distance coefficient was found to be much more sensitive to differences in multivariate structure of the sampled distributions than was the correlation coefficient.

C. REPRESENTATION OF TAXONOMIC STRUCTURE

An important part of the work of the taxonomist, is to represent relationships in some coherent taxonomic system. We have engaged in the following studies relating to this problem.

9. The use of stereograms in numerical taxonomy.

Rohlf (1968) described a technique for summarizing the results of a numerical taxonomic study by a principal components analysis (with projection of the OTU's onto the first three principal components). The resulting three-way scatter diagram can be constructed as a solid model, as perspective drawings, or as stereo pairs. Computer programs were written to compute the coordinates of the OTU's so that a perspective stereo drawing could be made. Programs were also written to prepare tapes so that the stereograms could be automatically drawn by a Benson-Learner incremental plotter. The data for the *Hoplitis* complex (Michener and Sokal, 1957) were used as an illustrative example in this paper. The principal advantage of these techniques over conventional cluster analyses is that they enable one to ascertain visually which OTU's are intermediate between clusters, and to distinguish these from OTU's distant from all clusters in the study. It is sometimes difficult to distinguish these two types of OTU's when only a phenogram is available. This method also enables one to observe relations in hyperspace among clusters (phenograms contain no information about this).

10. The effect of correlated characters on numerical taxonomic data.

Rohlf (1967) studied the effect of using a set of rather highly intercorrelated characters in a numerical taxonomic study (74 external morphological characters of 45 species of North American mosquito pupae). It was found that the main effect of the character correlations was to cause the clusters to be elongated into hyper-ellipsoids. Standard methods of cluster analysis have a tendency to make the OTU's at the ends of these clusters appear to be much more isolated than they perhaps "should be". Methods for correcting for this bias were discussed, and the use of multidimensional scaling techniques such as principal components analysis were suggested for such data. As is typical of data with a large general factor (due to the overall correlation of characters), cluster analysis of the correlation rather than the distance matrix gave a much clearer result.

11. Analysis of hyper-ellipsoidal clusters.

Rohlf (1965, 1967, and 1968) has been interested in the effects of different types of data structures on the results of cluster analyses. The majority of present-day cluster analyses work most efficiently on clusters which are hyperspheroidal in shape, i.e., OTU's clustered around a point, the centroid of the cluster. But other shapes of clusters are possible and have been found to exist in taxonomic data. These are typically handled very poorly by conventional cluster analysis. Rohlf (1966, unpublished paper presented at the Brookhaven meetings of the Biometric Society; 1969 in preparation) has developed a clustering algorithm which allows for a wide variety of basic cluster shapes. The clustering

algorithm is basically a sequential agglomerative clustering procedure. A modified generalized distance function is used to compute the distance between a cluster and the OTU's which are candidates for addition to the cluster in the next clustering cycle. The use of this distance function rather than an ordinary average distance or a centroid distance function allows the clusters to become elongated into hyper-ellipsoids, since the generalized distance function considers distances in certain directions to be more important than others. OTU's at the ends of an ellipsoidal cluster are considered to be closer than OTU's around the side of the cluster. The generalized variance of a cluster (its P-volume) is used as an index of homogeneity of OTU's within a cluster. By a simple technique of adding dummy variables to the original variables included in the study one can compute distances in a way such that OTU's which are consistent with the trend of the shape of the cluster are considered close to the cluster, and OTU's inconsistent with the trend are considered further away. For example, in a ring-shaped cluster the addition of the dummy variables X^2 and Y^2 will cause the generalized distance function to represent distances of OTU's from the ring rather than the centroid of the cluster. Consequently OTU's near the centroid would also be considered farther away from the cluster than OTU's around the outer edge of the ring. The rate at which distances increase as one moves away from the cluster is a function of the closeness with which the OTU's in the cluster follow a definite curvilinear relationship. Clusters of any shape may be handled by this technique if the appropriate dummy variables are added to the matrix. As a by-product of the computations, the cluster analysis indicates the apparent shape of the cluster at each step in the clustering so that one is aware of the fact that, for example, a parabolic shaped cluster has been isolated. The principal limitation of this method is that one must decide a priori what class of dummy variables should be added to the matrix, and this in turn determines what classes of cluster shapes can be recognized by the program.

12. CABOD cluster analysis.

Drs. Rohlf and F. Möller (Istituto di Statistica, Università Bocconi, Milano, Italy) have worked on developing a form of cluster analysis which has the following properties: 1. It is based on the notion of the information content of a character. Characters are evaluated and used on the amount of information they contain about other characters in the study. 2. The method clusters by subdividing OTU's into groups (rather than by adding OTU's onto previously found clusters), i.e., it is a divisive type of clustering. It was hoped by this approach to give greater reliability to the important higher level clusters. 3. It attempts to reduce the effect of errors in sampling of OTU's, i.e., to allow for the fact that one may have selected a large number of OTU's from one genus, and a few from another. The procedure also allows for the simultaneous construction of an identification key for the clusters it finds. The programming of this method has been completed, but

it was found that it did not give satisfactory results when used in actual practice. The principal difficulty seems to be the fact that it uses a divisive clustering algorithm, and the fact that the clustering program divides OTU's into groups on the basis of a monothetic criterion. The principal result of the consultation with Dr. Möller during his stay in Lawrence in the summer of 1966 was to abandon this type of clustering.

D. THE EFFECT OF CONVERGENCE ON PHENETIC RELATIONSHIPS

13. A study of rodents.

In this study by A. J. Boyce of Oxford University, a research assistant with us for one year, the behavior of numerical taxonomic methods were investigated when applied to forms some of which are known to be convergent.

A jumping mode of progression has supposedly arisen independently in several rodent groups. Usually such groups also contain members which do not show this specialization and forms were chosen so that both types were represented. The material was taken mainly from the North American families which contain the kangaroo rats and mice, the spiny and pocket mice, the gophers and the jumping rats and mice. Representatives of the Old-World gerbils and spring-hares and the generalized murine forms, the wood and black rat, were also studied. Each of 17 genera was represented by a number of male and female specimens (64 in all). For each OTU 244 linear skeletal measurements were taken (122 cranial, and 122 post-cranial). Various similarity coefficients were computed. No coefficient produced a pattern of relationships in complete agreement with the existing taxonomic arrangement. The correlation matrix indicated relationship in functional adaptation and the distance matrix seemed indicative of overall size. The effects of size may be reduced by ignoring the first principal component of the variance-covariance matrix. The pattern of relationships with respect to the second and third principal components is in fairly good agreement with that produced by a cluster analysis of correlations between the forms.

The distribution of the families in the space defined by the second and third latent vectors clearly shows the differences in diversity within and between the families. Similarities between the members of different families adapted to a jumping mode of progression are not sufficient to produce any overlap of the families in this space.

The study was incorporated as part of Mr. Boyce's thesis for the doctorate at Oxford University (Boyce, 1965). It is to be published soon.

E. THE NONSPECIFICITY HYPOTHESIS

There is continuing interest in measuring the amount of congruence between classifications based on different suites of characters.

14. A new test of the hypothesis of non-specificity on the Hoplitis complex.

This is a study of the degree to which phenetic similarities among species of bees of the Hoplitis complex, are the same in the two sexes and in different body regions. A study was carried out on 97 species, partitioning the 122 recorded characters into 69 found in females and in both sexes versus 53 found only in males. In a companion study of the same species using the same 122 characters, 60 head characters were compared with 62 nonhead characters. Phenograms were computed showing the classifications of these bees based on all characters and on the subsets of characters described above. The cophenetic correlation based on correlation matrices are 0.71 for sexes and 0.61 for head versus nonhead characters. Our study discusses the partial failure of the nonspecificity hypothesis on the basis of the present findings and compares these with the results of other recent authors. This study has been published by Michener and Sokal (1966).

F. NUMERICAL CLADISTICS

There have been several recent attempts to develop methods for estimating cladistic relationships (branching sequences of evolutionary lineages). Our work at The University of Kansas on this subject originated from the investigation on the principles of taxonomic methodology reported in Section B.

15. The original study.

This study was based on the "Caminalcules", a group of imaginary animals generated by J. H. Camin according to principles intentionally known so far only to him but which he believes to be consistent with what is generally known of transspecific evolution. Genetic continuity was established by tracing the drawings of the animals from sheet to sheet, permitting the preservation of all characters except for such modifications as were desired.

This study has led to an empirical method based on the principle of parsimony of evolutionary steps which appears capable of estimating probable cladistics from the characters of existing organisms. This method has been successfully applied to known phylogenies of hypothetical animals and to assumed phylogenies of well known groups of organisms such as fossil horses. A first publication on this work has appeared (Camin and Sokal, 1965).

Several studies on other organisms have been completed and are awaiting publication. These include the Fusulinidae (with R. Kaesler), the Mecoptera (by G. W. Byers), the Hoplitis complex of bees (R. R. Sokal and C. D. Michener), the vipers (H. Marx and J. Felsenstein), the genus Arctostaphylos in plants (P. V. Wells), chromosomal inversions in blackflies (R. Hansell and D. Chant), and tiger beetles (H. Willis). A computer program has been prepared by Mr. Ron Barcher to carry out the procedures of numerical cladistics and present the results in easily intelligible fashion. This program is generally available (see Section G).

16. Some recent improvements.

In a doctoral dissertation, Hendrickson (1967) examined the pattern cladogram as defined by Camin and Sokal and found that certain anomalies could be traced to the presence of an empty internode in some places in this pattern cladogram. The pattern cladogram was therefore redefined by him as "the included subset pattern cladogram". A procedure for determining the maximum number of extra steps a given character can take in being fitted to another pattern is given, and the possible use of this information in a weighting function has been discussed. Since the compatibility matrix constructed from the redefined pattern cladogram is symmetrical with respect to compatibilities, it is possible to make a number of inferences about the structure of the cladogram from this compatibility matrix. This led to the development of the restricted extra step partitioning procedure.

The character space is presented in terms of graph theory, and the question of parsimony is examined as a problem in this field. The most parsimonious cladogram is a tree of minimum total length, constrained to a directed city-block lattice. This is discussed in detail in Hendrickson (1968).

Some suggestions have been made for the use of techniques of "data analysis" in locating cases in which certain of the assumptions of character coding should be relaxed. In particular, cases of probable convergence and reversals have been considered.

Several sets of data have been analyzed using these methods, and these have pointed out some of the shortcomings of this method.

In general, the conclusions that can be reached are that many more characters need to be considered than are normally considered for cladistic analyses, and that in view of this conclusion, it is paradoxically more likely that satisfactory cladistic reconstructions can be made on the basis of studies of recent organisms than from fossil organisms.

Since the reconstruction of the cladistic sequence is subject to considerable bias and possible distortion in most sets of data, the use of cladistics as a basis for classification seems unwarranted. The most effective use of numerical cladistics seems to be as a test of the internal consistency of evolutionary speculation.

This work is being prepared for publication.

G. THE DEVELOPMENT OF COMPUTER PROGRAMS FOR NUMERICAL TAXONOMY

Since the inception of the project, numerous computer programs for numerical taxonomy have been prepared. These programs have been organized into two comprehensive systems of programs. The NT-SYS programs (developed by Dr. F. J. Rohlf in collaboration with Mr. John Kishpaugh, Mr. Ron Bartcher, and Mr. Jerry Crow) are for numerical phenetics. The CLADON programs perform computations for numerical cladistics (developed by Mr. Ron Bartcher).

17. The NT-SYS programs.

These have been organized into an integrated system of programs with provisions for simplified free format parameter cards which call up the various subprograms and specify the data to be acted on. Provisions are also made for storage of data matrices (and matrices which result from the computations) on a master "stack" tape along with symbolic identification codes. The present version (October 1968) of NT-SYS contains over thirty basic entry points which allow the calling up of programs which constitute over 70 links.

This package of programs (prepared under the current grant) is one of the most sophisticated multivariate analysis packages available today. It was used extensively on the IBM 7040 and the present GE 625 computing system at The University of Kansas. Requests for this program have come from numerous institutions in this country and overseas. At present plans are being made to publish the system in the Computer Contributions of the Kansas State Geological Survey. Work has just started on developing a version of NT-SYS for the IBM 360/65 computing system at the State University of New York, Stony Brook.

The following is a brief listing of the types of computations allowed for within NT-SYS.

a. Matrix algebra. Input to this program consists of statements such as $X = (A + B) * C - 1$ which would result in matrices A and B being added together, their sum being multiplied by matrix C, and the inverse of the result stored as matrix X.

b. The coefficient of association program allows the computation of various association coefficients.

c. The CORDST program is used for the computation of correlation, distance, variance - covariance, and cosine θ data matrices.

d. An EDIT program is included which permits manipulation of the master stack tape containing the matrices.

e. A general FACTOR extraction program is included with provisions for estimation and iteration of communality estimates in factor analysis. Options are provided for using Thurstone's complete centroid method of factor extraction, the power method, the Jacobi method, and the Wilkinson-Householder methods for computation of eigenvalues and eigenvectors. Plans are being made for the inclusion of options for alpha, image, canonical, and maximum likelihood factor analyses.

f. The GTAXON program is a generalized sequential agglomerative cluster analysis program which takes into consideration within-group variation of OTU's. It also allows for elongated and curvilinear clusters.

g. The INPUT program is used for the inputting of matrices and storing the information on the master stack tape.

h. The MAXTAX program performs iterative manipulation of a phenogram to attempt to maximize the cophenetic correlation.

i. The MIAM program performs an oblique rotation of a factor matrix to simple structure using Sokal's mass modification of Thurstone's analytical method for simple structure.

j. The MIXCNG program allows for the deletion and reordering of rows and columns of matrices.

k. The MULDIS program is a multi-group generalized discriminant function analysis program.

l. The MXCOMP program performs a comparison of two symmetric matrices for congruence and computes the cophenetic correlation.

m. The MKPLOT program is used to plot rows or columns of a matrix against one another on a two-way scatter diagram on the line printer.

n. The OUTPUT program is used for the printing and punching out matrices.

o. The PHNGN1 and PHNGN2 programs are special phenogram output programs.

p. The PROJET program performs a projection of OTU's onto a given set of axes.

q. The SORTIT program prints a sorted listing of correlation or distance values for each variable in a correlation or a distance matrix.

r. The STAND program standardizes the characters of a data matrix.

s. The TAXON cluster analysis program has options for single and complete linkage, weighted and unweighted average linkage analysis, as well as Spearman's sums of variables method and centroid clustering.

t. The TEXT program is a general text editing routine which is used for formatting and outputting writeups of the NT-SYS system.

u. The TRANFM program performs various transformations of elements of matrices. Transformations such as arc sin, arc cos, Fisher's z , log, $\log(Y + \frac{1}{2})$, and others are allowed for.

v. The VARMAX performs the varimax orthogonal rotation of a factor matrix to simple structure.

w. The WLKLM program computes the Wilk's statistic.

x. The GOWER program is used to transform association and distance matrices in a way such that the eigenvectors computed from the transformed matrix corresponds to the projections of the OTU's onto the principal components which would result if one were to analyze the matrix of correlations among the characters. This program is used whenever there are many fewer OTU's than characters.

A number of other programs are currently in preparation.

18. The CLADON programs.

The CLADON programs for numerical cladistics consist of three separate programs for computing a compatibility matrix, obtaining an initial solution by the monothetic method, and an iterative method for finding the minimum length tree. The cladogram is close to the form in which taxonomists are used to looking at data of this sort, i.e., cladograms with characters marked across branches of a cladogram, and there is a total count of the number of required evolutionary steps at the bottom of a tree. This program has also been requested by a number of institutions. Among them are the University of Chicago, The Chicago Natural History Museum, Kansas State University, The University of Michigan, etc. It has been published by its developer (Bartcher, 1966).

H. NUMERICAL TAXONOMIC STUDIES OF VARIOUS TAXONOMIC GROUPS

19. A numerical classification of the orders of insects.

This research project was started from "seeding funds" by the General Research Fund of The University of Kansas. It has been financed from the current grant for two years. The aim is to see whether methods for numerical taxonomy are applicable to higher categories (using the exemplar method) and what can be learned about the methodology in such a case. A second aim is to see what light methods of numerical taxonomy can throw upon the phenetic similarities among representatives of most of the orders of insects. Thirdly we wish to apply the methods of numerical cladistics to this study as well. The study is based on 39 species representing the major insect groups. These data have been processed through the computer, and are awaiting analysis and preparation of a manuscript.

20. A numerical taxonomic study of the mite superfamily Celaenopsoidea.

A Ph.D. dissertation by Richard C. Funk, entitled "Revision of the family Euzerconidae and its relationships within the superfamily Celaenopsoidea (Acarina: Mesostigmata), aided by the techniques of numerical taxonomy," has been completed. It is a numerical taxonomic analysis of a relatively unknown group of mites, including 30 species of Euzerconidae and 16 additional species of non-euzerconid mites in the superfamily Celaenopsoidea. The study was conducted in two parts. The first part was a study of adult female specimens of 25 species of Euzerconidae in which individual specimens served as OTU's. The study was based on 76 OTU's and 116 characters. The second part was an exemplar study, using both males and females, of 46 species of Celaenopsoidea, the largest superfamily of Trignyaspid Mesostigmata, and included the Euzerconidae. This study was based on 179 characters, 80 of which had been utilized in the first study. In addition to yielding a useful classification of the family Euzerconidae and a tentative classification of the entire superfamily Celaenopsoidea, the study resulted in the synonymy of two and the reclassification of four previously described species, plus the description of 20 new species and seven new genera of Euzerconidae, four new species, three new genera and two new families of non-euzerconid Celaenopsoidea, and the raising of one previously described subfamily of Diplogyninae to family rank. The thesis (Funk, 1968) is being prepared for publication.

21. A numerical taxonomic analysis of the chigger mite genus Euschoengastia Ewing (Acarina: Trombiculidae).

This study, not anticipated in the original application, has been undertaken by William J. Wrenn. The classical taxonomy of chiggers is based almost exclusively on the characters of the parasitic larval instar and in most cases it has been possible

to match the predaceous post-larval instars with the larvae only by the tedious process of laboratory rearing of these mites through the entire life cycle. This study includes the larvae of more than 65 species and the reared and correlated nymphs of 30 of these species. Data on more than 450 characters have been tabulated for most of the larvae and more than 100 nymphal characters have been selected for analysis. In addition to analyzing the taxonomy of the genus *Euschoengastia* by the techniques of numerical taxonomy, we intend to assess the value of numerical taxonomy in matching the post-larval instars of chiggers with their respective larval instars and the possible application in matching the various stages of other organisms with complex life histories, such as parasitic trematodes, hydrozoan coelenterates, and holometabolous insects.

22. Numerical classification of mosquito pupae.

Rohlf (1967) reported on an analysis of 45 species of North American mosquito pupae using 74 external morphological characters. The resulting classifications agreed surprisingly well with the standard classification of the family considering the fact that the pupal stage is not typically taken into account when determining distinctions between genera and subfamilies. In addition, the characters used in this study were not ones ordinarily considered reliable by conventional taxonomists.

I. STATISTICAL METHODS IN SYSTEMATICS

23. A review.

Another activity under this project not anticipated at the time of the original grant application, concerned an extensive and intensive review of the application of statistical methods to systematics in general, including not only numerical taxonomy but the far wider application of statistics to the entire field of systematics. The occasion for such a review arose when R. R. Sokal spent a sabbatical year abroad as a Fulbright professor during 1963-64. This review (Sokal, 1965) is the only such review in the literature and should be of considerable value to diverse systematists contemplating the application of statistical methods to their problems. International interest in this paper, the publication of which was supported by the present grant, has been considerable. A supply of over 500 copies of the reprint was exhausted in a few months.

J. THE PARALLEL RESEARCH IN NUMERICAL TAXONOMY CARRIED OUT UNDER THE SPONSORSHIP OF THE NATIONAL INSTITUTES OF HEALTH

There is close interaction between the activities in numerical taxonomy on the two projects. The development of computer programs for instance is sponsored and partially funded by NSF and by NIH. However, in terms of actual research projects, the work supported by NSF has been

kept separate from that supported by NIH. The latter projects can be divided into the following four general activities: (1) the application of numerical taxonomy to new taxonomic groups, (2) work on the principles and theory of taxonomy, (3) new approaches in numerical taxonomy and the application of numerical taxonomy to new fields, (4) experiments with the techniques of numerical taxonomy.

(1) Among the applications of the techniques to new groups is a study of the mosquito genus Psorophora undertaken and published by Hendrickson and Sokal (1967). In this study the phenetic relationships among 29 species of culicine mosquitoes were based on 158 adult characters. The species included 15 in the genus Psorophora and (for purposes of comparison) 14 in Aedes. The findings were reported as phenograms and as graphs representing the position of the species in a three-dimensional principal components space. Although the various subgenera appear distinct, there seems little justification for uniting three of the established subgenera in the genus Psorophora as distinct from Aedes. From the analysis it is obvious that the classification of these 29 species in the conventional, hierarchic, mutually exclusive manner is unsatisfactory because it does not represent the phenetic relationships revealed by this study.

A second study just completed by Huber (1968) is a revision of 30 genera in 10 families of cockroaches (Blattariae). The purpose of this study (which included first instar and last instar nymphs as well as adults) was to study the changes in phenetic relationships within this group during ontogeny. It was found that classifications based on last instar nymphs are much closer to those of adults than of first instar nymphs. Huber's study included a very thorough reanalysis of a recent revision of the roaches by McKittrick and is perhaps the first extensive reanalysis of a conventional taxonomic study by numerical taxonomy.

(2) Important recent theoretical developments have involved a re-examination of the principles of classification. This was prompted by the development (at Kansas and elsewhere) of numerical methods for analysis of phylogenetic sequences (cladistic analysis). In an extensive review of the principles of classification, Sokal and Camin (1965) examined and contrasted conventional and operational taxonomy and subdivided relationships into those of phenetics and cladistics. They demonstrated that phenetic methods are to be preferred for general classificatory work but that phenetics and cladistics are necessary for a complete understanding of the systematics of a group of organisms. In another paper Sokal et al. (1965) replied in detail to criticism of the phenetic method in numerical taxonomy by Ernst Mayr. The advent of numerical taxonomy has had considerable repercussions on general taxonomic practices. For museum procedures considerable automation has been suggested and numerous projects in this regard are underway. A detailed discussion of efficiency in taxonomy and the application of computers to taxonomic work was prepared by Sokal and Sneath (1966) under this grant. A special analysis of the problems of using electronic

data processing techniques in biological collections was presented by Crovello (1967). Among the projects still underway is that of the "intelligent ignoramus" in which two untrained persons were asked to reclassify a group of specimens of bees that had been previously classified by orthodox taxonomy by C. D. Michener and by numerical taxonomy by Michener and Sokal. The purpose of this experiment was to investigate whether untrained personnel could define characters sufficiently accurately and rapidly so that numerical taxonomy could proceed and establish useful classifications from such data. The results were generally satisfactory, the error between the two "ignoramuses" and between Michener's classification being no greater than that between numerical taxonomic classifications based on different sets of characters. This study by Sokal, Rohlf and Michener has been completed and will be written up and published in the near future.

(3) An application of certain techniques of numerical taxonomy and of trend surface analysis to the classification of the terrestrial vertebrate fauna of Kansas was undertaken by Fisher (1968). By principal component analysis five factors could be extracted from a matrix of correlations among the localities based on the species. Factor scores were mapped by trend surface analysis. These were shown to be of interest where they yielded overlapping faunal zones.

By applying numerical taxonomy to thirteen biotic parameters of eight weevil strains, Fujii (1968) in our laboratory was able to show that taxonomic units clustered on the basis of these characters yielded a satisfactory and consistent classification of these strains and that one could use numerical taxonomy in this way to define similarities between ecological properties of these strains. These properties define imperfect images of ecological niches, whose similarities can then be measured. This application of numerical taxonomy to ecology has been extended and applied to a series of protozoan populations collected in fresh waters by Cairns. Study and analysis of these data by Sokal, Cairns, and Fujii are underway.

Among the new techniques employing numerical taxonomy was a first attempt at simulating the scanning of biological images randomly as well as systematically by means of perforated punch cards superimposed on the images. These studies by Sokal and Rohlf (1966) and Rohlf and Sokal (1968) were applied to 29 species of hypothetical animals (the *Cardinalcules*) and pupae of 32 species of culicine mosquitoes. Black lines appearing through a hole were scored "1" and empty holes were scored "0". Similarity matrices were constructed on the basis of matching scores for corresponding masks and holes. The resulting phenograms agreed well with previous numerical taxonomic studies obtained by deliberately looking for and scoring conventional characters. Complete scanning of the entire image also gave satisfactory results but classifications were not necessarily better than by the random scanning methods. Increasing the size of the apertures was also attempted. It slightly degraded the classifications obtained from the finer scan.

(4) Among studies investigating the techniques of numerical taxonomy, Crovello (1968) examined the effects of standardization versus condensation and of using only an estimate of central tendency as well as that plus an estimate of variation in a basic data matrix. Changes in these techniques appear to result in different relationships. In another analysis of character correlations of the *Hoplitis* complex of bees first worked on by Michener and Sokal (1957), Rohlf (1968) found that when a three-dimensional model was constructed of these groups the operational taxonomic units were clustered into five clusters corresponding well with the previous classification of the group both by numerical taxonomy and conventional taxonomy. The analysis of these correlations is essentially completed and should be written up in the near future. This is a joint study by Rohlf, Michener, and Sokal.

A review of various methods of cluster analyses was undertaken at the invitation of Drs. Maurice Lorr and Samuel B. Lyerly and published as Sokal (1967). It permitted the interchange of considerable information between social scientists and biological numerical taxonomists on common problems and approaches.

(4) Bibliography of publications resulting from this project since its inception.

1. Bartoher, R. L. 1966. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040. Computer Contr. 6. State Geological Survey, The University of Kansas. 54 pp.
2. Basford, N. L., J. E. Butler, C. A. Leone and F. J. Rohlf. 1968. Immunologic comparisons of selected coleoptera with analysis of relationships using numerical taxonomic methods. Systematic Zool. 17: (in press).
3. Boyce, A. J. 1965. The methods of quantitative taxonomy with special reference to functional analysis. Ph.D. Thesis. Oxford University. 190 pp.
4. Camin, J. H. and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. Evolution 19: 311-326.
5. Fisher, D. R. and F. J. Rohlf. 1969. Robustness of numerical taxonomic methods to errors in homology. Systematic Zool. 18: (in press).
6. Funk, R. C. 1968. Revision of the family Euzerconidae and its relationships within the superfamily Celaenopsidae (Acarina: Mesostigmata) aided by the techniques of numerical taxonomy. Ph.D. Thesis. The University of Kansas. 258 pp. + illustrations.
7. Hendrickson, J. A., Jr. 1967. A methodological analysis of numerical cladistics. Ph.D. Thesis. The University of Kansas. 88 pp. + appendices.
8. Hendrickson, J. A., Jr. 1968. Clustering in numerical cladistics: a minimum-length directed tree problem. Mathematical Biosciences (in press).

9. Michener, C. D. 1963. Some future developments in taxonomy. *Systematic Zool.* 12: 151-172.
10. Michener, C. D. 1964. The possible use of uninomial nomenclature to increase the stability of names in biology. *Systematic Zool.* 13: 182-190.
11. Michener, C. D. and R. R. Sokal. 1966. Two tests of the hypothesis of nonspecificity in the Hoplitis complex (Hymenoptera: Megachilidae). *Ann. Entomol. Soc. Amer.* 59: 1211-1217.
12. Rohlf, F. J. 1967. Correlated characters in numerical taxonomy. *Systematic Zool.* 16: 109-126.
13. Rohlf, F. J. 1968. Stereograms in numerical taxonomy. *Systematic Zool.* 17: 246-255.
14. Rohlf, F. J. and F. R. Fisher. 1968. Tests for hierarchical structure in random data sets. *Systematic Zool.* 17: (in press).
15. Sokal, R. R. 1963. The principles and practice of numerical taxonomy. *Taxon* 12: 190-199.
16. Sokal, R. R. 1965. Statistical methods in systematics. *Biol. Rev. (Cambridge)* 40: 337-391.
17. Sokal, R. R. 1967. Recent ideas on the theory of systematics [in Russian]. *Zhurnal Obshchye Biologii* 28: 658-674.
18. Sokal, R. R. 1968. Numerical taxonomy: methodology and recent developments [in Russian]. *Zhurnal Obshchye Biologii* 29: 297-315.

Bibliography of publications in numerical taxonomy supported by the parallel NIH grant.

- Crovello, T. J. 1967. Problems in the use of electronic data processing in biological collections. *Taxon* 16: 481-494.
- Crovello, T. J. 1968. The effect of alteration of technique at two stages in a numerical taxonomic study. *Univ. Kans. Sci. Bull.* 47: 761-786.
- Fisher, D. R. 1968. A study of faunal resemblance using numerical taxonomy and factor analysis. *Syst. Zool.* 17: 48-63.
- Fujii, K. 1968. Numerical taxonomy of ecological characteristics and the niche concept. *Systematic Zool.* : (in press).
- Hendrickson, J. A., Jr. and R. R. Sokal. 1967. A numerical taxonomic study of the genus Psorophora (Diptera: Culicidae). *Ann. Entomol. Soc. Amer.*, 61: 385-392.
- Rohlf, F. J. and R. R. Sokal. 1967. Taxonomic structure from randomly and systematically scanned biological images. *Systematic Zool.* 16: 246-260.

- Sokal, R. R. 1967. A review of clustering methods in biological taxonomy. In Lorr, M. A. (ed.), Proceeding Conference on Cluster Analysis of Multivariate Data. Bur. Soc. Sci. Res., Inc. pp. 3.01-3.10.
- Sokal, R. R. and J. H. Camin. 1965. The two taxonomies: Areas of agreement and conflict. *Systematic Zool.* 14: 176-195.
- Sokal, R. R., J. H. Camin, F. J. Rohlf and P. H. A. Sneath. 1965. Numerical taxonomy: some points of view. *Systematic Zool.* 14: 237-243.
- Sokal, R.R. and F. J. Rohlf. 1966. Random scanning of taxonomic characters. *Nature*, 210: 461-462.
- Sokal, R.R. and P. H. A. Sneath. 1966. Efficiency in taxonomy. *Taxon* 15: 1-21.

(5) Proposed research if support is continued.

A. CONTINUATION AND COMPLETION OF SEVERAL PROJECTS LISTED IN SECTION (3).

It will have been noted that several of the projects discussed under previous research have not been entirely completed. These include projects where observational and experimental work is over, the analysis has been carried out, and only the manuscript must be prepared. Projects 5, 13, 16 and 20 fall in this category. In others (Projects 6, 11 and 19) the analysis must still be completed, and yet others (Projects 3, 17 and 21) are continuing projects where further experiments or observations are planned. Some of these undertakings will be accomplished before the termination of the present grant, but in a continuing enterprise, such as our numerical taxonomy program, it is inevitable that a few of these projects will carry over into the new grant period. This is so not only because it is impossible to apportion research exactly into segments corresponding to grant periods, but also because we know from past experience that, with reasonable frequency, new ideas and opportunities crop up which tempt us to pursue them and upset previous time schedules for the various projects.

In the first place we therefore request support to continue and complete the projects enumerated above and described in detail in Section (3).

B. A GENERAL RESEARCH PROGRAM IN NUMERICAL TAXONOMY AND RELATED FIELDS

In the paragraphs below we outline our proposed general approaches to research in the field of numerical taxonomy and in certain related subject areas. Significant overlap will be found between the research proposed in this part of our application and that portion of a grant renewal application by Drs. Sokal and Rohlf to the National Institutes of Health, concerned with numerical taxonomy. As pointed out in Section (11), the NIH grant "Research in Statistical Biology" covers also four areas quite apart from numerical taxonomy. The overlap in the two proposals is intentional. The research proposed below would probably require a five-year period and full funding on a scale requested in the budgets submitted to either NIH or NSF. It appears to us to be more likely, however, that both agencies will fund this work partially. In that case,

as individual projects are initiated and delimited, they will be identified with one or the other granting agency. This will be done roughly in proportion to the support received for this work from these sources. Often the subject matter of the research will suggest which agency shall be credited with the support. Thus in the past numerical classifications of mosquitoes and roaches has been funded by NIH; cladistic studies by NSF. Only the development of general taxonomic computer programs will be jointly acknowledged to both agencies. The proposed research in part A above will be solely attributed to this NSF project.

The following four areas will receive our attention: (1) The further development of general systematic theory; (2) Improvement of the techniques of numerical taxonomy; (3) Tests of the various hypotheses underlying the theory of numerical taxonomy; and (4) The application of numerical taxonomy to various exemplary groups of organisms.

1. The further development of general systematic theory.

Underlying all of the technical work outlined below there must be a solid foundation of general systematic theory, integrating theoretical and practical aspects of the conventional and numerical schools. These endeavors have accompanied our efforts in the past (as can be seen from the co-principal investigators' bibliography) and must continue to do so in the future. Development of the theory must range over the entire field of systematics from homology to the definition of taxonomic categories, from the species concept to paleontological research, and from nomenclature to numerical cladistics. In connection with the preparation of the revised edition of "Principles of Numerical Taxonomy", all of the relevant areas will be considered and worked on during the forthcoming grant period but it is impossible to predict in which area major progress will be made. In the past this has often been the consequence of an accidental discovery or of some chance conversation with colleagues or graduate students. The following sample of topics may be considered representative of proposed projects. Work on all of these is currently underway to varying degrees.

a. An appraisal to what degree numerical taxonomy can be practiced at present as a routine classificatory procedure. This will involve a discussion and evaluation of automatic procedures for character recording and coding.

b. An analysis of the interrelation between phenetic and cladistic evidence at the species level as well as at higher taxonomic levels. This is a critical issue in the continuing controversy between "phylogenetic" systematics and numerical taxonomy.

c. Some recently tested programs have proven to be quite efficient in determining the contents and exact limits of taxa thought to be a biological species. Techniques are already envisioned by which existing programs could be modified to identify "species" clusters with high probability of being

"biologically" correct. Future research along these lines will consist of developing the modified programs and testing them in a variety of "different" groups.

d. An investigation is planned of theoretical properties of different criteria of optimality used in choosing classifications, with the objective of finding a theoretically most preferable optimality criterion. At present such a criterion is conceived of as being chosen on the basis of finding the classification with the highest information content and the best predictive ability of the classification, as well as the greatest robustness of the classification under the addition of new information. The investigation of optimality under these considerations is intended to cut across the current dichotomy between phenetic and cladistic approaches to numerical taxonomy.

e. The development of numerical measures for phenetic diversity. This field is little explored so far, yet has great importance in the present theoretical framework of population biology. It directly leads to the next topic.

f. The definition of the ecological niche in terms of the phenetic dimensions of its inhabitants.

g. The development of new models and patterns for taxonomic categories and for nomenclature. As a result of findings to date it has become obvious that present day categories and nomenclature are inadequate to represent taxonomic relationships. New forms compatible with recent findings and with electronic data processing must be found.

2. Improvement of the techniques of numerical taxonomy.

We plan to continue our development and evaluation of methods for summarizing and displaying multivariate relationships. These methods include those of cluster analysis, multidimensional scaling, factor analysis, three-dimensional models, stereograms, networks, and some considerations of nomenclatural problems connected with these classificatory procedures.

Much additional work in cluster analysis is desirable. Two sorts of activities will be especially emphasized. The first is an extensive evaluation of cluster analytical techniques with a view towards developing a standard terminology and set of descriptors to be used in comparing various methods. We hope that this will also lead to a useful classification of the different types of cluster analyses. At present it is very difficult to evaluate the utility of the very large number of cluster analyses which have been proposed since they are usually described using a different set of terms (often subject-matter oriented) and are applied to new sets of data. Relatively little work has been done to compare different cluster analyses done on the same set of data. In view

of the importance of cluster analysis to many areas of biology, medicine, psychology, sociology and economics we feel that the preparation of such a standard terminology is a very important task.

The second project will evaluate the usefulness of factor analytical procedures in taxonomy (both with and without rotation to simple structure). Much work has been done recently investigating the usefulness of three-dimensional models and stereograms to represent projections of OTU's on the first three principal components, resulting from an analysis based on correlations among the characters. This method serves as a useful complement to cluster analysis since it enables one to obtain a visual impression of the structure of the data without first having to assume that clusters are present. A disadvantage of the technique is that it does not yield a classification in the conventional sense. Preliminary work (reported in Rohlf, 1967) will be continued to investigate the usefulness of rotation of factors to simple structure. In particular it is of interest to investigate whether the equal weighting of rotated factors will compensate for biases in sampling of characters from various dimensions of variation in the organisms.

Possible modes of character weighting fall into two categories. Data analysis prior to any actual classification of the group in a study can evaluate characters on the basis of predictions from evolutionary theory and provide weighting factors for initial use in the process of classifying. Theoretical studies of the best forms of this type of weighting are planned. The second type of weighting is that done by the classification algorithm itself in the process of forming a classification. Such weighting can provide evaluative information about characters that cannot be predicted before a classification is formed. Some such weighting procedures have already been programmed; future studies will be devoted to improving the efficiency of these techniques.

Other methodological approaches include techniques for constructing various sorts of networks, for example Prim networks, which will be investigated as alternative methods for displaying some of the structure present in the data. The objective is to develop algorithms that can form most parsimonious networks without assuming discreteness of character coding, directionality or irreversibility of characters in evolution, or the nature of the primitive form for a group. It is also desirable to form the algorithms in such a way that they can be employed on arbitrarily-weighted characters. Each of these aims has been accomplished in some programs already in use, but more effective algorithms are theoretically possible, and seem desirable.

We plan to investigate various schemes for the verbal description and discussion of phenetic relationships as revealed by some of the techniques described above. The conventional binominal nomenclature is based on the concept of non-overlapping systems of nested clusters. Evidence is accumulating that such a system is a simplistic view of the complexities inherent in taxonomic relationships. Methods need to be considered which allow the description and communication of more complex systems of phenetic relationships.

3. Tests of the various hypotheses underlying the theory of numerical taxonomy.

We would like to test further several of the assumptions of numerical taxonomy propounded by Sokal and Sneath. Specifically, we would like to test the hypothesis of the matches asymptote. This states that as the size of a set of characters increases, an asymptotic similarity value is reached and that beyond that point the increase of information obtained by sampling more characters becomes quite negligible; hence search for additional characters becomes relatively unprofitable. In the bee data of Michener and Sokal and in those on the orders of insects (Sokal, Michener and Huber) information for such a study has already been collected. The analysis will therefore comprise mainly computer processing and interpretation.

An important aspect of numerical taxonomic work to which we plan to devote an appreciable amount of time is the subject of the magnitude and effect of correlation among characters. So far we have little information as to the extent to which characters will be correlated in a data matrix of taxonomic units of intermediate rank, as for species and genera. Such correlations can sometimes be quite extensive and hence may introduce considerable redundancy of information employed in numerical taxonomic work. It would be desirable to study the effect of removal of such redundancy by one of the various methods of multivariate statistical analysis suitable for such problems, as for example factor analysis. An application of factor analysis to geographic variation data by Sokal (1962) is an analogous procedure. A large number of data matrices are available from studies already completed or currently underway on which these investigations may be carried out.

Models of parallelism and convergence should be investigated. We hope to take several groups of organisms which are apparently convergent for several notable adaptations and then try to estimate the quantitative degree of convergence between these forms. By comparing the similarities of these so-called convergent forms and of their supposedly close phylogenetic relatives with each other, a measure of the amount of convergence should be possible. Choice of the particular taxonomic groups will have to depend on personnel available at the time. An earlier study by Boyce on gerbils is awaiting publication (see Section (3)D above).

In this connection it is of interest to simulate various models of evolution on the computer to generate large amounts of data with known cladistic relationships. This will be valuable for two reasons. First, phenetic relationships can be generated, which result from an evolutionary process rather than from simpler models in which data are random samples from a multivariate normal distribution as done by Rohlf and Fisher (1968; see Section (3)B above). Experiments can be made in which the parameters in the evolutionary

model are varied in order to see which methods of cluster analysis yield the best representation for various models. The model might be populations performing a random walk in a p -dimensional phenetic hyperspace. Variance-covariance matrices of these populations would be computed and the multivariate means clustered. Populations will be allowed to split into two independently evolving groups or to become extinct. It is quite likely that certain procedures of cluster analysis will be more efficient in reflecting the data of various structures arising from different patterns of evolution. Such a program would also be useful to estimate the efficiency of various numerical cladistic procedures in estimating the true cladistic sequences. A large number of pseudo-phylogenies will be generated and their end points submitted to taxonomic programs with the object of estimating the probability with which the programs can deduce the correct cladistic relationships of organisms. It is very important that such studies be done in order to evaluate the usefulness of procedures which have been proposed for numerical cladistics, since there are few fossil data complete enough to serve as checks on the accuracy of these methods.

These activities require access to a large-scale computing system for the heavy computational load involved. Equally important is convenient access to the computer during the time in which one is developing new algorithms and when one is performing experiments studying the effects of variations in various parameters in the models studies. It is therefore expected that both the teletypewriter and the Cullen-Fried teleputer system will be used extensively.

4. Application of numerical taxonomy to various groups of organisms.

One of the urgent problems in numerical taxonomy is to enlarge the material available so far for numerical taxonomic study by including several new groups of plants and animals as material for such studies. Through such work we could obtain larger suites of characters and sets of taxa, which should enable us to reach better founded generalizations on the various points to be investigated. An in-depth study of a few taxonomic groups also seem to be very important.

It would be difficult and injudicious to predict the exact groups on which such work will be done. This will depend largely on the interests and aptitudes of research associates and graduate students to be employed. Special numerical taxonomy, more than most fields of science, requires a certain degree of rapport between the investigator and the group to be studied. Thus it would be unwise to urge a man interested in lice or other ectoparasites to work on beetles, for example. The interests of two of the co-principal investigators have been on insects (aphids and mosquitoes) while the third has worked with and continues to be interested in cyprinodontoid fishes. Since no extensive systematic collections are located at Stony Brook, we expect to work out arrangements with nearby

institutions with large taxonomic holdings, such as The American Museum, or the Smithsonian, whereby our students and research associates could carry out the data-gathering plans of their work at these places.

So far most numerical taxonomic studies have either dealt with a very restricted taxon or else were based on only a very sparse sampling of species from a variety of taxa. It would be of great interest in itself to study phenetic diversity in a circumscribed group for the entire world, as well as of a somewhat broader taxonomic group for a given region. In this manner, one could eventually compare the phenetic diversity in groups of organisms in various geographic regions, for example, North and South America or Africa. One could attempt to correlate this with size and ecological diversity of the areas studied.

C. SIGNIFICANCE OF THIS RESEARCH

One of the dramatic developments in biology in this decade has been the rapid emergence of the newly integrated field of population biology. Just as advances in microbiology are largely of a chemical nature, so in the field of population biology many of the new concepts are formulated mathematically. It is also characteristic that the so-called "new mathematics" is largely involved in this work. The application of various methods of biostatistics, cybernetics, information theory, linear programming, the theory of games, and other methods, many of them intimately tied up with the development of computers, is currently revolutionizing many of our previously held views on the nature of the interactions among organisms in population dynamics and evolution. Closely connected with this development are the technical advances being made in automatic data recording, processing, and storing, so that procedures and conventions in fields, such as taxonomy, which date back to the scientific capabilities of the 19th century are becoming obsolete. A new taxonomy will play an important role in the new population biology which is currently being founded and for this reason alone the research above is of importance. However, it is self-evident that biologists of all kinds will continue to need a systematic foundation on which to base their researches. Modern population biology needs a taxonomy which will fit in with its general concepts; and this present-day taxonomy does not do. The current re-evaluation of taxonomy is valuable also in subjecting present-day systematic theory to a thorough-going review. Even if the answers proposed by numerical taxonomists are not always the correct ones, the challenge they have put forward to the established views has caused orthodox taxonomists to re-examine their own principles and this should lead to an improvement in systematics.

Since the original application for this grant was written in 1961, numerical taxonomy has progressed from a programmatic proposal for taxonomy to a widely accepted and practiced aspect of systematic work, whose impact is being felt in ever increasing measure by systematists in all fields. Admittedly, some of the early hopes for numerical taxonomy as the infallible panacea for taxonomic work have not been

realized. New problems have been uncovered, leading in time to new theoretical formulations and to new methodologies. But to an ever increasing extent the measurement of phenetic similarity and diversity has become a valuable approach in evolutionary studies, and numerical cladistic approaches promise to follow suit. The successes of numerical taxonomy in ordinary taxonomic work such as monographs can be easily documented and the literature has become so voluminous that Sokal and Sneath, authors of the "Principles", have had great difficulty in keeping up with it for the second edition of their book.

Beyond biological systematics in the strict sense, numerical taxonomy has contributed to the methodology of numerous fields: ecology - classification of habitats, leading to an understanding of common factors determining the composition of plant and animal communities; medicine - disease classification, providing objective measures for common symptomatology and leading to efficient keys for automatic diagnosis; psychology - classification of personality types and mental disorders; pedology - soil classifications; and applications to machine parts, geological strata, legislatures, political systems, archeological artifacts and many more.

(6) Facilities.

For numerical taxonomic research, general office and laboratory space is available in the Biological Sciences Department. A special laboratory (about 700 square feet) is being set up to house desk calculators, remote computer terminals, IBM card punch and IBM card file cabinets.

The facilities of the University's computation center are available and essential for this project. The computer is an IBM 360/67 but running as a 65, with 512 K core memory, 6 tape drives, an IBM 2314 disk, and 2 IBM 1100 lpm printers. From our previous experience we expect these facilities to be adequate for our computing needs for the next several years. The center has plans for short and long range improvements. By September 1969 it intends to have 30 IBM 2740 remote terminals. Of these at least one will be available to this project to be used for the development of computer programs for numerical taxonomy and for the calling up of programs for executing sets of data.

A Culler-Fried remote terminal useful for graphic display of mathematical functions will also be available.

(7) Personnel.

Co-Principal Investigator: James S. Farris

1. Academic biography: Born Baltimore, Maryland, 1942
 B. S. University of Massachusetts, 1963.
 M.S. University of Michigan, 1965.
 C. Ph. University of Michigan, 1967.
 Ph. D. University of Michigan, 1968.

2. Positions.

Teaching Assistant (Zoology), University of Massachusetts 1962-63.
 Research Assistant (Zoology), University of Michigan 1963-64.
 Teaching Fellow (Zoology), University of Michigan 1964-65.
 NSF Trainee (Zoology), University of Michigan 1965-67.
 Statistical Consultant, University of Michigan Museum of Zoology 1965-67.
 NSF Pre-doctoral Fellow (Zoology), University of Michigan 1967-68.
 Assistant Professor of Biological Sciences, State University of New York at Stony Brook, 1968-

3. Publications.

James S. Farris has authored eight papers, seven of which are relevant to the present application and are listed in the joint bibliography below. He is working on a book on quantitative phylogenetic systematics.

Co-Principal Investigator: F. James Rohlf

1. Academic Biography: Born Blythe, California, 1936.
 A. B. (Zoology) San Diego State College, 1958.
 Ph. D. (Entomology) The University of Kansas, 1962.

2. Positions.

Research Associate in Entomology, The University of Kansas, summers 1962, 1965.
 Assistant Professor of Biology, University of California, Santa Barbara, 1962-1966.
 Visiting Assistant Professor of Entomology, The University of Kansas, 1965.
 Associate Professor of Statistical Biology, The University of Kansas, 1966-1969.
 Associate Professor of Biological Sciences, State University of New York at Stony Brook, 1968-

3. Publications.

F. James Rohlf has published 26 papers in several fields of biological research, 18 of which are relevant to the present project and are listed in the joint bibliography below. A textbook on biometry and a set of statistical tables, in co-authorship with Robert R. Sokal, will be published in February 1969.

Co-Principal Investigator: Robert R. Sokal

1. Academic Biography: Born Vienna, Austria, 1926 (U.S. Citizen).
 B. S. (Biology) St. John's University, Shanghai, China, 1947.
 Ph. D. (Zoology) University of Chicago, 1952.

2. Positions.

Assistant in Zoology, University of Chicago, 1950-1951.
 Instructor in Entomology, The University of Kansas, 1951-1953.
 Assistant Professor of Entomology, The University of Kansas,
 1953-1958.
 Associate Professor of Entomology, The University of Kansas,
 1958-1961.
 Professor of Statistical Biology, The University of Kansas, 1961-1969.
 Professor of Biological Sciences, State University of New York at
 Stony Brook, 1968-
 Senior Postdoctoral Fellow, NSF, University College, London, 1959-1960.
 Fulbright Visiting Professor, Hebrew and Tel Aviv Universities, Israel,
 1963-1964.
 NIH Career Investigator, 1964-1969.

3. Publications.

Robert R. Sokal has published 64 papers in several fields of biological research, as well as one book on the principles of numerical taxonomy. Twenty-eight of these publications are directly relevant to the present project and are listed in the joint bibliography below. A textbook of biometry and a set of statistical tables in co-authorship with F. James Rohlf will be published in February 1969. A revision of "Principles of Numerical Taxonomy" with P. H. A. Sneath is well underway.

Systems Programmer: Gerald P. Crow

1. Academic Biography: Born Clarksdale, Miss., 1943.
 B. S. (Chemistry) Kansas State University, 1966.

2. Positions.

Programming Consultant, Computation Center, The University of Kansas,
 1966-1968.
 Systems Programmer, Department of Entomology, The University of
 Kansas, 1968-

Joint Bibliography of the Co-Principal Investigators Pertinent to this Application.

- Basford, N. L., J. E. Butler, C. A. Leone and F. J. Rohlf. 1968. Immunologic comparisons of selected coleoptera with analysis of relationships using numerical taxonomic methods. *Systematic Zool.* 17: (in press).
- Camin, J. H. and R. R. Sokal. 1965. A method for deducing branching sequences in phylogeny. *Evolution* 19: 311-326.
- Farris, J. S. 1966. Estimation of conservatism of characters by constancy within biological populations. *Evolution* 10: 587-591.

- Farris, J. S. 1967. The meaning of relationship and taxonomic procedure. *Systematic Zool.* 16: 44-51.
- Farris, J. S. 1967. Definitions of Taxa. *Systematic Zool.* 16: 174-175.
- Farris, J. S. 1967. Comment on psychologism. *Systematic Zool.* 16: 345-347.
- Farris, J. S. 1968. Categorical ranks and evolutionary taxa in numerical taxonomy. *Systematic Zool.* 17: 151-159.
- Farris, J. S. 1968. Significance testing and confidence intervals for fixed mortality rates. *Ecology*. (In press).
- Farris, J. S. 1969. On the use of most parsimonious trees in evolutionary taxonomy. *Evolution*. (in press).
- Farris, J. S. and A. G. Kluge. 1969. On the phylogeny of anurans. *Systematic Zool.* (in press).
- Fisher, D. R. and F. J. Rohlf. 1969. Robustness of numerical taxonomic methods to errors in homology. *Systematic Zool.* 18: (in press).
- Michener, C. D. and R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11: 130-160.
- Michener, C. D. and R. R. Sokal. 1966. Two tests of the hypothesis of nonspecificity in the *Hoplitis* complex (Hymenoptera: Megachilidae). *Ann. Entomol. Soc. Amer.* 59: 1211-1217.
- Rohlf, F. J. 1961. Numerical taxonomic study of the genus *Aedes* of America north of Mexico (Abstract). *Proc. N. Cent. Branch E.S.A.* 16: 12-13.
- Rohlf, F. J. 1963. Congruence of larval and adult classifications in *Aedes* (Diptera: Culicidae). *Systematic Zool.* 12: 97-117.
- Rohlf, F. J. 1963. Classification of *Aedes* by numerical taxonomic methods (Diptera: Culicidae). *Ann. Entom. Soc. Amer.* 56: 798-804.
- Rohlf, F. J. 1964. Methods for checking the results of numerical taxonomic studies. *Systematic Zool.* 13: 102-104.
- Rohlf, F. J. 1965. Character Correlations in numerical taxonomy. *Proc. 12th Inter. Cong. Ent.* pp. 109-110.
- Rohlf, F. J. 1965. Multivariate methods in taxonomy. In *Proceedings of the IBM Scientific Computing Symposium on Statistics*, October 21-23, 1963. IBM, White Plains, New York. pp. 3-14.
- Rohlf, F. J. 1965. A randomization test of the nonspecificity hypothesis in numerical taxonomy. *Taxon* 14: 262-267.

- Rohlf, F. J. 1967. Correlated characters in numerical taxonomy. *Systematic Zool.* 16: 109-126.
- Rohlf, F. J. 1968. Stereograms in numerical taxonomy. *Systematic Zool.* 17: 246-255.
- Rohlf, F. J. and D. R. Fisher. 1968. Tests for hierarchical structure in random data sets. *Systematic Zool.* 17: (in press).
- Rohlf, F. J. and R. R. Sokal. 1962. The description of taxonomic relationships by factor analysis. *Systematic Zool.* 11: 1-16.
- Rohlf, F. J. and R. R. Sokal. 1965. Coefficients of correlation and distance in numerical taxonomy. *Univ. Kans. Sci. Bull.* 45: 3-27.
- Rohlf, F. J. and R. R. Sokal. 1967. Taxonomic structure from randomly and systematically scanned biological images. *Systematic Zool.* 16: 246-260.
- Sneath, P. H. A. and R. R. Sokal. 1962. Numerical taxonomy. *Nature* 193: 855-860.
- Sokal, R. R. 1958. Quantification of systematic relationships and of phylogenetic trends. *Proc. 10th Inter. Cong. Entom.* 1: 409-415.
- Sokal, R. R. 1959. Comments on quantitative systematics. *Evolution* 11: 420-423.
- Sokal, R. R. 1959. Die Grundlagen der numerischen Taxonomie. *Proc. 11th Inter. Cong. Entom.* 1: 7-12.
- Sokal, R. R. 1961. Distance as a measure of taxonomic similarity. *Systematic Zool.* 10: 70-79.
- Sokal, R. R. 1962. Typology and empiricism in taxonomy. *J. Theor. Biol.* 3: 230-267.
- Sokal, R. R. 1963. The principles and practice of numerical taxonomy. *Taxon* 12: 190-199.
- Sokal, R. R. 1964. The future systematics. In C. A. Leone [ed.], *Taxonomic Biochemistry and Serology*, Ronald Press, New York. pp. 33-48.
- Sokal, R. R. 1964. Numerical taxonomy and disease classification. In J. A. Jacques [ed.], *The Diagnostic Process*. U. Mich. pp. 51-68.
- Sokal, R. R. 1965. Current progress in numerical taxonomy. *Proc. 12th Inter. Cong. Ent. Entomol.* p. 108.
- Sokal, R. R. 1965. Statistical methods in systematics. *Biol. Rev. (Cambridge)* 40: 337-391.
- Sokal, R. R. 1966. Numerical Taxonomy. *Scientific American* 215: 106-116.

- Sokal, R. R. 1967. A review of clustering methods in biological taxonomy
in M. Lohr and S. B. Iyerly [eds.], Some Critical Issues and Problems
in Cluster Analysis. Bur. Soc. Sci. Res., Inc. pp. 3.01-3.10.
- Sokal, R. R. 1967. Recent ideas on the theory of systematics [in Russian].
Zhurnal Obzhenye Biologii 28: 658-674.
- Sokal, R. R. 1968. Numerical taxonomy: methodology and recent developments
[in Russian]. Zhurnal Obzhenye Biologii 29: 297-315.
- Sokal, R. R. and J. H. Camin. 1965. The two taxonomies: areas of agreement
and conflict. Systematic Zool. 14: 176-195.
- Sokal, R. R., J. H. Camin, F. J. Rohlf, and P. H. A. Sneath. 1965. Numerical
taxonomy: some points of view. Systematic Zool. 14: 237-243.
- Sokal, R. R. and C. D. Michener. 1958. A statistical method for evaluating
systematic relationships. Univ. Kans. Sci. Bull. 38: 1409-1438.
- Sokal, R. R. and C. D. Michener. 1967. The effects of different numerical
techniques on the phenetic classification of bees of the Hoplitis complex
(Megachilidae). Proc. Linn. Soc., Lond. 178: 59-74.
- Sokal, R. R. and F. J. Rohlf. 1962. The comparison of dendrograms by objective
methods. Taxon 11: 33-40.
- Sokal, R. R. and F. J. Rohlf. 1966. Random scanning of taxonomic characters.
Nature 210: 461-462.
- Sokal, R. R. and P. H. A. Sneath. 1963. The principles of numerical taxonomy.
W. H. Freeman, San Francisco. 359 pp.
- Sokal, R. R. and P. H. A. Sneath. 1966. Efficiency in taxonomy. Taxon 15:
1-24.

	<u>Requested from NSF</u>	<u>SUNY Contribution</u>	<u>Total Cost</u>
<u>PERSONNEL</u>			
Co-principal investigators:			
J. S. Farris, Assistant Professor (Approx. 20% academic year) (100% of time, summer, 2½ months)	None 3,092	650*	650 3,092
F. J. Rohlf, Associate Professor (Approx. 20% academic year) (100% of time, summer, 2½ months)	None 5,139	1,150*	1,150 5,139
R. R. Sokal, Professor (Approx. 15% academic year) (50% of time, summer, 2½ months)	None 3,611	1,600*	1,600 3,611
Systems Programmer	9,000		9,000
3 Graduate Research Assistants	12,000		12,000
1 Key punch Operator	5,000		5,000
1 Technician (100%, 12 months)	6,000		6,000
1 Secretary	5,500		5,500
Student Help (hourly)	1,000		1,000
SALARY SUBTOTAL	50,342	3,400	53,742
<u>Fringe Benefits</u>			
13.8% of salaries and wages for Research Foundation Employees	6,947	469	7,416
PERSONNEL TOTAL	57,289	3,869	61,158
<u>PERMANENT EQUIPMENT</u>			
4 disk packs	2,000		2,000
Computer printout holders	200		200
PERMANENT EQUIPMENT TOTAL	2,200		2,200
<u>CONSUMABLE SUPPLIES</u>			
(Stationary, photo & art supplies, Microcopy supplies)	750		750
<u>PUBLICATION COSTS</u>	800		800
<u>TRAVEL</u> to scientific meetings for 3 investigators	1,200		1,200
<u>OTHER:</u> Postage and Communication	200		200
<u>COMPUTER CHARGES</u>	10,000		10,000
SUBTOTAL (Direct Costs)	72,439	3,869	76,308
<u>INDIRECT COSTS</u> (44.9% of direct costs excl. computer charges)	28,035	1,737	29,772
TOTAL BUDGET	100,474	5,606	106,080

Proposed Budget

For Year 1971

	<u>Requested from NSF</u>	<u>SUNY Contribution</u>	<u>Total Cost</u>
<u>PERSONNEL</u>			
Co-principal investigators:			
J. S. Farris, Assistant Professor (Approx. 20% academic year) (100% of time, summer, 2 months)	None 3,278	700*	700 3,278
F. J. Rohlf, Associate Professor (Approx. 20% academic year) (100% of time, summer, 2 months)	None 5,447	1,200*	1,200 5,447
R. R. Sokal, Professor (Approx. 15% academic year) (50% of time, summer, 2 months)	None 3,828	1,700*	1,700 3,828
Systems Programmer	9,600		9,600
3 Graduate Research Assistants	12,600		12,600
1 Keypunch Operator	5,350		5,350
1 Technician (100%, 12 months)	6,425		6,425
1 Secretary	5,885		5,885
Student Help (hourly)	1,050		1,050
SALARY SUBTOTAL	53,463	3,600	57,063
<u>Fringe Benefits</u>			
13.8% of salaries and wages for Research Foundation Employees	7,378	497	7,875
PERSONNEL TOTAL	60,841	4,097	64,938
<u>PERMANENT EQUIPMENT</u>			
	None		None
<u>CONSUMABLE SUPPLIES</u>			
(Stationary, photo & art supplies, microcopy supplies)	800		800
<u>PUBLICATION COSTS</u>	900		900
<u>TRAVEL</u> to scientific meetings for 3 investigators	1,200		1,200
<u>OTHER:</u> Postage and Communications	200		200
<u>COMPUTER CHARGES</u>	10,000		10,000
SUBTOTAL (Direct Costs)	73,941	4,097	78,038
<u>INDIRECT COSTS</u> (44.9% of direct costs excl. computer charges)	28,710	1,840	30,550
TOTAL BUDGET	102,651	5,937	108,588

Digitized by the Hunt Institute for Botanical Documentation

*Represent firm commitment by University although it is anticipated the co-principal investigators will spend 15-20% of their time on the project.

2 YEAR SUMMARY

	<u>Requested from NSF</u>	<u>SUNY Contribution</u>	<u>Total Cost</u>
Direct Costs	\$146,380	\$7,966	\$154,346
Indirect Costs	56,745	3,577	60,322
TOTAL COSTS	\$203,125	\$11,543	\$214,668

(9) Explanation of budget.

The budget outline in the previous pages is mostly self-explanatory. Some comments on special aspects of the budget follow.

Summer salaries (2½ months) are provided for the three co-principal investigators. J. S. Farris and F. J. Rohlf will spend full time on this project during the summer, while R. R. Sokal will spend half time on this. All three co-principal investigators will spend 15-20% of their time each month on this project during the rest of the academic year.

Robert R. Sokal is preparing two renewal applications for his other present research support. One of these is to the National Science Foundation to support work in ecological genetics, the other (jointly with Dr. F. J. Rohlf) to the National Institutes of Health to support several research activities collectively called "Research in Statistical Biology". This latter project also is intended to support some of the research in numerical taxonomy.

If both of these other renewal applications are approved, the budget of the present project can be very significantly reduced. Only a quarter time summer salary would be needed for Sokal, and a half time summer salary would suffice for Rohlf. Half time salaries for the entire year would suffice for the systems programmer, the keypunch operator, the secretary, and the technician. Two, rather than three, graduate student salaries would also suffice. Over the two years a reduction of \$73,588 in total costs would be effected.

The employment of a full-time systems programmer is most important to the success of the present proposal. The progress that has been made during the past period is largely due to the fact that we have had the assistance of two able systems programmers (one furnished by the present project and one by NIH) to write the many complex computer programs involved in this research. See Section (3)G for details of the NT-SYS system of computer programs. Mr. Gerald Crow, our present programmer paid from funds of this grant, plans to move to SUNY Stony Brook with us and will be implementing NT-SYS and CLADON at Stony Brook starting 1 February 1969. His work time in the past and as planned under this grant is divided as follows: approximately one fourth of his time is taken up processing research data in numerical taxonomy generated by the co-principal investigators and research assistants. The second quarter of the programmer's time is taken up with activities which can best be described as public service. An increasing number of persons at

CASE Manpower Information: (2 year total)

NSF FUNDED

UNIV FUNDED

Co-principal investigators:

12 summer man-months

6 academic man-months

Graduate res. assistants:

72 calendar man-months

All others secretarial or technical.

The University of Kansas and all over the country are employing numerical taxonomic programs and are in constant need of advice and help in running these programs. Our present programmers spend a considerable amount of time answering replies from other institutions, sending out tapes or card decks, printouts of programs, and aiding researchers and students from various departments of the university who have difficulties with their programming. Our research group has effectively become the center of numerical taxonomic activity in this country, and such activity will steadily increase. We are certain that the demand for such services will continue at our new location. By devoting a portion of the programmer's time to such activities, the granting agency is in effect subsidizing a wide range of taxonomic endeavors by scientists as a whole. We could easily document this with 20 to 30 research studies in biological systematics, ecology, sociology, psychology, political science, history, and even the humanities, which have directly benefitted from the developmental work undertaken under this grant.

The remaining half time of the programmer's work is devoted to the writing of new types of numerical taxonomic programs as these are developed by statistically minded investigators. By concentrating the programming investment at one locality, writing these programs in FORTRAN and then widely distributing them to the community of taxonomists, considerable saving of funds and conservation of effort can be achieved. Our programmers (who are persons with a minimum of a bachelor's degree in mathematics or a related field) also carry out some of the more mathematically oriented research in numerical taxonomy.

Our present systems programmer is Mr. Gerald P. Crow who has a bachelor's degree in chemistry from Kansas State University, and much programming experience. He is highly qualified for this position.

The three graduate students would be employed at half time for ten months and full time for two months. We expect one of these graduate students to be assigned to each of the co-principal investigators but in any one year this distribution need not necessarily be equal.

The lab technician would help with the numerous technical details of the various projects and be assigned to these as the need arises. Her duties would consist of mounting specimens on microscope slides, measuring structures, checking and proofreading data and the like. It is expected that the many activities of the proposed project would more than keep one full time technician occupied.

Keypunch operator. For the very voluminous data handling required by our research we have needed a full time keypunch operator in the past. In her spare time she has helped with proofreading, measurement of specimens, and various other work.

Secretary. Sokal and Rohlf currently swap one secretary with their manuscripts and research correspondence. The addition of Farris to the group would further increase the load. The departmental secretary would not be able to keep up with this work.

In all salaries we have allowed for expected increases in pay scales.

Permanent equipment. The four disk packs for the 2341 IBM disk are needed to serve for storage of our voluminous data and large system of programs. We have far too many data to handle these by cards and disks are more reliable than tapes and permit faster access to programs. We cannot expect the Computation Center to furnish us with data storage facilities of such magnitude.

Computer costs. The computer costs of \$10,000 per year are only a portion of the estimated computer costs of our research. We estimate the total computer charges for our research under this grant will be in excess of \$20,000 per year. The additional computer time will be furnished by the SUNY Stony Brook Computation Center as an added University contribution to this research.

(10) Current research support of co-principal investigators.

Annual support

F. J. Rohlf

NSF On-line computation for statistical analysis
(equipment only, no salaries)
Expires November 1970.

\$64,000
(for 4 years)

R. R. Sokal

NSF Natural selection during population growth
of Tribolium. 1 June 1964 - 31 May 1970.

Out of funds

NIH Research in statistical biology.
Terminates 31 August 1969.

\$61,000

Robert R. Sokal is currently supported by a
Career Development Award of the NIH which will
terminate 31 August 1969.

The following applications are pending

R. R. Sokal and F. J. Rohlf

Research in statistical biology

\$120,000

R. R. Sokal

NSF The ecology of natural selection of
Tribolium and houseflies

\$44,989

(11) Application for related research to other agencies.

No research proposal for the above projects has been made to any other agency.

Some work in numerical taxonomy will be supported by NIH if the grant entitled "Research in Statistical Biology" will be renewed. Only a limited portion of that grant is devoted to numerical taxonomy, the other topics of investigation being geographic variation analysis, factor analysis, ecological genetics, and mathematical modelling. Individual projects in numerical taxonomy will be assigned to one or the other of the grants, as discussed in detail in Section (5) on proposed new research.



THE UNIVERSITY OF KANSAS · LAWRENCE, KANSAS · 66044

DEPARTMENT OF ENTOMOLOGY · SNOW ENTOMOLOGICAL MUSEUM

21 November 1968

Dr. James S. Farris
Department of Biological Sciences
State University of New York at Stony Brook
Stony Brook, New York 11790

Dear Steve:

Enclosed please find the original copy of the triumvirate NSF application. Would you please sign it, ask Ray Jones to do likewise, then make a copy for yourself and carry the original to Dr. Ackerman in the Graduate School office who will handle it thereafter. If there are any problems, please call us.

Sincerely,

Robert R. Sokal
Professor of Statistical Biology

RRS/mka

Enclosure