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The Hunt Institute Archives contains hundreds of thousands of pages of historical content, writing and images, created by thousands of individuals connected to the botanical sciences.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time and social context in which these materials were created, some of the collections contain material that reflect outdated, biased, offensive and possibly violent views, opinions and actions.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 Archival records are historical documents, and the Hunt Institute keeps such records unaltered to maintain their integrity and to foster accountability for the actions and views of the collections'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

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

September 12, 1963

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

Dear Dave:

We've been back here less than a week now and my hay-fever is running true-to-form. It was certainly a pleasant visit to New York, though, and I want to thank you and your charming wife for taking me on such an entertaining (and gastronomically satisfying) tour of Manhattan. It was really a very profitable visit.

Hold onto your hat(s), but it looks as though we may have two taxometric problems for you. Kim Miller is working on a list of code-characters for Tragia, as I had promised, but Wayne Milstead is even further along with a similar line-up for Prenanthes. Within a week I, or they, will be sending you the preliminary lists for your comments. Then we may even need some more cards, but I presume your supply is not yet exhausted.

Who knows, maybe even Margaritaria (as originally promised) will get ready this school year. If the Annals of the Missouri Botanical Garden folds up, I can quit wasting my time struggling with the euphorbs of Panama.

Give my regards to all, including Pierre Dansereau, if he's back.

Sincerely,

Grady

Grady L. Webster
Associate Professor
Biological Sciences
Curator, Kriebel Herbarium

GLW/HW

December 13, 1963

Mr. Wayne L. Milstead
Department of Biological Sciences
Purdue University
Lafayette, Indiana

Dear Mr. Milstead:

Thanks for your letter of the 6th. We have had a little slow down in the operations of the computer here. Unfortunately, they are making shifts in the various components and are expecting that the facilities will be out of operation during most of the month of January. They have already shut down one of the satellites and therefore at this moment are running only on part time.

I fear, therefore, that we will not have anything to say to you during January. It is unfortunate that this is the case. We hope that there will be an opportunity in February to continue, but we have not yet ascertained when they will be able to start again.

Sincerely yours

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDF

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

December 6, 1963

Dr. David J. Rogers
Curator of Quantitative Taxonomy
The New York Botanical Garden
Bronx Park
New York 58, New York

Dear Dr. Rogers,

I must apologize for my delay in answering your last letter. I am at present in the midst of writing my thesis and time flys.

Regarding your suggestion that I spend a few days in New York on the results of the Prenanthes runs, I wonder if it would be possible to delay this until perhaps the semester break in January? As I mentioned, I am trying to get my thesis completed this semester and would perfer to plan a visit after I see this thing pretty well through. I had hoped to be able to include some of the computer information in my thesis, but as it stands I think that I would be better off trying to work it into a paper which I anticipate writing during the Spring. If it is convenient for you, I might be able to get to the Garden on about January 22 or sometime during the week of January 19 to the 25.

As to money, there does not seem to be any available here, but I believe this problem can be overcome in some way. We have not had a chance yet to investigate the availability of departmental funds, which holds some hope.

Sincerely yours,

Wayne L. Milstead

Wayne L. Milstead
Dept. of Bio. Sci.
Purdue University
Lafayette, Indiana

October 1, 1963

Mr. Wayne L. Milstead
Department of Biological Sciences
Purdue University
Lafayette, Indiana

Dear Mr. Milstead:

I hasten to say that because one is not working on Euphorbs we do not consider him to be in the enemy camp. We are just suspicious as to why anybody would find that particular group of plants more interesting than the Euphorbs!

It seems to me that your most important question is the one on the number of states or attributes that might best be used to represent continuous measurements. Unfortunately, we have no specific advice on this particular item. This is a moot question and one that can best be answered by you for your particular problem. The way that I have worked this sort of thing is to make a graph of though of the samples to give me the total spread of the continuous measurements. Then from this graph you can rather subjectively decide at what points to break the graph into separate attributes. You will get some help on this by reading the Andersonian method in his various publications concerning introgressive hybridization. This is of course a trial and error procedure and one which you will have to "manipulate" to get the most satisfactory sets of discrete attributes.

You asked about the form in which information concerning the relative similarity of specimens will appear in final form after processing by the computer. We haven't settled on the final form of the output as yet but would suggest that what we are aiming for is a listing of clustered specimens with their "distances" from their neighboring clusters and a hierarchical ordering of the clusters. In other words, what we would have would be clusters representing generic, sub-generic and species categories. Of course, if one is looking for sub-specific categories, I would suggest that this be a separate investigation because the nature of the sub-specific categories requires considerable more analysis among the taxa within a species than when measuring taxa at the species or a higher level. In other words, the various levels require varying amount of computational activity, and I think that the kinds of ~~varieties~~ that are useful information in the differentiation of sub-specific taxa are a different order ~~of categories~~ than those used for the differentiation of specific or higher categories.

If you are worrying about sub-specific categories, you may wish to think of these problems; but if species or higher levels are your concern, don't worry too much about it. It just so happens that my own group (Manihot) has a fantastic number of cultivars an one species, and this is why I use for sub-specific differentiation an entirely different order from the categories used for specific differentiation.

You are disturbed to some extent by the unconscious weighting of characters due to the use of too small a number of states. I think that this will be overcome by the techniques I mentioned earlier, but we really are not sure what weighting systems are involved and can only try out empirically the various pieces of information as to their value. Perhaps once we have a working model we can begin to experiment with this kind of information to determine just what does provide objective information and which represents subjective differences. I have no answers for you on these questions. I am not sure that anybody has bothered to investigate these important points as yet, but you may be sure that these are problems of ours and we hope to get around to them as soon as possible.

I enclose a reprint of our AIBS article which described the Port-a-Punch technique. You should be able to follow from that. Good luck!

Sincerely

David J. Rogers
Curator of Quantitative Taxonomy

DJR:MDP
Enclosure

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

September 26, 1963

Dear Dr. Rogers,

I might start by giving a general run-down on the genus Prenanthes and that portion of the genus that I am primarily working with. The genus has a world wide distribution almost entirely in the northern hemisphere, largely in Asia and N. America, and has been divided into two subgenera; Euprenanthes in Europe and W. Asia, and Nabalus in N. America and Asia. I am working with the 12 N. American species of subgenus Nabalus, but would like to compare them to Asian and European species as far as the availability of specimens makes this possible. There are about seven subspecific taxa among these 12 species.

At this point I have examined a considerable number of specimens, at least of the N. Americanspecies, for most of the characteristics listed here. By this I mean that the availability of information on some of the charscteristics is somewhat less than for others depending on the condition of specimens, stage of maturation, time and complications involved in obtaining information on some, etc. Therefore this list gives a possible maximum number of characteristics I might be able to finally use. In order to make use of a maximum number of these characteristics in the problem, I intend to select 25 to 30 specimens of each species that are distributed as randomly as possible over the ranges of the species, and where necessary using those specimens that give maximum information. This reduction in sample size, from what I have actually examined in most cases, and a careful but, hopefully, somewhat unbiased selection of "good" specimens should up the amount of information available for the program. In addition I will use what information I have on some Asian and European species, plus a few Lactuca species.

Any criticism and/or suggestions on my use of characteristics and attributes would be helpful. Also, information would be

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

appreciated on the method of actually entering this information on the Port-a-Punch cards, especially when the use of more than one card is necessary, and on the form in which information concerning the relative similarity of specimens will appear in final form after processing by the computer. One additional point that you might comment on is the number of states or attributes that might best be used to represent continuous measurements of such characteristics as phyllary length, petiole length, etc. I am thinking of the possibility of unconscious weighting of some characters due to the use of a small number of states in relation to others with larger numbers of states.

Sincerely;

Wayne L. Milstead

Wayne L. Milstead

After typing this I have read your letter to Dr. Webster of Jan. 15, 1962 in which you describe the actual use of the Port-a-Punch, which clears up a lot in reference to my questions in this letter. Also, I begin to wonder if I might be considered to be in the camp of the enemy since I am working on a Composite rather than the Euphorbs as the rest of his student are.

Leaf length/width ratio
(segments of range)

Leaf pubescence, upper surface
1-main vein only, type 1
2-all veins, type 1
3-entire surface, type 1
4-main vein only, type 2
5-all veins, type 2
6-entire surface, type 2
7-main vein only, type 3
8-all veins, type 3
9-entire surface, type 3
10-absent over all

Leaf pubescence, lower surface
(as for upper surface)

Marginal pubescence
1-type 1
2-type 2
3-type 3
4-absent

Leaf shape
(will use about 15 generalized forms grouped on the basis of observations)

Upper epidermal cell shape, of leaves
1-lobed
2-sinuate
3-near round

Lower epidermal cell shapes, of leaves
(as for upper surface)

Stomate number, upper surface
(segment of range)

Stomate number, lower surface
(segments of range)

Petiole length
(segments of range)

Peduncle length
(Segments of range)

Inner phyllary length
(segments of range)

Inner phyllary number
(segments of range)

Inner phyllary pubescence
1-midrib only, type 1
2-entire surface, type 1
3- midrib only, type 2
4-entire surface, type 2
5-midrib only, type 3
6-entire surface, type 3
7-pubescence absent

Inner phyllary papillation
1-none
2-base only
3-tips only
4-tips and base
5-scattered over all
6-dense over all

Inner phyllary papillation type
1-black
2-waxy
3-none

Outer phyllary length
(segments of range)

Outer phyllary number
(segments of range)

Length of inner phyllaries to pappus
1-shorter than pappus
2-equal to pappus
3-longer than pappus

Inner phyllary color
1-tan-yellow
2-green
3-dark-green
4-red plus green
5-mostly green plus purple
6-mostly purple plus green
7-purple
8-dark purple

Flower number per head
(segments of range)

Corolla to corolla limb length ratio
(segments of range)

Flower color
1-green-yellow
2-yellow
3-cream (yellow-white)
4-white
5-pink-white
6-red
7-purple
8-lavender
9-blue

Corolla pubescence
1-unicellular
2-one cell wide
3-multicellular
4-absent

Pollen diameter
(segments of range)

Pappus length
(segments of range)

Pappus color
1-white
2-pale yellow
3-yellow
4-tan
5-cinnimon
6-gold-brown

Stigma branch length
(segments of range)

Achene length
(segments of range)

Vascular bundle number in achene
(segments of range)

Achene crystals, location
1-in embryo only
2-sclerenchymatous cap
3-upper ovary
4-lower ovary
5-embryo plus upper ovary
6-embryo plus lower ovary
7-embryo plus cap
8-none of above

Achene crystals, type
1-druse
2-simple
3-raphid
4-aggregate of simple
5-none of above

Achene surface
1-smooth
2-papillate
3-near pubescent

Apical cap in achene
1-present, sclerenchyma mainly
2-present, parenchyma mainly
3-absent or very small

Plant height
(Segments of range)

Stem color (also location of colors on stem)
1-green
2-green-purple
3-green-purple-green
4-purple-green
5-purple-green-purple
6-purple
7-red-purple

Stem pubescence
(classes established from observations)

Stem pubescence types
1-glabrous
2-puberulent
3-hispidulous
4-hispid
5-none of above

Plant habitat
(established from records and observations)

Inflorescence and branching types
(classes established from observations)

Number lobes per leaf
(segments of range)

Chromosome number (N)
1-8
2-9
3-16
4-other

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

28 March 1963

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

Dear Dave:

Please excuse my long delay in acknowledging the shipment of cards + portapunch which you sent. Things seem to have closed in on me after that week you were here.

I had intended to include a copy of the Margaritaria program for your perusal, but it is still not ironed out. I will let you a look at it before we spoil too many cards, however.

We have an interesting situation developing here this fall. Someone in the department is getting a small lab computer which will be available in this building. If the Margaritaria problem goes promisingly, I may try out something very simple (perhaps data from the taxonomy course) on the lab machine. It would sure be handy if each herbarium were equipped with a computer to grind out the monographs & revisions!

I would like to thank you again for visiting us and giving the seminar. It created quite a bit of interest, and now that Purdue has a 7090 on campus a lot of people are getting computer-happy!

Sincerely,

Grady

Grady L. Webster

February 25, 1963

Dr. Grady Webster
Department of Biology
Purdue University
Lafayette, Indiana

Dear Grady:

I returned safely to New York after a very exciting visit with you.

We are sending under separate cover a portapunch and about 2,000 cards. We hope this will be enough to keep you in business for the time being.

Enclosed herewith is a card you can use to get yourself on the mailing list for Taxometrics. Just fill it in and send it off, and it will be taken care of.

Please give my best regards to your wife and thank her for a very fine dinner.

Sincerely

DJR:MDF
Enclosure

Jan. 15, 1967 3

Dr. Grady Webster
Dept. Biological Sciences
Purdue University
Lafayette, Ind.

Dear Grady:

Your responsiveness to the idea of computer processing of your data leads me to wonder if you'd like to cooperate with us in a trail run of your info in our early efforts to work out some of the bugs in the program. We are still in an experimental stage, and do not consider that we have a program that does all it should do for the taxonomist, nor necessarily does correctly that which we think it should. It will take many tests, on many different groups, before we can have the confidence we would like to have. I hope you will be interested--it won't cost you anything but blood, sweat and tears if you are interested. At this stage of the game, we have enough funds for the computer operation, and under our navy grant, have set it up so that we will try different botanist's data. When and if we ever get out of the experimental stage, then running the computer (or meeting the charges for the computer) will be something else again.

One of the problems we took up early, and have a solution to, is just the question you asked--how to get the raw data into the computer. The two enclosures--one a brochure from IBM explaining their "Port-a-Punch" devise, and two, our own IBM card designed to be used with the Port-a-Punch--will be needed to understand the method.

The data we use consists of the properties of the specimens under consideration. These properties may be morphological, anatomical, physiological, cytological, genetic, ecological, geographical, etc. If you are familiar with Anderson's method for setting up his "whisker" diagrams, where the character states of any one characteristic are set up as different classes, then you know how to proceed with our method.

We use the term "characteristic" as a group term which defines the parameters of a related set of properties. These properties we call "attributes". Characteristics are numbered with Roman numerals. The Roman numerals of the characteristics are only mnemonic and organizational aids to assist the investigator in assembling the data. The attributes are numbered with arabic numerals sequentially in groups from one to 70 (each of our punched cards having only seventy spaces per card). If we have 280 attributes, we will have four sets of attributes numbered from 1 to 70. There is no limit on the number of characteristics and attributes we can handle. For example, one of our collaborators has 420 attributes, thus requiring 6 cards. For convenience' sake the attributes may be numberd to indicate which card the information goes on. Card one, with attributes 1 to 70 can be numbered 101 to 170; card two, 201 to 270; card three 301 to 370, etc.

The attributes may be either quantitative or qualitative. The only restriction is that each set of attributes within each characteristic is mutually exclusive. This restriction means that no specimen may be recorded as possessing two or more

attributes within any one characteristic. To this extent we may think of the attributes as having the same requirements as a xx dichotomous key (even though the characteristic or attribute may not be considered as practical in the construction of a key).

Example extracted from my Manihot study:

I Color of stem
101 silver
102 silver-brown
103 brown
104 yellow

Note that the attribute silver brown will provide for those specimens that are both silver and brown and that specimens that are only silver or only brown have their respective attributes.

VIII Median lobe length
311 Less than 14 cm
312 14-17 cm.
313 greater than 17 cm

This is a standard manner of handling quantitative data.

A characteristic may have any number of attributes, but must have at least two. The fine-ness or broad-ness of the attributes is determined by the taxonomic significance of the property.

This concludes the basics of data organization. We have found in practice that the easiest way of assembling the data is simply to type the attribute numbers in sequence under their respective characteristic headings. There is another enclosure that shows what I mean here.

Now, to get this information into the computer. The IBM Port-a-punch, mentioned on page one, enables the investigator to transfer his data conveniently to a form suitable for machine processing. The enclosed card is one that we designed to be used with the port-a-punch. Obviously, we don't have 80 columns, nor do we used all the rows, for several practical reasons. We have a little proprietary interest in the card, though obviously it should be used wherever it will facilitate taxonomic work. I think the advantages are obvious--you can take it in the field with you, carry it around in the lab, have students helping you with the data recording, etc. You could even use it in a library. You merely punch out the appropriate holes with a stylus.

Notice first the extreme right-hand column. This column is to signify to the computer which card of any particular species (or specimen) the data has been punched for. The first card with its seventy numbers will have number one punched out, the second card, number two, etc. The second set of numbers on the right of the card, labelled "Specimen number" is self-explanatory. The four columns allows us up to 9999 specimens.

The enclosed card is card 3 for specimen number 3029. The numbers on the remainder of the card are the actual attribute numbers. The topm row of each of the double rows is to record the presence of the attribute for the respective specimen. We have punched out 4, 15, 29, 44 and 60. The lower row of each of the double rows is only punched out in the event that one or two conditions are true, namely, either the specimen is damaged and we lack data on a characteristic, or else the characteristic is such that some specimens (species, etc.) cannot possibly possess any of the attributes assigned. In the first instance, if the specimen does not

have flowers (it was not collected with themxxx, they have been lost, etc.) we cannot state under the characteristic "flower color" whether they are (attributes) 7, white; 8, blue; or 9, yellow. We have, therefore, punched out 7, 8, and 9 in the lower of the double rows. An example of the second case could be as follows: The species or part of the genus possess petals while the remainder does not. Those that do not have petals can not have an characteristic of color ascribed to them. Consequently, we would, as before, punch out in the lower of the double rows 7, 8, and 9. When the lower row is punched out, the computer will ignore the characteristic in any comparison made between this specimmn, for that characteristic, and the remainder of the specimens.

There are other details we have built into the program so that each attribute is identifiable at all stages of operation and in the final print-out, thus giving you something to tie the information to when the final work is presented. We need not go into details of this sort here.

I think that we have covered the saliant points, but there may be some points that are not as clear to you. One question--I think I remember sending you a memeographed copy of the actual program. Did I? If not, I'll send one along.

We haven't published anything on this method, but we aren't quite ready. So, please consider it as unpublished information, not for duplication.

I hope that you will be able to work with us--not only for our sake, but for good old Euphorbs, too.

Yours,

PURDUE UNIVERSITY
DEPARTMENT OF BIOLOGICAL SCIENCES
LAFAYETTE, INDIANA

January 7, 1963

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

Dear Dave:

Thanks very much for the copy of your paper on computer taxonomy. I think we would definitely be interested in considering one of your programs for processing our data. The details of how you transfer the raw data into the computer still escape me, however. So anything you have on your new system would be appreciated.

During my last trip to the Field Museum, I noted that you have all the specimens of Manihot out on loan. This provides me with a good excuse for sending you duplicates of our summer's collection for determination. The half-dozen sheets will get off in the mail to you within a few weeks.

Sincerely,

Grady

Grady L. Webster
Associate Professor,
Biological Sciences

GLW/fd

P.S. Thanks for checking on Reverchonia.