Uses of Computerized Floristic Data of Indiana for Plant Geography[1]

THEODORE J. CROVELLO and CLIFTON KELLER
Department of Biology
University of Notre Dame, Notre Dame, Indiana 46556

Abstract

The work reported here is an attempt to demonstrate how computers can make data available in floras of more value to plant geography. Deam's *Flora Of Indiana* was used to extract county distribution data on the 81 taxa of the Brassicaceae growing in the State. These data were used to generate county by county checklists, summary statistics of the number of species per county, and two indices of similarity among all pairs of counties. The results of the similarity studies were depicted as geographic contour maps, with each isopleth a level of similarity; as a phenogram; and as a computer generated geographic map of the State in which counties with high similarity with each other are differentially shaded with the same symbols. The pattern revealed clusters of counties that corresponded with known environmental characters in north-south or east west gradients. Four counties appeared unexpectedly dissimilar from others. This seems explainable either by inordinate amounts of collecting there, or by the association (of two) of the counties with prairie communities of Illinois. Computerized analysis of floristic data has potential to permit further insights into the phenomena of plant geography.

In the 33 years since the publication of Deam's (2) *Flora Of Indiana*, systematic biology has experienced an increase of research activity. It has incorporated new kinds of data into its store of information. These include chemosystematic data and research using the scanning electron microscope. Of equal importance is the use of computers to serve as an efficient means to store and to retrieve desired aspects of these voluminous data. In the few years that have elapsed since the beginning of the use of computers for biological information retrieval, we already realize that the accumulation of reliable data is essential to sound progress in plant systematics. For this reason we were anxious to explore ways to unlock those data in printed floras that are the result of diligent, professional, floristic studies.

The purpose of the present work was to analyze a part of Deam's Flora to demonstrate its additional value to plant geography and ecology. While the data have been available since 1940, we had to wait for the computer as the medium through which we could rearrange and analyze it efficiently. Specifically, we concentrated on the data available for the geographic distribution of each species of one family in the State.

Materials and Methods

The materials used were the county distribution maps of the 81 species and varieties of the Brassicaceae (mustard family) recorded by Deam as growing in Indiana, both native and introduced. For this

[1] This project was supported partially by National Science Foundation, Office of Science Information Services, Grant GN-878, to Theodore J. Crovello.

preliminary study no attempt was made to update the distribution maps of Deam with later collections. Our resources did not permit this. The Brassicaceae was chosen because it is the family of greatest interest to the authors and because it contains both native and weedy taxa in Indiana.

Data from each map were accumulated on computer punch cards. These included the coded species name and counties in which Deam had recorded its occurrence. To facilitate data capture, a transparent "mask" containing county identification numbers was placed over each map. The number of each county in which the taxon was collected was then recorded. The data on the cards were then verified for correctness.

Data analysis was accomplished using our own computer programs written in the PLI language. Phenograms, designed to depict the similarity among all 92 counties simultaneously, were calculated using NTSYS, a set of programs developed by F. J. Rohlf and R. R. Sokal. All computations were made on Notre Dame's IBM 370/158 digital computer.

Readers can get a deeper insight into our methods by a brief statement of what each computer program does. For simplicity we name each program in capital letters.

INDIANA 1 takes the distribution data from cards, arranges them in a species by county table, and stores them on magnetic tape for later, efficient use.

INDIANA 2 uses the above data table to produce an outline map with the number of species per county.

INDIANA 3 uses county data to produce an Indiana map wherein each county is shaded (by computer) to represent different values (*e.g.*, number of species per county, or the value of some biogeographic diversity index for each county).

CHECKLIST uses the data of INDIANA 1 to produce county checklists of species recorded from the county. Conversely, it can produce checklists by taxa. Either scientific or common names can be requested.

HUHEEY also uses the data of INDIANA 1 to calculate an average divergence value of each county with each of its neighbors. As given by Huheey (4), the divergence value of counties a and b, $D_{a,b}$ is:

$$D_{a,b} = \frac{N_a + N_b}{N_a + N_b + N_c}$$

where N_a is the number of species in county a but not in county b, and N_b is the converse. N_c is the number of species common to both counties. Hence, if two counties are identical, their D value will be 0.0. If they are absolutely *different* ($N_c = 0$), then their D value will be 1.0.

The numerical taxonomy programs of Rohlf and Sokal, NTSYS, were used to calculate the simple matching coefficient between all pairs of counties, and to create a phenogram using the Unweighted Pair Group Method. Consult Sneath and Sokal (7) for computational details.

Results

It required 1 hour to extract and keypunch the distribution data from five of Deam's maps. Recent changes in procedure allow us now to capture data on 20 maps per hour. To conserve space, we shall not present those results which the reader can visualize easily.

Figure 1, a result of INDIANA 2, is an actual computer printout of a county outline-map of Indiana which contains the number of species of the Brassicaceae collected in each county. Such maps are useful to discover readily which parts of the state have a rich flora for this family. They also can be used to ascertain which counties might be poorly collected to date. For example, Fayette and Union Counties surely have more than one and two species, respectively. The computing cost to calculate such statistics and to print the summary map was under $5.00. Actual computer time was 20 seconds. Similar costs cover the production of shaded county maps by the INDIANA 3 program.

CHECKLIST'S results are simply the inverted files of the basic species by county table. The cost to produce a county checklist for all of the 92 counties of Indiana was $3.00. This also included a second list, one species at a time, of which counties from which it had been recorded.

Another form of output is a base map of Indiana to which has been added the value of Huheey's Divergence Value for each county. Recall that a county absolutely similar with all of its contiguous counties would have a value of 0.0 (no divergence). A county that had no species in common with any of its neighboring counties would have a value of 1.00 (complete divergence). No county has a divergence value lower than 0.6, indicating that a considerable degree of dissimilarity exists among all counties. The greatest divergence values, hence the most likely area for a biogeographic boundary, occur roughly from east to west in the central part of the State. Another transect of high divergence is present in the southwest quarter of the State, running from south to north, before turning westward.

Figure 2 summarizes the relationships between the counties using the simple matching coefficient. It is a summary of the relationships obtained using NTSYS and its resulting phenogram. Since there are so many counties, it is impossible to present the detailed phenogram here. We believe that this is the first time that the results of a phenogram have been summarized by the differential shading of a geographic map. Computing cost to produce these results was under $5.00.

Figure 3 summarizes relationships among the counties according to Huheey's Divergence Index. The Index for each county appears by the county name, rounded off to the nearest tenth of a unit. To indicate directions and rates of change in the State, isopleth lines of equal Divergence Values were fitted and drawn on the printout by eye.

Discussion

Given the results of analyses like those described above, additional biological insights are provided by attempts to correlate environmental

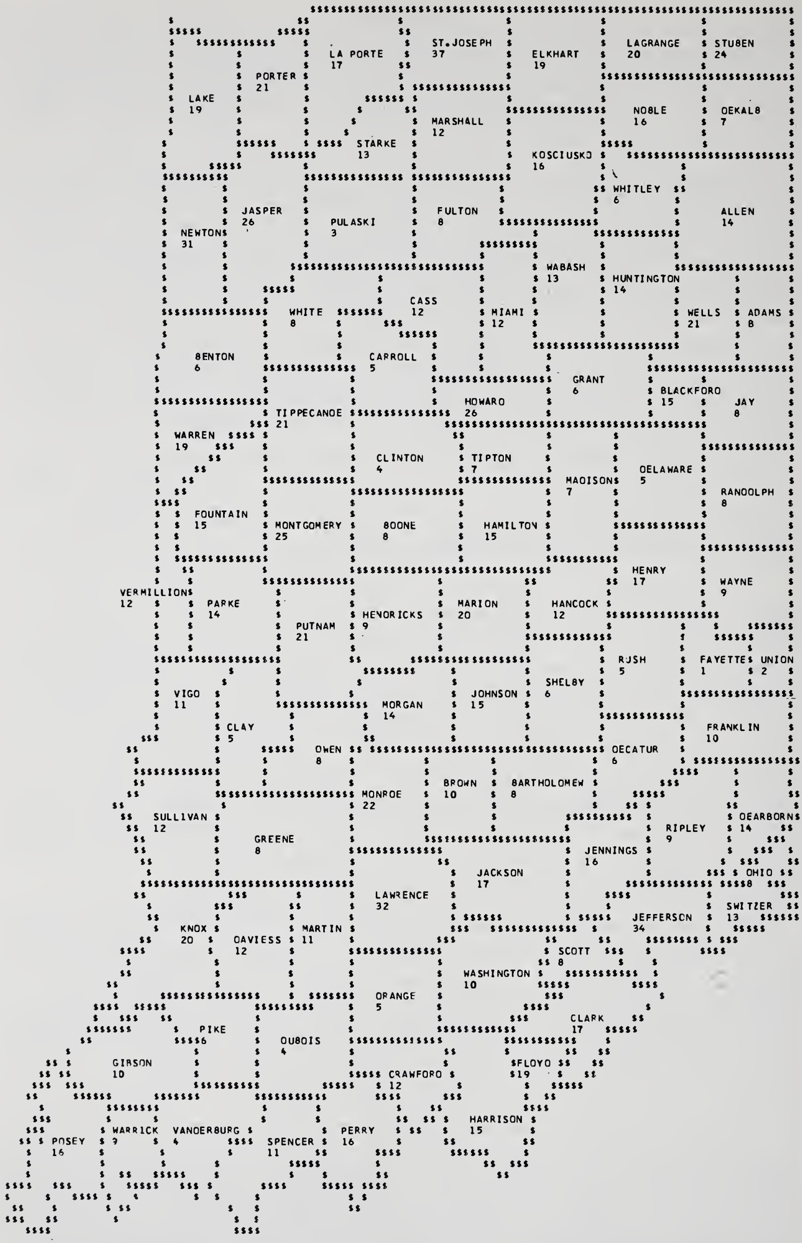

FIGURE 1. *Computer printout of Indiana showing the number of species of the Brassicaceae found in each county.*

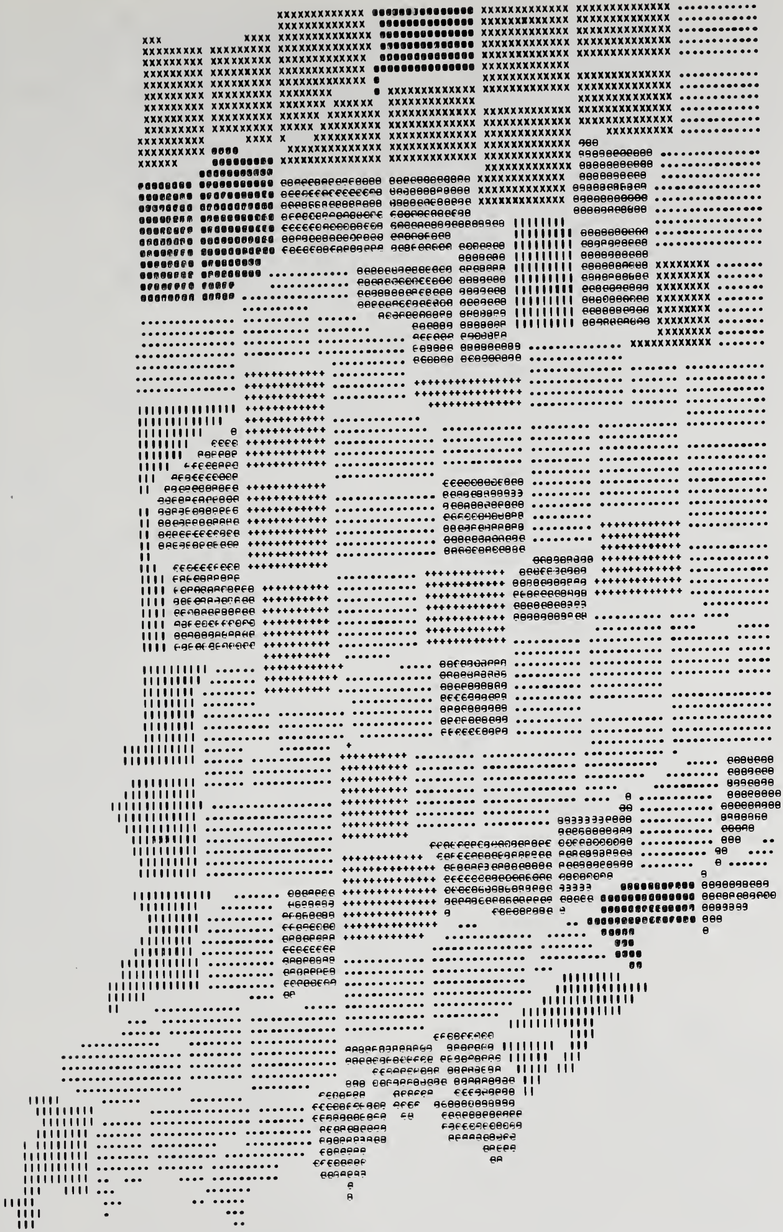

FIGURE 2. *Computer printout of the State of Indiana showing the relative similarity among counties as determined by numerical taxonomy. Counties shaded with the same symbol are more similar in the species that grow there than they are with counties shaded otherwise.*

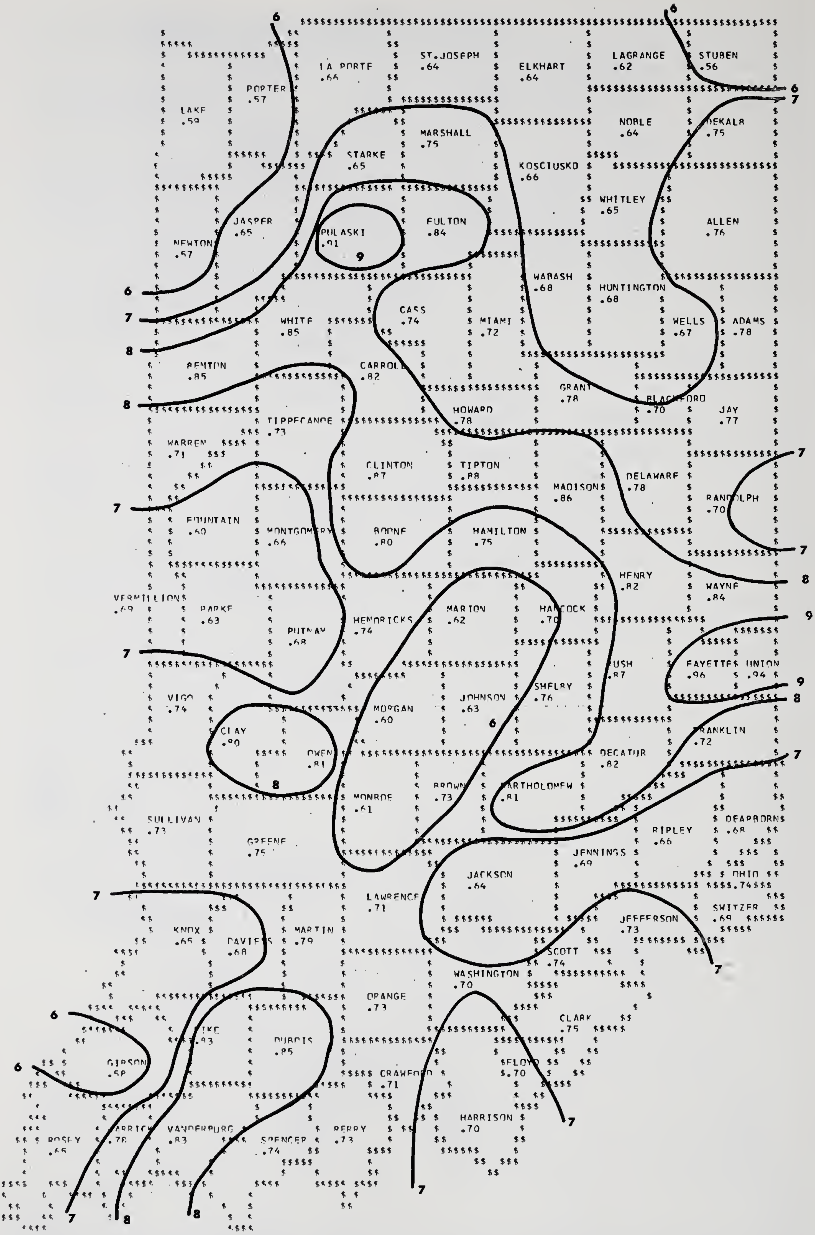

FIGURE 3. *Computer printout of the State of Indiana showing relationships among counties according to Huheey's Divergence Value. Isopleth lines of equal Divergence Values were fitted later by eye.*

data with the distribution pattern of one taxon at a time (using each map in Deam separately), or with the average distribution of a larger taxon (here the 81 species of the Brassicaceae). This can be done with transparent overlays (as in 1, 6), or it can be done via computer. We know of no other phytogeographic study which has calculated average phytogeographic similarities among geographic units. Zoogeographic workers such as Fisher (3) and Huheey (4) have made such analyses, however.

Figure 2 visually seems to be correlated with several climatic and geological characters. For example, there exist north to south latitudinal gradients on the figure which agree with climatic variables like length of frost free seasons and normal annual precipitation.

Even more interesting is the similarity among a "column" of counties in the southwest part of the state (Figure 2, counties with vertical line symbols; and those counties with the plus symbol). These coincide nicely with bedrock geology. Four counties were found to be very different from the others. Two, Jasper and Newton, are the familiar extension of the Illinois Prairie into northwestern Indiana. The third, St. Joseph County, may be anomalous due to the very intensive collecting by Notre Dame botanists, which extended over several decades. The authors are not familiar with Clark County, the fourth anomaly. But all of the 17 species reported there are very common. Also, it too was quite heavily collected when the Coulter brothers were resident at Hanover.

In this preliminary study readers must be careful not to put too much emphasis on the particulars of our results. Our caution stems from our current ignorance of which of the county-species combinations that are not positive are due to non-occurrence and which are due to incomplete collecting. Yet we can easily visualize the interesting questions that can be answered with such procedures. For example, does each plant family show the same phytogeographic boundaries in the State? If not, why not? Similarly, within a family like the Brassicaceae, do the many introduced weedy taxa show the same patterns as the native taxa? If not, what does this imply? Similar questions could be asked about different ecological succession stages. Such detailed studies await a more complete set of data.

The present study utilized only plant distribution data to demonstrate that the value of floristic data can be enhanced through selected use of the computer. Readers should be aware that such data are only a fraction of that available in a flora. Keller and Crovello (5) indicated how computers might be used to capture and analyze actual descriptions of taxa, including habitat data. These studies have led us to an even deeper conviction that reliable, descriptive taxonomy is essential to further progress in systematics and related fields.

Literature Cited

1. CADBURY, D. A., J. G. HAWKES, and R. C. READETT. 1971. A computer mapped flora. Academic Press, New York, N.Y. 788 p.

2. DEAM, C. C. 1940. Flora of Indiana. Indiana Dep. Conserv., Indianapolis. 1236 p.

3. FISHER, D. R. 1968. A study of faunal resemblances using numerical taxonomy and factor analysis. Sys. Zool. 17:48-63.

4. HUHEEY, J. E. 1965. A mathematical method of analyzing biogeographical data. I. Herpetofauna of Illinois. Amer. Midland Natur. 73:490-500.

5. KELLER, C., and T. J. CROVELLO. 1973. Procedures and problems in the incorporation of data from floras into a computerized data bank. Proc. Indiana Acad. Sci. 83:116-122.

6. PERRING, F. H., and S. M. WALTERS. 1962. Atlas of the British Flora. Thomas Nelson, London, Eng. 437 p.

7. SNEATH, P. H. A., and R. R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, and Co., San Francisco, Cal. 573 p.