

**A New Method of Multivariate Analysis Distinguishing Members
of Closely Related Taxa with an Application to Six Species
of the Trematode Genus *Haematoloechus***

GÜNTHER O. W. KRUSE

(Received for publication; April 28, 1980)

Key words: endoparasites, lung worms, frog, multivariate analysis, taxonomy

A taxonomist develops the ability to observe, to compare and to position specimens within the vast array of known taxa. At first the process may be tentative and relatively slow, but experience leads to more sure and deft decisions. The increased efficiency may be attributed to the programming of the taxonomist's natural computer, the brain. Because this computer may become biased, distracted or tired, mechanical computers are welcome tools which produce fast and consistent results.

The computer program described in this paper will *not* make taxonomic decisions. The analysis described in this paper would take two workers the better part of one year if done by hand; it was accomplished by one worker in one week using a computer.

Fuller (1969) described the original set of computer programs which were prepared by Dale Nelson, University of Nebraska Computer Center, and analysed variations of botanical material. A modified version adapted for zoological specimens has been developed and is used in this analysis. It enables the taxonomist to compare many specimens and their character ranges at one time and aids in sorting them according to

their similarities. Since no one character always expresses itself in a given manner, the taxonomist must recognize reliable trends within or between variable populations from a multitude of data.

The genus *Haematoloechus* Looss, 1899 (Trematoda: Haematoloechidae) consists of 40 species of lung flukes distributed throughout the world wherever anurans are found. Brooks (1976) reported *H. coloradensis* (Cort, 1915) Ingles, 1932, *H. complexus* (Seely, 1906) Krull, 1933, *H. longipectus* Stafford, 1902, *H. medioplectus* Stafford, 1902, *H. parviplectus* (Irwin, 1929) Harwood, 1932, and *H. varioplectus* Stafford, 1902, from Nebraska.

The Brooks Collection is available in the Harold W. Manter Laboratory, Division of Parasitology, University of Nebraska State Museum. One hundred sixty-four specimens of *Haematoloechus* from Nebraska were selected at random to test for the first time the applicability of this program to the analysis of zoological material.

Analysis

The first step in the analysis was to sort out any specimens which were not mature (uterus not filled with eggs). It is important to compare individuals of about the same developmental state.

The Harold W. Manter Laboratory, Division of Parasitology, 529-W Nebraska Hall, University of Nebraska State Museum, Lincoln, Nebraska 68588 U.S.A.

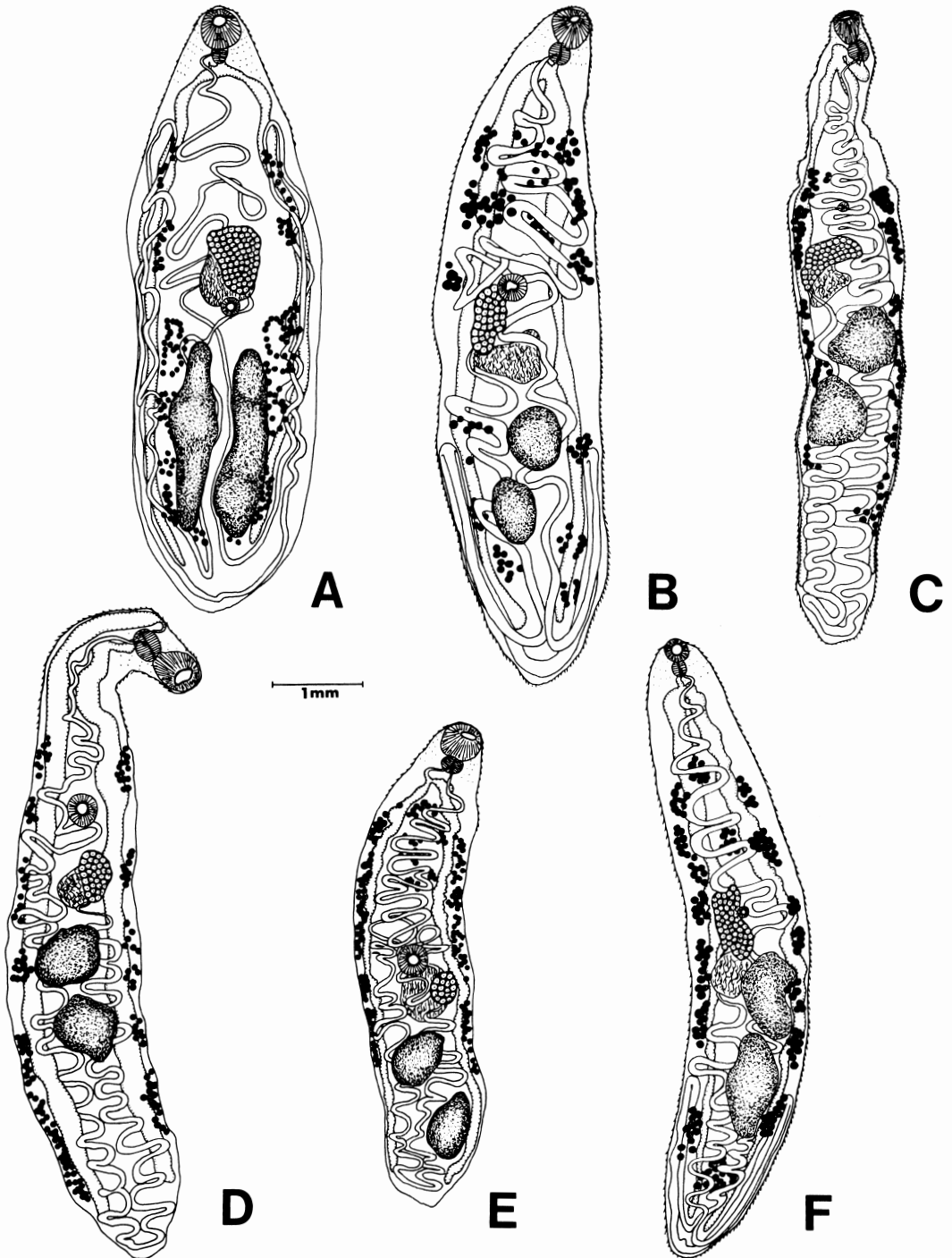


Plate I Fig. A: *Haematoloechus longiplexus* Stafford, 1902, Fig. B: *H. varioplexus* Stafford, 1902, Fig. C: *H. medioplexus* Stafford, 1902, Fig. D: *H. coloradensis* (Cort, 1915) Ingles, 1932, Fig. E: *H. complexus* (Seely, 1906) Krull, 1933, Fig. F: *H. parviplexus* (Irwin, 1929) Harwood, 1932. All figures are drawn to the same scale and are ventral views.

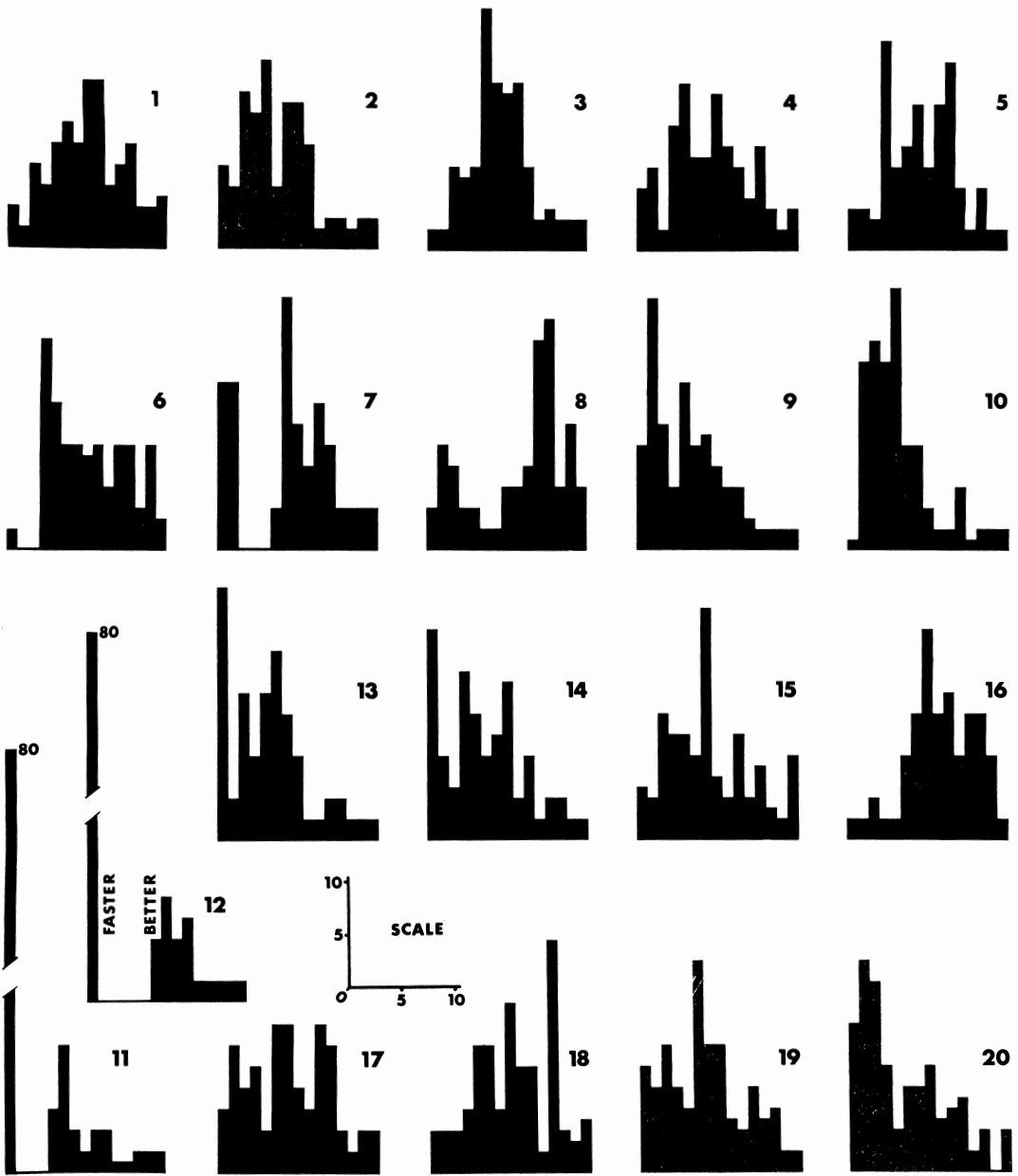


Plate II TOGETHER. Histograms resulting from the first NUTAL analysis. Abscissa of each character is divided into 15 cases; ordinate represents number of individuals in each case.

The second step is aspect sorting. In this process any groups that are obviously different are separated from the original cluster. On this basis individuals belonging to the

group with morphological characters shown in Figure A on Plate I were considered a species and dropped from further analysis.

Measurements

The remaining specimens, with all identification concealed, were used for the computer analysis. Morphological characters with sufficient variation were chosen and each specimen was measured for them: (1) body length; (2) body width; (3) body shape (=width/length); (4) forebody length; (5) oral sucker length; (6) oral sucker width; (7) acetabulum width; (8) sucker-width ratio; (9) pharynx length; (10) pharynx width; (11) uterus recoil; (12) uterus recoil/body length; (13) distance of vitellaria from anterior end; (14) #13/body length; (15) distance of ovary from anterior end; (16) #15/body length; (17) testes length; (18) testes width; (19) distance from anterior testis to posterior end of worm; (20) post-testicular space; (21) egg length; (22) egg width; (23) egg shape (=width/length); (24) ovary length; (25) ovary width. While they are relatively conspicuous, these characters do not necessarily represent the accepted taxonomic characters for *Haematoloechus*.

NUTAL Program

The population data were prepared for the first stage of the computer analysis, the NUTAL program.

The range of variation of each character was divided into 15 classes. This number of classes was chosen because it is the smallest multiple of five which has a median

class, it gives enough categories to retain precision of measurement, and it facilitates the conversion of the range into the five classes necessary for a later program.

The data were transferred to IBM cards and the NUTAL program was run. The computer calculated and printed out the frequency distribution, the mean variance, and the standard deviation of each character. The results of the first run (TOGETH) were plotted and are shown in Plate II. Arbitrary designations were assigned to the results of each analysis to remain unbiased throughout the program.

The shapes of the histograms varied strongly from character to character indicating that a mixed population was sampled. Only characters 11 and 12 permit separation into two distinct groups. The group on the left of the histogram (with no uterine recoil) was named FASTER, the other group was named BETTER. Each group was used for further analysis.

Group BETTER (with uterine recoil) was again analysed by the NUTAL program (Plate III).

There are several breaks in the distribution of the individuals throughout the 20 characters (e.g., character 9 and 13), but only character 8 (sucker ratio) gives a clear separation into two distinct groups. These clusters, named NEITHER and NOR, were again subjected to NUTAL and the histograms resulted in normal curves, indicating that each was a homogeneous population



Plate III BETTER. Histograms of characters 8, 9 and 13. Only character 8 shows a clear separation dividing group BETTER into NEITHER and NOR.

which could not be subdivided or for our purposes, a species.

Group **FASTER** (without uterine recoil) was again checked by the **NUTAL** program (Plate IV). Characters 11 and 12 were eliminated because the variables were zero. There is a break in characters 7 and 8 (acetabulum diameter and sucker ratio), and the two groups are named **BUT** and **NOT**. Notice here these characters had also been employed to separate **NOR** from **NEITHER**. Thus, character 7 could have been used earlier to divide **TOGETHER** showing that the sequence of groups dropping out has no importance and that the final results have no phylogenetic meaning. Further analysis of **BUT** was unproductive and therefore it was considered to represent a species.

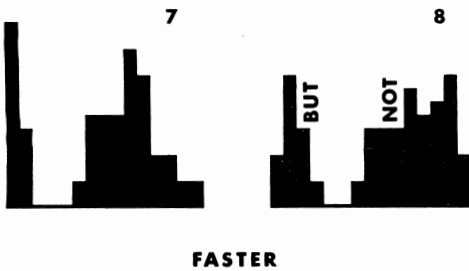


Plate IV **FASTER**, subdivided into **BUT** and **NOT** on the basis of the break in the histograms of characters 7 and 8.

Applying **NUTAL** to **NOT** resulted in histograms of a variety of shapes of curves (Plate V). Some of them were normal (fig. 1), skewed to the left (fig. 3), or bimodal (fig. 4).

A normal curve indicates that the genes for the character are normally distributed in the population as a result of interbreeding. This is evidence of either (1) a variable species or (2) introgressive hybridization between two species. These trematodes are hermaphroditic and probably self-fertilizing, so we have evidence of a variable species.

A skewed curve indicates that some of the specimens are juveniles, that the character is not distributed normally in the population, or that an inadequate sample of the population was taken. The juveniles had been eliminated during the first step of this analysis. It was concluded that an inadequate sample was taken. Those curves skewed to the left did so because there were fewer individuals in the one group (see below). This could not be foreseen because specimens had been chosen at random.

A bimodal curve for all characters indicates that the specimens were an adequate sample of a population with two different gene pools. If only a few histograms exhibit bimodality, these characters should be considered diagnostic and may be used to separate the taxa.

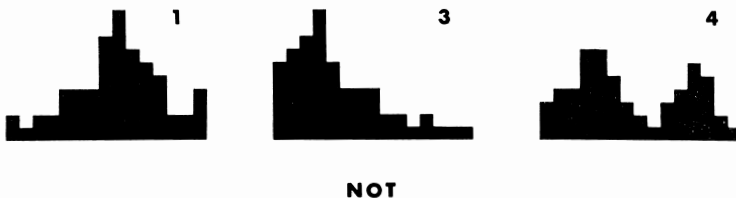


Plate V **NOT**. Examples of differently shaped histograms resulting from the application of the **NUTAL** program to this group.

Since the histograms showed this variety of shapes the next step was to determine if the specimens belonged to one or two species. This was accomplished by a second computer program, GOCOR, followed later by a third one, SIMIX.

GOCOR Program

The GOCOR program computes the Pearson's correlation coefficients, the sums, the sums of squares, the means, and the standard deviations.

The output consists of a table of correlation coefficients (Plate VI). The absolute values range from 0.000 to 1.000 (some being negatively correlated), representing the tendency of two independent variables to vary together.

It was necessary to determine systematically the acceptability of each character based on the correlation. To be significant, the value of correlation had to be above .195 (Snedecor and Cochran, 1967). Each positive character had to positively correlate with every positive character; every negative character positively correlate with every negative character. There were in-

consistencies in which some values were negative between two positive characters or where the values were lower than .195. Character 6 showed inconsistencies with many other characters; characters 12, 14, 15, and 16 also demonstrated irregularities, while character 3 was negative with all others. This procedure resulted in the elimination of five characters. The remaining characters are: 1, 2, 3 (-), 4, 5, 7, 8, 9, 10, 13, 15, 19, and 20.

A logical explanation for the consistent correlation of the remaining thirteen characters is to postulate close relationship. Further evidence for the explanation, especially with respect to the kind of remaining population sampled, was obtained by using the similarity index (=SIMIX).

SIMIX Program

The data were set up for the final computer program, SIMIX. SIMIX is a substitute name for Fuller's (1969) HYBIX (=Hybrid Index) because this author made some changes in the original program to accommodate zoological material and because the presence of hybrids is not established among the Trematoda.

The steps for setting up this final program were: (1) the determination of the range of variation for each character; (2) the division of this range of variation into five classes; (3) the assignment of 0 to the low class and the similarity value of 4 to the other extreme (reversed for negative characters); and (4) the totalling of the similarity values for each specimen to obtain the similarity index for the individual.

The limits of the sum are 0 and four times the number of characters used in the final program (i.e., 0 to 52, since 13 characters are involved). The frequency distribution of the similarity indices of the sample is presented in Plate VII.

The histogram shows a tendency toward a bimodal curve with a clear separation

**INTERCORRELATIONS
(NOT)**

13	-	4963
7	-	5892
-10	-	9882
13	-	1859
-18	-	17859
-18	-	9858
-17	-	7855
6	-	15626
-11	-	8625
③	-	13	-.625
neg				
12				
14				
15				
16				
		} below .195		

Plate VI Intercorrelations of the characters in NOT which result from the GOCOR program.

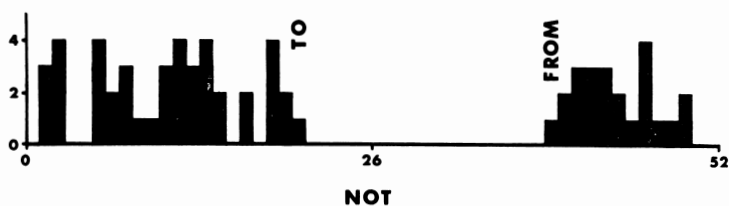


Plate VII NOT. Histograms resulting from the SIMIX program. The length of the absciss is derived from the highest value assigned to the class times the number of characters still in operation (i.e., 4 times 13).

around the mean value of 26. It can be seen that there are two separate, but similar, groups, probably indicating close relationship. These groups are named TO and FROM. The individuals at the extreme values would represent the types; the more a specimen deviates from these values the more it differs from the type. The kind of difference we are dealing with, convergence or divergence, can not be determined with this program. At this point, the taxonomist will have to make a decision. It may be interesting to note that the two specimens at the extremes were collected at almost the extreme east-west ends of Nebraska (about 400 miles).

Conclusion

The computer analysis of 164 specimens of *Haematoloechus* from Nebraska resulted in the identification of six groups. Comparing the individuals in the groups derived in this analysis to the original descriptions of the species in the genus *Haematoloechus*, we find (Plate VIII) that the six groups correspond with the six species previously reported from Nebraska: *H. longiplexus* (=OUT), *H. varioplexus* (=NOR), *H. parviplexus* (=NEITHER), *H. medioplexus* (=BUT), *H. complexus* (=TO), and *H. coloradensis* (=FROM).

Brooks (1976) reported *Haematoloechus coloradensis* from localities in both eastern and western Nebraska. According to Kruse

(1976) specimens which were collected in eastern Nebraska should be identified as *H. complexus*.

It should be pointed out that this program is a useful tool to distinguish similarity groups within a population. The taxonomic level is of no importance. Even though, the set of programs identifies and separates closely related taxa its results do not indicate phylogenetic relationship (see plate VIII). Also, the number of characters used is not critical as long as a sufficient number is retained for the final step of the program (about 10) to show the variation. Some characters which in the beginning appeared to be valuable had to be dropped from the program because they could not be observed in every specimen or did not show enough variation (like egg size = characters 21–23). For this analysis only morphological characters were used, but any data that can be scored may be included. It is important to have large numbers of individuals. The more specimens included, the smoother the curves of the histograms will be and the easier it becomes to distinguish breaks in the characters for separation.

The set of computer programs is available and may be obtained from the author. The programs are adapted for the IBM 360/65J computer with a FORTRAN IV compiler. The system runs on OS and HASP combined.

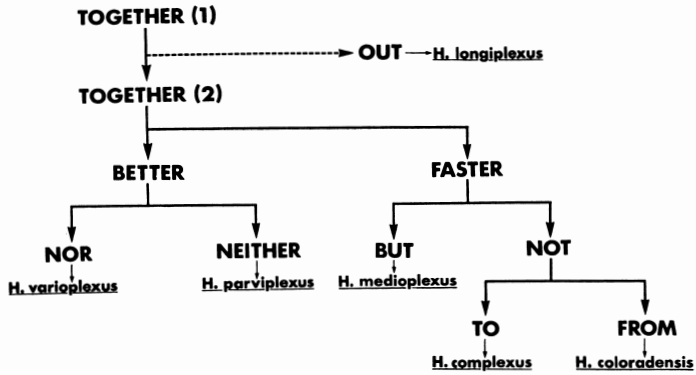


Plate VIII Summary chart showing the various groups as separated by the computer program. In the final analysis the end-groups corresponded with existing species of the tested genus, *Haematoloechus*.

Acknowledgements

The author wishes to express his appreciation to the following who have assisted in the work: J. F. Davidson under whose direction this study was made and who provided the computer time; M. H. Pritchard for use of the museum specimens and for reading the manuscript; D. R. Brooks who collected and prepared the specimens used in this computer analysis; M. Haack for designing the histogram layout.

References

- 1) Brooks, D. R. (1976): Parasites of amphibians of the Great Plains. Part 2. Platyhelminths of amphibians in Nebraska. Bull. Univ. Nebr. State Mus. 10, 65-92.
- 2) Fuller, M. J. (1969): The Genus *Carduus* L. in Nebraska. Univ. Nebr. Studies 39, 1-57.
- 3) Kruse, G. O. W. (1976): The *Haematoloechus* complex—a computer analysis of trematodes found in the lungs of anurans in Nebraska. Proc. Nebraska Acad. Sci. 86, 20 (Abstract).
- 4) Snedecor, G. W. and W. G. Cochran (1967): Statistical Methods. The Iowa State University Press. Ames, Iowa 6th ed. 593 pp.

近似分類構成単位を識別する新多変量解析法と
その *Haematoloechus* 属吸虫 6 種への応用

Günther O. W. Kruse

(ネブラスカ州立大学博物館寄生虫研究室)

多数の標本について行った多種形質の計測値（その形質の選定が適当であった場合）の度数分布をつくりその峰でグループ別けし、(NUTAL プログラムによる) さらに種別の困難な場合には、計測値が有意な正あるいは負の相関を示す形質をとりあげ、(GOCOR プログラムによる) それぞれの変異を 5 段階にわけて 0 から 4 までの値を与えてその合計を指標とする。再びこの指標値の度数分布をとってグループ別けをする

といった手順を踏むと (SIMIX プログラムによる) それぞれ同一性をもったいくつかのグループに到達することが出来る。

この操作はコンピューターに依らねばならないが実際にネブラスカ州産の吸虫 (*Haematoloechus* 属) の標本 164 個体に応用し、既知の 6 種の記載に一致する 6 つのグループに選別することが出来たので、この方法の有用性が確かめられた。