

## Metodologia (Methodology)

A numerical criterion for assessing  
the discriminatory/aggregative potential of a taxonomic  
character with a Fortran 90 program for calculations

Um critério numérico para avaliar  
o potencial discriminatório/agregativo  
de um caráter taxonômico com um programa  
Fortran 90 para cálculos

JOAQUIM CARLOS SENA MAIA<sup>1</sup>

One of the steps in a taxonomic study concerns the identification and design of groups that are outlined (OTU's Taxonomic Operational Units) which need to be analyzed by taxonomists (see SNEATH & SOKAL, 1973).

In all steps, two basic types of characters are taken into account: those that add simpler units and form hierarchically superior units, usually by similarities (*e. g.*, gathering of related species to form a taxomic group) and the ones discriminating different groups.

The basic intention of this work is to present an alternative method for, using numerical criteria to allow an assessment of the discriminatory or aggregative potential of a taxonomic character.

## MATERIAL AND METHODS

Six species of the genus *Plebeia* (Hymenoptera, Apidae) were used, namely: *Plebeia juliani*, *Plebeia droryana*, *Plebeia emerina*, *Plebeia meridionalis*, *Plebeia remota* and *Plebeia saiqui*, whose data were used in the author's master's thesis (MAIA, 2017), in the creation of an evaluation criterion for the interspecific structure of the *Plebeia* using

---

<sup>1</sup>Professor Adjunto (Aposentado) do Departamento de Estatística da UFPR. E-mail: jcsenamaia@gmail.com.

qualitative and quantitative variables. In this case, data were submitted to analysis of variance (MAIA, 2017).

To exemplify the method now proposed, we selected four morphological characters (punctuation of propodium, labrum color, pronotum color and scape color) with their respective status, such as, for example, "yellow" and "black" status for the character color of the scape. Of 17 attributes used in the thesis (MAIA, 2017) we choised four to demonstrate the functioning of the numerical model.

Each status of a given character works with a grouping criterion that can include one or more species. Thus, considering the color of the labrum, the status "white" is present in *P. juliani* and *P. meridionalis*, the "yellow" status includes *P. droryana* and *P. emerina*, while the "black" status includes *P. remota* and *P. saiqui* species.

The number of species "housed" with each status is represented by the letter K. Thus, in the analysis of the character color of the labrum, we have the value of  $K=2$  (two species) in each of the statuses (white, yellow and black). As three groups were formed (each group corresponding to a status) therefore  $n=3$ .

Then we calculate the product of K values (multiply  $K_1 \times K_2 \times \dots \times K_n$ ) to obtain the value of P. Extracting the root of order n of the value of P we obtain the geometric mean G of the product under consideration. For P is the producer.

$$G = \sqrt[n]{P} = \sqrt[3]{2 \times 2 \times 2} = 2$$

The value found (2) is geometric mean of the coefficients that indicate the number of species in each group (or status).

The geometric mean is the statistical quantity that best describes, in this methodology, the internal structure species distribution in relation to the status of the analyzed characters (see PIZANI ET AL. 1966).

To facilitate the interpretation of the results, we adopted the criterion of making the results oscillate between 0 (zero) and 5 ( $0 \leq D \leq 5$ ), calculating the value of D (degree of discrimination / aggregation) using the formula below, where N is equal to the number of species listed in this analysis.

$$D = 5 \frac{N - 2G}{N - 2}$$

where

D = degree of discrimination/aggregation of the analyzed character,

N = number of species (or other taxonomic unit),

G = geometric mean of the K values.

The value found (G=2) is the geometric mean of the product of the K values, that is, the value average of the coefficients that indicate the number of species in each group (or status). Using the data obtained on the previous page, we can calculate the value of the discriminatory potential/aggregative potential of the color taxonomic character of labrum:

$$D = 5 \frac{N \cdot 2G}{N - 2} \quad D = 5 \frac{6 \cdot 2(2)}{6 - 2} = 2,5$$

The interpretation of the D value can be done as follows:

0 >= D <= 1 — very aggregative characters

1 < D <= 2 — aggregative characters

2 < D <= 3 — intermediate characters

3 < D <= 4 — discriminatory characters

4 < D <= 5 — very discriminatory characters

## RESULTS AND DISCUSSIONS

We present in the Table 1 as an example of the calculations made for

Table 1. Grouping formations.

Characters	Status/grouping formation			G	D				
Propodium punctuation	rugous and reticulate	rugous just in the base		P	G	D			
	J - M - E (k=3)	D - R - S (k=3)		9	3	0			
Labrum color	white	yellow	black	P	G	D			
	J - M (k=2)	D - E (k=2)	R - S (k=2)	8	2	2.5			
scapum color	yellow	black		P	G	D			
	D (k=1)	J - M - E - R - S (k=5)		5	2.2	2			
pronotum color	entirely black	incompl yellow band	complete yellow band	incompl striae	complete striae	entirely yellow	P	G	D
	J (k=1)	M (k=1)	D (k=1)	E (k=1)	R (k=1)	S (k=1)	1	1	5

4 morphological characters of 6 species of the genus *Plebeia*. Each species is represented by the initial letter of respective specific name as follows: *juliani*: J, *meridionalis*: M, *droryana*: D, *emerina*: E, *remota*: R, *saiqui*: S.

The Table 1 shows that the character “color of the pronotum” can be classified as “very discriminatory”, as the value of  $D$  reaches the maximum possible, that is 5. This value occurs when a taxonomic character has the property of being able to identify all the members of the analyzed taxon. The characters “color of the labrum” and “color of the scape” have an intermediate degree of discrimination and the  $D$  values are in the range between 2 and 3, more precisely 2.5 and 2.0.

The character “punctuation of the propodium” forms 2 groups of equal size indicating a minimum grade discriminatory (or maximum aggregation) because the value of  $D = 0$  (zero). This occurs when we have a taxon formed by an even number of taxonomic units, with two statuses, and each status “hosts” exactly half of the members.

## CHARACTER WEIGHTING

BURT (1964) warned that numerical taxonomy need not necessarily assign equal weights to all characters (isocratic classification). SNEARTH (1973) argues that the most recommended is to use equal weights for all characters. He agrees, however, that some complex characters are given weights in proportion to their complexity or informational content. MICHENER AND SOKAL (1957) conclude that, even if it is desirable, there is no rational way to determine character weights and, in practice, we should assign the same weight to all. According to them, when many characters are used, the similarity analyzes are only slightly affected by the weighting effects. FARRIS (1966) suggests that characters that vary little within populations are more reliable indicators of cladistic relationships than more variable characters and therefore should have greater weights. GOODMAN (1969) believes that a character should be inversely weighted to the variance within the taxonomic unit. In his studies of plant reproduction he found that variation can be largely attributed to environmental rather than genetic effects. Thus, genotypic differences can be better evaluated if highly variable characters are not emphasized by weighting criteria.

The calculation method we present in this work applies to the application of qualitative (or discrete) variables in which the different status of the analyzed characters can be coded by letters or numbers that identify them, without ranking them in taxonomic importance.

Thus, we discuss the possibility of using the values of  $D$  as a weight-

Table 2. Weighting scale of D values emphasizing discriminative potential.

D value	Degree of discrimination
$0 \leq D \leq 1$	Character very little discriminatory
$0 \leq D \leq 2$	Little discriminatory character
$0 \leq D \leq 3$	Moderate discriminatory character
$0 \leq D \leq 4$	Discriminatory character
$0 \leq D \leq 5$	Very discriminatory character

Table 3. Weighting scale of D' values emphasizing aggregative potential.

D' value	Degree of aggregation
$0 \leq D' \leq 1$	Character very little aggregative
$0 \leq D' \leq 2$	Little aggregative character
$0 \leq D' \leq 3$	Moderate aggregative character
$0 \leq D' \leq 4$	Aggregative character
$0 \leq D' \leq 5$	Very aggregative character

ing factor, as it provides information that evidences the potential for aggregation or discrimination of a taxonomic character.

The calculation of D proposed in this article allows the creation of two weighting scales, emphasizing either the discriminative potential or the aggregative potential. If we intend to emphasize the discriminative potential, just calculate the values of D with the interpretation given in table 3.

However, if the objective is to emphasize the aggregative potential of the character, we must calculate D' complementary value of D, as follows:  $D' = 5 - D$ . The interpretation of the values of D' is shown in table 3 the capacity of aggregation or discrimination of a taxonomic character.

Some methods of numerical taxonomy can use weights that allow to differentiate the contribution of each character in the formation of a given

taxonomic group. What we are discussing is the possibility of using the values of D as a weighting factor.

For calculus see the applicative (see pág. 14).

## CONCLUSIONS

1) In a taxonomic study, morphological, physiological, behavioral and ecological characters can be used.

2) Whatever the nature of the character used, as long as it receives an appropriate encoding, it can be analyzed as to its discriminative or aggregative potential for the configuration of a given undetermined taxon.

3) The contribution of character variations in each group allowing the individualization of each taxonomic unit can be expressed numerically through the methodology here presented.

4) The geometric mean can be an adequate way to show taxonomic relationships.

5) The calculation of D makes it possible to evaluate the informative potential of a taxonomic character in terms of discrimination or aggregation of taxonomic units.

6) It would be important to assess the relevance of using D as a weighting criterion in taxonomic studies.

Acknowledgements—To Prof. Graci Maria Fernandes de Oliveira (*in memoriam*) for suggestions that had contributed for the methodology formalization and to Dr. Sebastião Laroca for encouragement and suggestions.

## SUMMARY

An alternative method of assessment of the discriminatory and aggregative potentials of a taxonomic character as contribution to numerical taxonomy.

KEY WORD: numerical method; taxonomy; aggregative potential; discriminatory character.

## SUMÁRIO

Um método alternativo de avaliação dos potenciais discriminatórios e agregativos de caráter taxonômico, como contribuição à taxonomia numérica.

PALAVRAS CHAVE: método numérico; taxonomia; agregativo; discriminatório..

## BIBLIOGRAPHY

- BURTT, B. L. 1964. Angiosperm taxonomy in practise. Phenetic and phylogenetic classification, pp 5-16. Sys. Ass. Pub. 6. 164 pp in Sneath P. H. A. & R. R. Soka. 1973. *Numerical Taxonomy* — W. H. Freeman and Company, San Francisco, 573 pp.
- FARRIS J. S. 1966. Estimation of conservatism of character by constancy with biological biological populations. *Evolution* 20: 587-591.
- GOODMAN, M. M. 1969. Measuring evolutionary divergence. *J. J. Genet.* 44 (suppl. 1): 310-316 in *Numerical Taxonomy*
- MAIA, J. C. S. 2017. Análise estatística interespecíficas espécies paranaenses (Brasil) de *Plebeia* (Apidae, Meliponi). *Acta Biol. Par., Curitiba*, 46 (1-2): 39-58.
- MICHENER, C. D. & R. R. Sokal. 1957. A quantitative approach to a problem in classification. *Evolution* 11: 130-162.
- PISANI, J. F., J. S. MOURE, B. CRESTANA, D. AILY, M. L. LORENZETTI. 1966. Análise estatística da estrutura interespecífica dos *Bombus* do Brasil (Hymenoptera, Apidae). *Bol. Univ. Fed. Paraná. Zool.* 8: 101-120.
- SNEATH, P. H. A. & R. R. SOKAL. 1973. *Numerical Taxonomy*, F. H. Freeman and Company, San Francisco 573 pp.
- SOKAL, R. R. & C. D. MICHENER. 1958. A statistical method for evaluation systematic relationship. *Univ. Kansas Sci. Bull.* 38: 1409-1438.

Fortran 90 program to calculate D and D' values. A tutorial is also presented. An executable file may be obtained by the email [jcsenamaia@gmail.com](mailto:jcsenamaia@gmail.com). To run this program, the computer must have a Fortran language compiler, which can be downloaded freely the website: <http://force.lepsch.com>. (Force 2.0.9 + GNU Fortran (GFortran)).

## TUTORIAL

This tutorial shows step by step how to provide the requested informations the GDDL.F program.

Step by step way of using the program.

- "How many OTU's (species)? "

Number of OTU's of groups to be analysed.

- "Names of OTU's." / "Name of OTU ", J
- "Which character will be analysed?"

Name

- "How many statuses? "
- "Inform Status "
- "Species " / "STATUS code "

For each species give the corresponding status code

- "You informed " / "Is it correct? (0 = no)"

The program informs the digitalized and asks for confirmation. In the case of mistakes, give "zero" to correct.

- "CALCULATION RESULTS " / "D =", "DL =", "G ="

The results will be given at "D" e "DL" and "G"

- "Will you analyse another character? YES = 1 "

for additional character write 1. Qualquer outro número encerra o programa



## 1. PROGRAM GDDL.F

```
1:  Program Calculo GDDL
2:  Implicit none
3:  Real, Dimension(100, 100) :: Mat
4:  Real, Dimension(100) :: soma, somasta
5:  Real :: G, D, DL, Rnusta, Rnuto
6:  Integer :: N, I, J, K, L, M, nuto, icod, nusta, inf, prod
7:  Integer :: Nova
8:  Character(len=15), Dimension(100) :: nomuto, nomstatus
9:  Character(len=15) :: nomcarac
10: 4  FORMAT(I3)
11:  Print*, "How many UTOs are part of the analyzed group?"
12:  Read(*,4) nuto
13:  Do J=1, nuto
14:  Print*, "UTO name (specie ?) =", J
15:  Read*, nomuto(J)
16:  Print*
17:  end Do
18: 7  Print* ; Print*
19: 10 Do 30 J=1,100
20:  Do 20 K=1,100
21:  Mat(J,K) = 0
22: 20 Continue
23: 30 Continue
24:  nomcarac = " "
25:  nusta = 0
26:  Do L = 1,100
27:  nomstatus(L) = " "
```

```
28:   End Do
29:   Print*, "Character to be analyzed"
30:   Read(*,36) NOMCARAC
31: 36  Format(A15)
32:   Print* ; Print*
33:   Print*, "How many status does this character have?"
34:   Read*, nusta
35:   Print*
36:   Print*
37:   Do K=1, nusta
38:   Print*
39:   Print*, "What is the STATUS associated with the code", K
40:   Read*, NOMSTATUS(k)
41:   Print*
42:   End Do
43:   Print* ; Print*
44:   Do J=1,nuto
45: 40  Print*, "Status code of *", NOMCARAC, "*for OTU = *",
NOMUTO(J), "*"
46:   Read*, icod
47:   Mat(icode,J) = 1
48:   Print*, "You informed *", nomstatus(icode), "*"
49:   Print*, "Is this correct? 0 = NO /// 1 = YES"
50:   Read*, inf
51:   Print*
52:   IF (inf == 0) GOTO 40
53:   End Do
54:   Do k=1,nuto
55:   Soma(K) = 0
56:   End Do
57:   Do J=1,nuto
58:   Do I=1, nusta
59:   Soma(J) = Soma(J) + Mat(I,J)
60:   End Do
61:   End Do
62:   Do K=1,nuto
63:   If (soma(K) .NE. 1) GOTO 50
64:   End do
65:   GOTO 60
```

```

66: 50 Print*, "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
67:   Print*, "Error typing codes. Start over!!!"
68:   Goto 7
69: 60 Do K = 1, nusta
70:   somasta(k) = 0
71:   End Do
72:   Do I=1,nusta
73:   Do J=1,nuto
74:   somasta(I) = somasta(I) + Mat(I,J)
75:   End Do
76:   End Do
77:   DO J=1, NUSTA
78:   IF(somasta(j) .EQ. 0) then
79:   goto50
80:   End IF
81:   End Do
82:   Prod = 1
83:   Do J = 1,nusta
84:   Prod = Prod * somasta(J)
85:   End Do
86:   Rnusta = nusta
87:   Rnuto = nuto
88:   G = Prod ** (1/Rnusta)
89:   D= 5* (Rnuto - 2*G) / (Rnuto - 2)
90:   DL = 5 - D
91:   Print*
92:   Print*, "RESULT OF CALCULATIONS"
93:   Print*
94:   Print*, "D =", D, "DL =", DL, "G =", G
95:   Print*
96:   Print*, "Do you want to analyze another character? NO=0 //
YES=1"
97:   Print*
98:   Read*, Nova
99:   IF (Nova .EQ.1) goto 10
100:   Print*
101:   END Program

```

