

Evaluation of Coding and Scaling Techniques in Fungal Numerical Taxonomy

A. ISMAIL

Department of Biological Science, Faculty of Education, Suez Canal University, Port-Said, Egypt

E-mail: ahmey@myway.com

ABSTRACT

The genus *Stemphylium* which is phenetically closer to the genus *Ulocladium* and the genus *Stachybotrys* which is morphologically similar to the genus *Memmoniella* were isolated from plant debris in Egypt. The isolates were identified as: *Stemphylium sarciniforme* (1), *Stemphylium vesicarium* state of *Pleospora herbarum* (2), *Ulocladium botrytis* (3), *Stachybotrys* state of *Melanopsamme pomiformis* (4), *Stachybotrys atra* (5) and *Memmoniella echinata* (6). Fungal descriptions were digitized for machine computation. Mutual relationships between isolates (OTU'S) were calculated by Canberra distance coefficient (CD) and flexible sorting clustering method (G). Characters coding was carried out by two techniques, while scaling was done with five ones. Results showed that the coding techniques had no significant effects, while the scaling techniques were found to affect the classification. Scaling of data by standardization of data lines improved the resulted dendrograms. Results also revealed that techniques of normalization of data lines, normalization of data columns and standardization of data columns were recorded to produce classification disorders.

Key Words: Numerical taxonomy; Coding and scaling; Fungi

INTRODUCTION

Numerical taxonomy means the grouping by numerical methods of taxonomic entities into taxa on the basis of their character states. This, therefore, requires the conversion of informations about the taxonomic units into numerical quantities. Many techniques are reported by Sneath and Sokal (1973) to enable this. Coding and scaling convert the crude data into digits suitable for computation. They overlap to some extent, but coding is largely concerned with logical decisions and divisions, and scaling is some form of mathematical transformation. The problems of coding and scaling have been discussed by many authors (Sneath, 1957; Proctor & Kendrick, 1963; Sokal & Michener, 1967; Lance & Williams, 1967a; Moss, 1968; Ismail, 1986). Sneath and Sokal (1973) stated that many of the coding and scaling methods have not yet been explored in practice. They recommended for predominantly two-state characters, that it should be coded by 0, 1 and not standardized and for predominantly multistate characters, they advised standardization with removal of size factor, where it is necessary. Mohamed *et al.* (2005) stated that the numerical taxonomy proved valuable as a base for identification of *Streptomyces*, however further studies are needed for more evaluation of this method.

The aim of the present study was to assess the different coding and scaling techniques of numerical methods in fungal taxonomy.

MATERIALS AND METHODS

Several plant debris were collected from cultivated fields at Aga (East Delta, Egypt). Plant samples were

transferred on to potato dextrose agar and Czapek's agar media. Incubation was done at 30°C for 10 days. The isolated genera *Stemphylium*, *Ulocladium*, *Stachybotrys* and *Memmoniella* were purified and identified according to the manuals of Ellis (1971) and Moubasher (1993). A total of 29 character units were recorded for each fungal isolate (Operational Taxonomic Units, OTU^s). These include both quantitative (1 - 9) and qualitative (10 - 29) phenetic characters. Coding for quantitative characters is done by real numbers, while qualitative ones are coded by 0 (negative or absent) and 1 (positive are present) in the first data matrix (Table I) and coded by real numbers in the second data matrix (Table II). Scaling of data was carried out by various techniques including normalization of data lines, normalization of data columns, standardization of data lines and standardization of data columns. Normalization of data was performed according to the formula:

$$\bar{X}_i = \frac{X_i - X_{\min}}{X_{\max}} \quad (\text{Anderberg, 1973})$$

X_i = Score of character for OTU^s i

X_{\max} = Maximum score of character values

X_{\min} = Minimum score of character values

While standardization was done using the formula:

$$\bar{X}_i = \frac{X_i - \bar{X}}{S_x} \quad (\text{Anderberg, 1973})$$

X_i = Score of character for OTU^s i

Table I. Data matrix with characters coded as interval and bi-state

CHARACTERS OTU'S	1	2	3	4	5	6	Type
1- Maximum length of conidiophore	55	100	110	260	110	110	Interval
2- Minimum width of conidiophore	5	5	4	4	4	4	Interval
3- Maximum width of conidiophore	11	4	6	6	6	5	Interval
4- Minimum Length of conidia	35	30	14	7	10	4	Interval
5- Maximum Length of conidia	60	45	30	13	12	6	Interval
6- Minimum width of conidia	25	25	7	5	6	4	Interval
7- Maximum width of conidia	35	34	18	8	4	6	Interval
8- Number of transverse speta	3	3	3	0	0	0	Interval
9- Number of longitudinal septa	5	3	1	0	0	0	Interval
10- Colony velvety	1	1	1	1	1	0	Bi-state
11- Colony powdery	0	0	0	0	0	1	Bi-state
12- Colony olivaceous brown	1	1	0	0	0	0	Bi-state
13- Colony blackish green	0	0	0	1	1	0	Bi-state
14- Colony blackish brown	0	0	1	0	0	0	Bi-state
15- Colony black	0	0	0	0	0	1	Bi-state
16- Conidiophore unbranched	1	1	0	0	1	0	Bi-state
17- Conidiophore occasionally branched	0	0	1	1	0	1	Bi-state
18- Conidiophore granulated near apex	0	0	0	1	1	0	Bi-state
19- Phialides present	0	0	0	1	1	1	Bi-state
20- Conidia solitary	1	1	1	0	0	0	Bi-state
21- Conidia in chain	0	0	0	0	0	1	Bi-state
22- Conidia in groups	0	0	0	1	1	0	Bi-state
23- Conidia smooth	1	0	0	1	0	0	Bi-state
24- Conidia minutely verrucose	0	1	0	0	0	0	Bi-state
25- Conidia verrucose	0	0	1	0	1	1	Bi-state
26- Conidia spherical	0	0	0	0	0	1	Bi-state
27- Conidia subspherical	1	1	0	0	0	0	Bi-state
28- Conidia obovoid	0	0	1	0	0	0	Bi-state
29- Conidia ellipsoidal	0	0	0	1	1	0	Bi-state

Table II. Data matrix with characters coded as interval

OTU'S CHARACTER	1	2	3	4	5	6	Type
1- Maximum length of conidiophore	55	100	110	260	110	110	Interval
2- Minimum width of conidiophore	5	5	4	4	4	4	Interval
3- Maximum width of conidiophore	11	4	6	6	6	5	Interval
4- Minimum Length of Conidia	35	30	14	7	10	4	Interval
5- Maximum Length of Conidia	60	45	30	13	12	6	Interval
6- Minimum width of Conidia	25	25	7	5	6	4	Interval
7- Maximum width of Conidia	35	34	18	8	4	6	Interval
8- Number of transverse septa	3	3	3	0	0	0	Interval
9- Number of longitudinal septa	5	3	1	0	0	0	Interval
10- Colony colour	1	1	3	2	2	4	Interval
11- Conidiophore braching	1	1	2	2	1	2	Interval
12- Conidiophore granulated near apex	1	1	1	2	2	1	Interval
13- Phialides resent	1	1	1	2	2	2	Interval
14- Conidia arrangement	1	1	1	3	3	2	Interval
15- Conidia surface	1	2	3	1	3	3	Interval
16- Conidia shape	2	2	3	4	4	1	Interval

\bar{X} = Mean score of character values

Sx = Standard deviation for character

As a result of scaling techniques, data matrix is computer designed in 5 forms for both data matrix (I) and data matrix (II): Data matrix containing raw values for all characters and character states, data matrix with normalized data lines, data matrix with normalized data columns, data matrix with standardized data lines and data matrix with standardized columns. The mutual relationships on the basis of overall similarities between OTU^S were calculated by Canberra distance coefficient (CD) according to the formula:

$$CD_{ij} = \sum_{k=1}^n \frac{|X_i - X_j|}{|X_i + X_j|} \quad (\text{Anderberg, 1973})$$

X_i = Score of character for OTU^S i

X_j = Score of character for OTU^S j

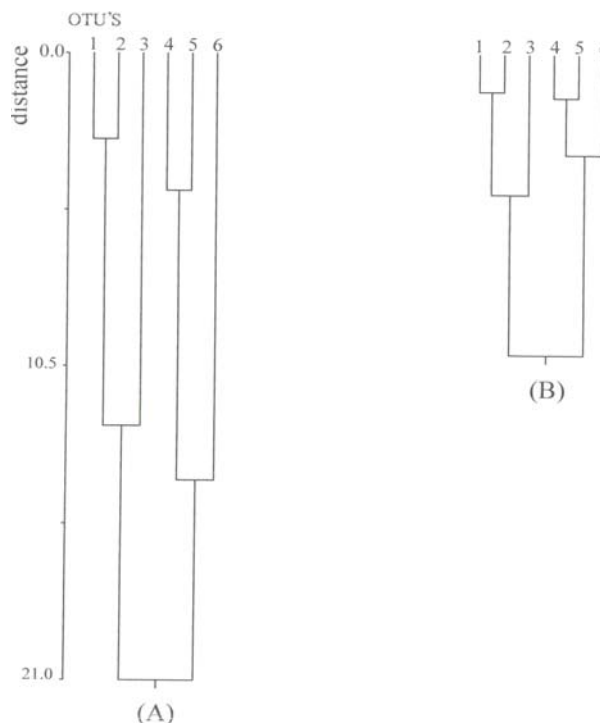
CD_{ij} = Canberra distance between OTU^S i and j

The distance matrix is analysed by flexible sorting method (Lance & Williams, 1967b) as recommended by Ismail (1998). All computer analysis were carried out by "quant" program (Ismail & Batko, 1996), running on "Gateway" PC computer.

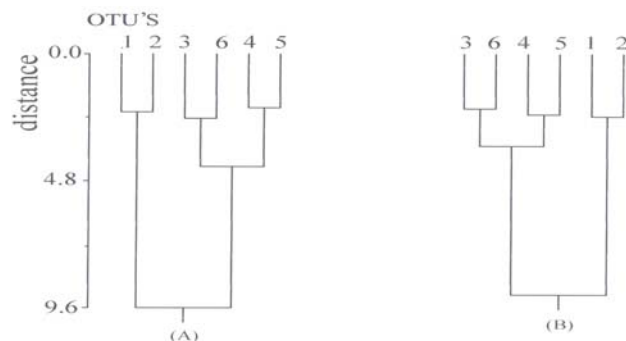
RESULTS AND DISCUSSION

Identification procedure resulted in the detection of the following species. *S. sarciniforme* (1), *S. state of Pleospora herbarum* (2), *U. botrytis* (3), *Stachybotrys state Melanopsamma pomiformis* (4), *Stachybotrys atra* (5) and *Memmoniella echinata* (6). Computer analysis revealed that the techniques of data coding has no significant effect on the resulted taxonomic structures. The main difference is the application of large number of characters improved the topology of dendrograms (Dend. 1). Grouping of fungal species using data without any kind of scaling "raw data" (Dend. 1) showed that *S. sarciniforme* (1) and *S. state of pleospora herbarom* (2) was grouped at similarity level of 97.2% and above. At similarity level of 87.4%, *U. botrytis* was added to this group. *Stachybotrys state of Melanopsamma pomiformis* (4) and *Stachybotrys atra* (5) were clustered at similarity level of 95.3% and above. *Memonoiella echinata* (6) was clustered with them at

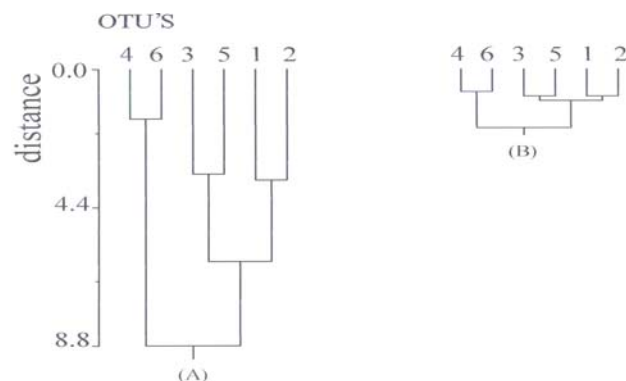
Dend. 1. Numerical classification based on raw data (A) Interval and bi – state coding (B) Interval coding



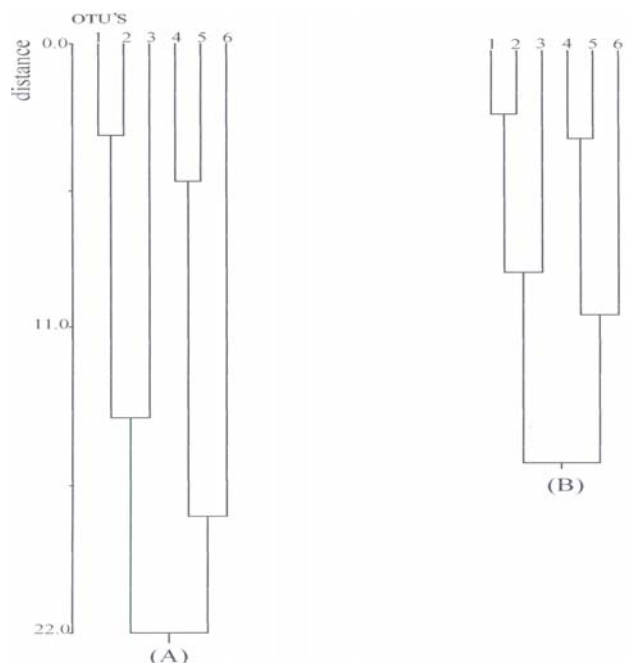
Dend. 2. Numerical classification based on normalized data lines (A) Interval and bi – state coding (B) Interval coding



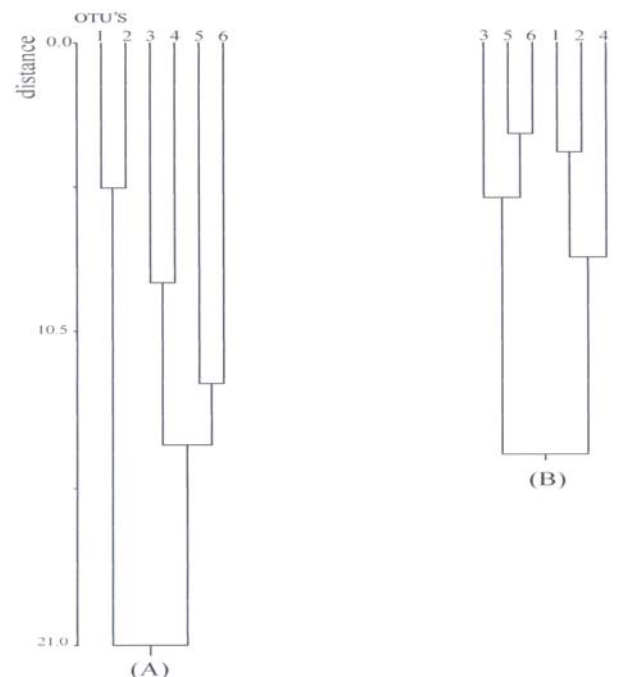
Dend. 3. Numerical classification based on normalized data columns (A) Interval and bi – state coding (B) Interval coding



Dend. 4. Numerical classification based on standardized data lines (A) Interval and bi – state coding (B) Interval coding



Dend. 5. Numerical classification based on standardized data columns (A) Interval and bi – state coding (B) Interval coding



similarity level of 85.8%. Two main groups were reported at similarity level of 78.8% and above; first group consists of 3 fungal species, which are: *Stemphylium sarciniforme*, *Stemphylium* state of *Pleospora herbarum* and *Ulocladium botrytis*. The second group contains *Stachybotrys* state of *Melanopsamma pomiformis*, *Stachybotrys atra* and *Memnoiella echinata*. Jarvis *et al.* (1998) reported two isolates of *Stachybotrys chartarum* and *Memnoiella echinata* to be closely related and were shown to produce a number of similar highly toxic compounds. More similar results were obtained by using the technique of standardization of data lines (Dend. 4), the only benefit of this method is that it produces dendrograms, which are topologically better (the level of similarities 99.3%, 91.1%, 89.5% & 87.6%, respectively). The used methods of normalization of data lines, normalization of data columns, standardization of data columns failed to group fungal species in correct manner (Dend. 2, 3 & 5). The author recommended using a large numbers of characters particularly the bi-state one as coding method and using the technique of standardization of data lines as the more efficient for data scaling.

REFERENCES

Anderberg, M.R., 1973. *Cluster Analysis for Application*, P: 359. Academic Press. New York and London
 Ellis, M.B., 1971. *Dematiaceous hyphomycetes*, P: 608. Commonwealth Mycological Institute, Kew, Surrey, England
 Ismail, A., 1986. The numerical study of Polish Cladosporia. *Ph. D. Thesis*, P: 305. Warsaw University, Poland

- Ismail, A., 1998. Evaluation of Clustering methods and numerical classification of three closely related genera of the family Mucraceae. *J. Union Arab Biol.*, 9: 195–208
- Ismail, A. and A. Batko, 1996. A Phenetic numerical taxonomy programs in Basic. *The 1st Conference on the Role of Science in the Development of Egyptian Society and Environment*, Pp: 21–3. October, 1996, Benha, Egypt
- Jarvis, B.B., W.G. Soreson, E.L. Hintikka, M. Nikulin, Y.H. Zhou, J.Jiang, S.G. Wang, S. Hinkley, R.A. Etzel and D. Deaborn, 1998. Study of toxin production by isolates of *Stachybotry hemosiderrosis* in infants. *Appl. Environ. Microbiol.*, 64: 3620–5
- Lance, G.N. and W.T. Williams, 1967a. Mixed-data classificatory programs. I-Agglomerative system. *Australian Computer J.*, 1: 15–20
- Lance, G.N. and W.T. Williams, 1967b. A general theory of classificatory sorting strategies. Hierarchical system. *Computer J.*, 9: 373–80
- Moss, W.W., 1968. Experiments with various techniques of numerical taxonomy. *Syst. Zool.*, 17: 31–47
- Mohamed, S., E.A. Saleh and M.M. Zaki, 2005. Identification of eight halotolerant Streptomycete isolates using a suggested numerical taxonomy. *Arab University J. Agric. Sci.*, 13: 641–68
- Moubasher, A.H., 1993. *Soil Fungi in Qatar and other Arabic Countries*. The Center for Scientific and Applied Research, Qatar
- Proctor, J.R. and W.B. Kendrick, 1963. Un-equal weighting in numerical taxonomy. *Nature*, 197: 716–7
- Sneath, P.H.A., 1957. The application of computers to taxonomy. *J. Gen. Microbiol.*, 17: 201–6
- Sneath, P.H.A. and R.R. Sokal, 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification*, P: 573. Freeman, San Francisco
- Sokal, R.R. and C.D. Michener, 1967. The effects of different numerical techniques on the Phenetic classification of bees of the Hoplitis complex. *Proc. Linn. Soc. London*, 178: 59–74

(Received 12 June 2006; Accepted 10 July 2006)