

## APPLICATION OF ORDINAL CLUSTERING TO THE TAXONOMY OF THE GENUS *ENTADA* (FABACEAE) IN TAIWAN

SHENG-ZEHN YANG<sup>1</sup>, YU-YING FENG AND FU-YA YEH<sup>2</sup>

*Department of Forestry, National Pingtung University of Science and Technology,  
1 Shuehfu Road, Neipu, Pingtung 91201, Taiwan*

*Key words:* *Entada*, Non-metric multidimensional scaling, Numerical taxonomy, Ordinal cluster analysis, Ordinal scaling

### Abstract

The importance of a mathematically correct treatment of ordinal information has not sufficient been emphasized. Since ordinal cluster analysis (OrdCIAn) and non-metric multidimensional scaling (NMDS) are order-invariant, an attempt has been made to illustrate the previous taxonomic treatments of three species of *Entada* Adans. (Fabaceae) of Taiwan by these two methods. The genus *Entada* is represented in Taiwan by *E. koshunensis* Hayata et Kanehira, *E. phaseoloides* (L.) Merr. and *E. rheedii* Spreng. In this study, 29 morphological characteristics and 14 operational taxonomic units (OTUs) representing these three species were selected to construct a primary mixed data matrix. Due to incomplete and limited herbarium specimens, the same local specimens were combined to form 14 OTUs from 62 specimens preserved in several Taiwan herbaria. The characters were measured on binary nominal, multistate nominal, ordinal, and quantitative scales. OrdCIAn and NMDS based on Gower's index were used to evaluate the relationship among the three *Entada* species from the primary matrix. The cluster dendrogram and the ordination of OTUs showed that the genus *Entada* obviously can be divided into three groups representing three existing species, therefore, agreed with previous taxonomic analyses.

### Introduction

Nominal, ordinal, interval and ratio data types are usually used in numerical taxonomy or vegetation ecology (Anderberg 1973). The methodological sequence implied by exploratory analysis of ecological data includes sampling, data collection, resemblance matrix calculation, and classification or ordination (Podani 2005). Each move between these levels of abstraction can be carried out by applying either an ordinal (O, sequence-/rank-based) or a metric (M, difference-/ratio-based) scale and methodological paths, such as M-M-O, O-M-O, O-M-M, O-O-M, M-O-M, M-O-O, O-O-O, are implied. In these paths ordinal and metric properties are confounded. Switching from M scale to O scale, some information is always lost and actual differences between distances become neglected, while changing from O scale to M scale, the increase of information may not be mathematically correct (Krauth 1986, Dale 1989, Podani 2005). Future multivariate analysis should be ordinal in nature so that the order of dissimilarities can be considered by algorithm. Thus, the sequence O-O-O is the best combination for the

<sup>1</sup>Corresponding author. E-mail: yangsz@mail.npust.edu.tw <sup>2</sup>Fooyin University. 151 Chihshueh Road, Taliao, Kaohsiung 831, Taiwan.

processing of ordinal data because this path can maintain the consistency of the analysis and the overall statistic precision. However, very few reports emphasized the importance of a correct treatment of ordinal information in explanatory data analysis in the past decade.

Multivariate analysis is most problematic when ordinal variables appear together with other scale types in the data. However, this obstacle can be solved by extending Gower's general coefficient of similarity to ordinal data types, facilitating cluster and multidimensional scaling (Podani 1999). Ordinal cluster analysis (OrdCIAn) proceeds in the same manner as agglomerative clustering method. Non-metric multidimensional scaling (NMDS) (Kruskal 1964) relies on the ordinal information and represents a good alternative to the metric procedures (Podani 2000, 2005). In NMDS, any change in the dissimilarity matrix of objects have no impact on the final ordination that the first two axes usually provide a fairly good representation of objects or variables. Gordon (1999) suggested that the name ordinal scaling is more appropriate than non-metric scaling. These two methods are order invariant and this property ensures that the results are independent of changes not affecting ordinal relationships (Podani 1999, 2005, 2006).

In Taiwan, three species of *Entada* Adans. (Fabaceae), namely *E. koshunensis* Hayata et Kanehira, *E. phaseoloides* (L.) Merr. and *E. rheedii* Spreng. occur (Yang *et al.* 2005), of which *E. koshunensis* is endemic. *Entada* is frequently found among riverine vegetation and the segments of its pods are dispersed by water. The terminal pinnae of the *Entada* species are modified into tendrils and the stems are often more than 30 cm in diameter. The plant can grow extremely quickly towards the light and has the ability to entangle trees (Nielsen 1992). As the leaflets and flowers of the three *Entada* species are very similar in shape and the specimens preserved in Taiwan herbaria are few and incomplete, these characteristics are the factors that obviously influence academic decisions in identifying the species. The aims of the present study are to provide a primary mixed data matrix (Podani 2001) and use multivariate analysis by OrdCIAn and NMDS to affirm the results of previous taxonomic treatments (Ho 1985, Huang and Ohashi 1993, Yang *et al.* 2005).

## Materials and Methods

Sixty-two voucher specimens of *Entada* species deposited in three herbaria of Taiwan, namely Provincial Pingtung Institute (PPI), National Taiwan University (TAI), Taiwan National Museum (TNM), were used and the same local specimens were combined to represent the fourteen operational taxonomic units (OTUs). Twenty-nine morphological characters (character states, variables) were selected to distinguish the three species of the genus (Table 1). OTUs 1-5 were the species identified as *E. koshunensis*, 6-10 identified as *E. phaseoloides*, and 11-14 identified as *E. rheedii*. The characters were measured on binary nominal, multistate nominal, ordinal, and ratio

scales. The binary nominal data type involves a categorization without numerical values or ranks. It involves two states, such as presence/absence of data or black/white coded as 1/0. The multistate nominal data type involves three states or more, such as red/blue/black. The ordinal data type can be placed in rank order along a continuum. Some of the abundance scale for recording vegetation data belongs to the ordinal data type. The ratio data type includes real quantitative values, and thus, the differences between values of objects can be compared (Podani 1994). For example, the presence of

**Table 1. Four scales describing 29 character states of the genus *Entada* in Taiwan. B: binary nominal, M: multistate nominal, O: ordinal, Q: quantitative.**

Character states	Data type	Values or scales	Character states	Data type	Values or scales
1. leaflet midrib, petiolelet hairs	B	0. absent 1. present	14. leaflet pairs no.	Q	no.
2. seed surface sculpturing	B	0. absent 1. present	15. leaflet thickness	Q	mm
3. seed surface luster	B	0. absent 1. present	16. leaflet length	Q	mm
4. seed around furrow	B	0. absent 1. present	17. leaflet width	Q	mm
5. inflorescence axis hairs	B	0. absent 1. present	18. leaflet petiole length	Q	mm
6. leaflet base	B	0. regular 1. oblique	19. pinnule rachis length	Q	mm
7. inflorescence	B	0. spike 1. raceme	20. inflorescence length	Q	mm
8. seed shape	B	0. round 1. elliptic	21. petal length	Q	mm
9. endocarp texture	B	0. parchment-like 1. woody	22. sepal length	Q	mm
10. seed color	B	0. brown 1. blackish-brown	23. stamen length	Q	mm
11. seed surface	B	0. flat 1. convex	24. legume section length	Q	mm
12. leaflet shape	M	1. oblique-ovate 2. obovate-elliptic 3. oblique-elliptic	25. legume section width	Q	mm
13. leaflet apex	O	1. mucronate	26. endocarp thickness	Q	mm
		2. mucronate-emarginate	27. seed length	Q	mm
		3. emarginate	28. seed width	Q	mm
			29. seed thickness	Q	mm

hairs is binary nominal (0 = absent, 1 = present); the outline of leaflets is multistate nominal (1 = ovate, 2 = oblique-ovate, 3 = obovate-elliptic, 4 = oblique-elliptic); the leaflet apex is ordinal (1 = mucronate, 2 = mucronate-emarginated, 3 = emarginate); and leaflet pairs, inflorescence length, petal length, legume section length or width, seed thickness, and endocarp thickness are ratio scale. The ratio scale variables are obtained as

**Table 2. The primary mixed data matrix composed of 29 character states and 14 operational taxonomic units (OTUs) of Taiwan *Entada* genus. States 1-11: binary nominal, 12: multistate nominal, 13: ordinal, and 14-29: quantitative.**

State	OTU													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	0	0	0	0	0	0	0	0	0	0	1	1	1	1
2	1	1	1	1	1	0	0	0	0	0	0	0	0	0
3	1	1	1	1	1	0	0	0	0	0	1	1	1	1
4	1	1	1	1	1	0	0	0	0	0	1	1	1	1
5	0	0	0	0	0	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	0	1	1	1
7	1	1	1	1	1	0	0	0	0	0	0	0	0	0
8	1	1	1	1	1	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	1	1	1	1
10	1	1	1	1	1	0	0	0	0	0	0	0	0	0
11	1	1	1	1	1	0	0	0	0	0	0	0	0	0
12	1	1	1	1	1	1	1	2	1	1	2	3	3	3
13	1	1	1	2	2	2	1	3	2	1	2	2	3	3
14	3	3	3	3	3	2	2	2	2	2	4	5	4	5
15	0.3	0.3	0.3	0.3	0.3	0.4	0.3	0.2	0.4	0.2	0.2	0.2	0.3	0.1
16	57.5	56.0	65.0	41.0	61.0	72.5	46.5	39.0	72.5	61.5	45.0	39.0	50.5	44.0
17	37.5	25.0	27.5	23.5	26.5	40.5	22.0	18.0	32.5	28.5	22.5	14.0	23.0	16.5
18	5.0	3.9	5.5	6.0	3.7	4.5	3.7	1.8	3.5	5.0	2.2	2.5	2.3	1.8
19	35.0	85.0	83.0	60.0	62.5	73.0	51.5	60.0	60.5	64.0	95.0	90.0	81.0	101.0
20	200.0	180.0	195.0	230.0	165.0	215.0	195.0	109.0	185.5	290.0	185.0	205.0	225.0	132.5
21	2.0	2.5	3.1	3.9	2.4	3.2	3.1	1.6	3.2	2.5	2.9	2.6	2.5	2.3
22	1.0	1.1	1.1	1.9	1.1	1.1	1.1	1.5	1.1	1.3	0.9	0.8	1.1	0.8
23	4.5	3.9	2.6	3.1	4.3	6.0	2.6	8.0	8.0	4.2	5.7	5.9	5.3	4.0
24	55.0	50.0	54.5	52.0	45.0	93.5	41.0	74.5	74.5	103.0	72.0	58.5	68.5	67.0
25	40.0	61.5	59.0	61.5	54.0	77.0	54.5	91.5	90.0	56.5	81.0	71.5	34.0	33.0
26	0.5	0.3	0.3	0.3	0.3	0.3	0.3	1.1	1.1	0.4	3.4	4.5	3.2	3.8
27	40.5	38.2	38.0	46.0	46.7	61.5	38.0	54.0	71.0	55.0	41.5	40.0	36.3	34.2
28	34.0	37.9	37.0	43.0	43.5	52.0	37.0	50.0	50.0	64.0	39.0	41.0	36.2	35.1
29	18.2	18.1	18.0	19.0	19.9	23.5	18.0	15.0	15.0	18.0	20.0	12.0	16.7	15.1

an average from five replicate samples from the same OTU. A primary mixed data matrix was made by 14 objects (OTUs) and 29 character states (Table 2) for the explanatory analysis.

Gower's formula was used to calculate the primary mixed data into dissimilarities, which were in turn subject to NMDS and the application of OrdClAn (program SYNTAX 2000, Podani 2001). Single link (SL, nearest neighbor) method was selected because SL possesses the property of order invariance (Hubert 1973, Boberg and Salakoski 1993). A dendrogram of ordinal cluster analysis of Taiwan *Entada* taxa based on 29 character states of 14 OTUs has been constructed. For evaluating the correlation between the dendrogram and the distance matrix, the cophenetic correlation coefficient was measured: the higher the correlation, the better the representation of distances in the hierarchy. If the value is larger than 0.9, the correlation is high, if the value is lower than 0.74, the correlation is not significant (Sneath and Sokal 1973). The results of NMDS will show the first two most important dimensions about the OTU distribution pattern and measure final stress value. The stress value of zero indicates perfect fit of the ranked ordination distances to the original distances.

### Results and Discussion

The dendrogram obtained by OrdClAn is presented in Fig. 1. Three groups are recognized from the threshold line between the dissimilarity values of 0.288-0.378. The first group is composed of the OTUs 1-5, the second group comprises OTUs 6-10, and the third group is composed of the OTUs 11-14. The cophenetic correlation coefficient of 0.933 shows a significant correlation between the dendrogram and the distance matrix.

The NMDS solutions for the first two dimensions are shown in Fig. 2. The final stress of 0.0304 indicates that the rank order of distance in the new space follows the original rank order as closely as possible. The points representing *E. koshunensis* are distributed along the middle-lower side; points in the upper-left of the ordination are *E. phaseoloides*; points in the upper-right of the ordination are *E. rheedii*. This OTU ordination (Fig. 2) is very close to the dendrogram obtained by OrdClAn (Fig. 1).

Given the few and incomplete herbarium specimens of *Entada* in Taiwan herbaria, there are controversial opinions concerning the taxa of the genus in Taiwan (Ho 1985, Huang and Ohashi 1993, Yang *et al.* 2005). The results from OrdClAn and NMDS indicate that three groups can be clearly distinguished from one another and the endemic species *E. koshunensis* is indeed different from the others and truly exists in Taiwan. This result is close to the results of Yang *et al.* (2005).

OTU 8 is composed of three specimens, which were collected from Mainland China and submitted to the TNM herbarium during the exchanging process. The point of OTU 8 is distributed in the upper-left near the second group (*E. phaseoloides*) (Fig. 2). Indeed, these three specimens collected from Mainland China have some different characteristics,

such as leaflet size and shape, however, the evidence is not clear enough to be able to separate OTU 8 from the second group. The difference of the characteristics may be influenced by the geographical location or the environmental conditions of the microhabitat. We suggest that these three specimens belong to the species *E. phaseoloides*. Nielsen (1992) showed the ranges of various species of *Entada* in Asia, but

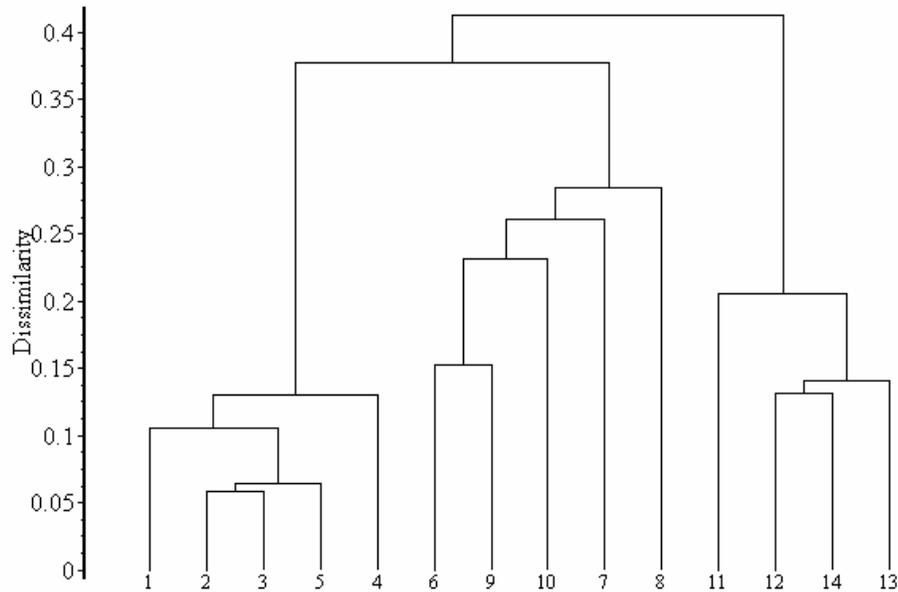


Fig. 1. Dendrogram of ordinal cluster analysis of Taiwan *Entada* taxa based on 29 character states of 14 operational taxonomic units (OTUs). The cophenetic correlation coefficient = 0.933, showing a significant correlation between the dendrogram and the distance matrix.

in his work *E. phaseoloides* was not distributed in Taiwan. Despite few different characters among five OTUs of the second group (Fig. 2), the present study confirms that the range of *E. phaseoloides* includes the Taiwan area.

Deciding on an appropriate correct treatment of ordinal information is still complicated by contrasting opinions presented in the literature (e.g. Podani 2006, van der Maarel 2007). However, the dissimilarity coefficient should be compatible with ordinal variables and the subsequent ordination of clustering methods should consider only the rank order of dissimilarities (Podani 2005). The idea of O-O-O sequence is very important for multivariate exploratory analysis and extending Gower's general coefficient of similarity to OrdCIAn and NMDS will meet this excellent path and obtain good results. According to the OrdCIAn and NMDS results, it is obvious that Taiwan *Entada* surely includes three independent taxa. *Entada koshunensis* truly exists in the fragmented land of Hengchun Peninsula with few populations, *E. phaseoloides* is widely distributed

in central Taiwan and more abundant than the other two species, and *E. rheedii* is rare just found in southern Taiwan.

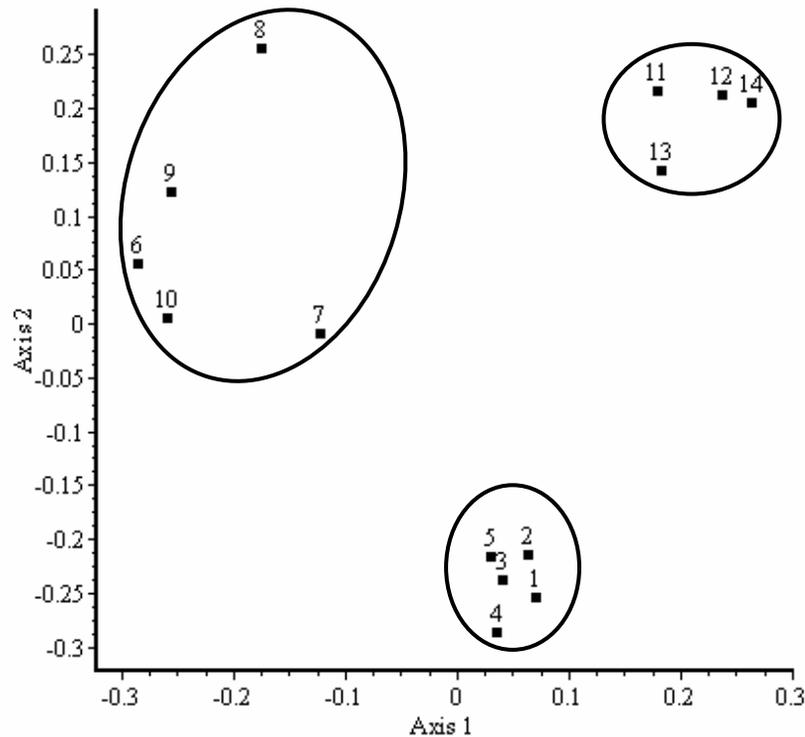


Fig. 2. Non-metric multidimensional scaling, for dimensions 1 and 2, of Taiwan *Entada* taxa based on 29 character states of 14 operational taxonomic units (OTUs). OTUs 1-5: *E. koshunensis*, 6-10: *E. phaseoloides*, and 11-14: *E. rheedii*.

### Acknowledgements

The authors are grateful to the herbaria of PPI, TAI and TNM for loaning the specimens. They express their appreciation to Mrs. M. Willis for the manuscript revision. They also thank the anonymous reviewer for his comments. Most of all, the authors really appreciate Dr. J. Podani for discussions and comments on methodological sequence.

### References

- Anderberg, M.R. 1973. Cluster analysis for applications. John Wiley, New York, USA, pp. 1-359.
- Boberg, J. and Salakoski, T. 1993. General formulation and evaluation of agglomerative clustering methods with metric and non-metric distances. *Pattern Recogn.* **26**: 1395-1406.
- Dale, M.B. 1989. Dissimilarity for partially ranked data and its application to cover-abundance data. *Vegetatio* **82**: 1-12.
- Gordon, A.D. 1999. Classification. (2nd. ed.) Chapman and Hall, London, UK, pp. 1-252.

- Ho, F.C. 1985. Notes on the genus *Entada* of Taiwan. *J. Taiwan Museum* **38**(1): 75-80.
- Huang, T.C. and Ohashi, H. 1993. Leguminosae. In: The Editorial Committee of the Flora of Taiwan (2nd ed.), *Fl. Taiwan* **3**: 169-171. The Editorial Committee of the Flora of Taiwan, Taipei, Taiwan.
- Hubert, L.J. 1973. Monotone invariant clustering procedures. *Psychometrika* **38**: 47-62.
- Krauth, J. 1986. Classification procedures for ordered categorical data. In: Gaul, W. and Schader, M. (eds.), *Classification as a tool of research*, pp. 249-255. Elsevier, Amsterdam, The Netherlands.
- Kruskal, J.B. 1964. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**: 115-129.
- Nielsen, I.C. 1992. Mimosaceae (Leguminosae- Mimosoideae). In: *Foundation Flora Malesiana. Fl. Malesiana ser. I*, **11**(1): 176-181. Rijksherbarium/ Hortus Botanicus, Leiden University, The Netherlands.
- Podani, J. 1994. Multivariate analysis in ecology and systematics- A methodological guide to the SYN-TAX 5.0 package. The Hague, SPB Academic Publishing bv, pp. 1-316.
- Podani, J. 1999. Extending Gower's coefficient of similarity to ordinal characters. *Taxon* **48**: 331-340.
- Podani, J. 2000. Introduction to the exploration of multivariate biological data. Backhuys, Leiden, The Netherlands, pp. 1-407.
- Podani, J. 2001. SYN-TAX 2000. Computer programs for data analysis in ecology and systematics. User's manual. Scientia, Budapest, Hungary, pp. 1-53.
- Podani, J. 2005. Multivariate exploratory analysis of ordinal data in ecology: Pitfalls, problems and solutions. *J. Veg. Sci.* **16**: 497-510.
- Podani, J. 2006. Braun-Blanquet's legacy and data analysis in vegetation science. *J. Veg. Sci.* **17**: 113-117.
- Sneath, P.H.A. and Sokal, R.R. 1973. *Numerical taxonomy: the principles and practice of numerical classification*. W. H. Freeman and Company, San Francisco, USA, pp. 1-573.
- van der Maarel, E. 2007. Transformation of cover-abundance values for appropriate numerical treatment- Alternatives to the proposals by Podani. *J. Veg. Sci.* **18**: 767-770.
- Yang, S.Z., Hueng, S.H., Huang, C.C. and Chiang, S.S. 2005. The taxonomy of *Entada* genus in Taiwan. *Q.J. Chinese For.* **38**(3): 255-266. (in Chinese with English summary)

*(Manuscript received on 4 September 2007; revised on 17 October 2007)*