TAXONOMY AND MATHEMATICAL METHODS TO EVALUATE THE DIFFERENCES AND SIMILARITIES OF TAXONOMIC GROUPS

Sr. HEMA[1], Dr. K. MANIKYA KUMARI[2]

[1]Lecturer, Department of Botony, St. Joseph's College for Women(A), Visakhapatnam
[2]Head, Department of Botony, St. Joseph's College for Women(A), Visakhapatnam

INTRODUCTION

Taxonomy is the area of the biological sciences devoted to the identification, naming, and classification of living things according to apparent common characteristics. It is far from a simple subject, particularly owing to many disputes over the rules for classifying plants and animals. In terms of real-life application, taxonomy, on the one hand, is related to the entire world of life on Earth, but on the other hand, it might seem an ivory-tower discipline that it has nothing to do with the lives of ordinary people. Nonetheless, to understand the very science of life, which is biology, it is essential to understand taxonomy. Each discipline has its own form of taxonomy: people cannot really grasp politics, for instance, without knowing such basics of political classification as the difference between a dictatorship and a democracy or a representative government and one with an absolute ruler. In the biological sciences, before one can begin to appreciate the many varieties of organisms on Earth, it is essential to comprehend the fundamental ideas about how those organisms are related—or, in areas of dispute, *may be* related—to one another.

Taxonomy in Context

The term taxonomy is actually just one of several related words describing various aspects of classification in the biological sciences. In keeping with the spirit of order and intellectual tidiness that governs all efforts to classify, let us start with the most general concept, which happens to be classification itself. Classification is a very broad term, with applications far beyond the biological sciences, that simply refers to the act of systematically arranging ideas or objects into categories according to specific criteria.

While its meaning is narrower than that of classification, even taxonomy still has broader applications than the way in which it is used in the biological sciences. In a general sense, taxonomy refers to the study of classification or to methods of classification—for example, "political taxonomy," as we used it in the introduction to this essay. Literary critics sometimes refer to a writer's taxonomy of characters. Within the biological sciences, however, the term designates specifically a sub discipline involving the process and study of the identification, naming, and classification of organisms according to apparent common characteristics.

Cladistics and Numerical Taxonomy

Cladistics is a system of taxonomy that distinguishes taxonomic groups or entities on the basis of shared derived characteristics, hypothesizing evolutionary relationships to arrange them in a tree like, branching hierarchy. The expression *derived characteristics* in this definition means that the characteristics that unite two types of organism are not necessarily present in a shared evolutionary ancestor. Rather, they have developed over the course of evolutionary history since the time of that shared ancestor.

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

152

NUMERICAL TAXONOMY

Cladistics, the most widely applied approach to taxonomy, has undergone considerable change since it was introduced by the German zoologist Willi Hennig (1913-1976) in the 1950s. Particularly important has been the marriage of cladistics with another taxonomic idea born in the mid-twentieth century, phenetics, or numerical taxonomy. Introduced by the Austrian biologist Robert Reuven Sokal (1926-) and the English microbiologist Peter Henry Andrews Sneath (1923-), numerical taxonomy is an approach in which specific morphological characteristics of an organism are measured and assigned numerical value, so that similarities between taxa (taxonomic groups or entities) can be compared mathematically. These mathematical comparisons are performed through the use of algorithms, or specific step-by-step mathematical procedures for computing the answer to a particular problem. The aim of numerical taxonomy is to remove all subjectivity (such as the taxonomist's "intuition") from the process of classification. Initially, many traditional taxonomists rejected numerical taxonomy, because its results sometimes contradicted their own decades-long studies of comparative morphological features. Nearly all modern taxonomists apply numerical methods in taxonomy, although there is often heated debate as to which particular algorithms should be used.

Identification, Classification, and Nomenclature

Earlier, taxonomy was defined in terms of its relationship to the identification, classification, and nomenclature of taxa. Let us now briefly consider each in turn, with the understanding that they are exceedingly complex, technical subjects that can be treated here in the most cursory fashion. The process of identification is a particularly complex one. When an apparently new taxon is discovered, a taxonomist prepares an organized written description of the characteristics of similar species, which are referred to as a taxonomic key. Instead of using pictures, which often poorly convey the natural variations in morphological features, taxonomists prefer to use a taxonomic key in written form, which provides much more detail and exactitude.

To put it in colloquial terms, by referring to a taxonomic key, a taxonomist may determine that if an organism "looks like a duck and quacks like a duck, it must be a duck"—only, in this instance, the taxa being compared are much more specific than the common term *duck* and the characteristics much more precisely described. (For one thing, there are several dozen species in the genus *Anas,* which includes all "proper" ducks, and many more species in the family Anatidae, or waterfowl, that are commonly called by "duck names"—including such amusingly named species as the ruddy duck, lack duck, freckled duck, and comb duck.) If there is no already established "duck" that the species in question resembles, the taxonomist may have discovered an entirely new genus, family, order, class, or even phylum.

A taxonomist may use what is called a dichotomous key, which presents series of alternatives much like a flow chart. For example, if the flowers of a sample in question are white and the stem is woody, then (depending on additional alternatives) it could be either species *A* or species *B.* If the flowers are not white and the stem is herbaceous (non-woody), then, presented with another set of additional alternatives, it is possible that the plant is either species *C* or species *D.*

Procedures Adopted by Numerical Taxonomists

Since numerical taxonomy is an operational science, the procedure is divided into a number of

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

153

repeatable steps, allowing the results to be checked at every step.

i) Choice of units to be studied: The first step is to decide what kind of units to study. In numerical taxonomy, the basic unit of study is called the "operational taxonomic unit" (OTU). Thus the OTU can be an individual plant if the taxonomist is studying a single population of plants to find out the range of variations in its characters. Similarly, you may treat an entire population of plants as an OTU if you are studying a single species represented by different populations existing in nature; or the OTU may be different species when genus' is being evaluated. Therefore, in numerical taxonomy, the OTU varies with the material being studied, and this' helps the taxonomists in making an objective study.

ii) Character selection : After selecting the OTU's, it is necessary to select characters by which they are to be classified. By experience, you will lCarr1 that characters which vary greatly amongst the OTU's are clearly more useful in numerical taxonomy; and we know that as many characters as possible may be used.

Preferably a minimum of GO and generally 80 to 100 or more characters are needed to produce a fairly stable and reliable classification. The selected characters have then to be coded or given some symbol or mark.' There are 2 methods of coding taxonomic information.

a) Binary coding or two-state coding-This is the simplest form of coding adopted in numerical taxonomy where the characters are divided into + and -, or as 1 and 0. The positive characters are recorded as + or 1 and the negative characters as - or 0. It is possible to use this method of coding for all characters studied.

In case a particular character is not present in an OTU being examined, the symbol or code NC is used, indicating that there is no comparison for that characters. However, we find that by using this method of coding, we tend to increase our work because there are large variations in the plant, and very often a single character such as colour of flower can be represented in a wide range We can have white, pink, red, yellow and other colours in roses. If we are to use this data in a binary coding, then we will have to use each colour as a character and it would be coded as + or -, as the case may be.

b) Multi-state coding-An alternative method would be to use multi-state coding where a single character can be coded in a number of states, each being represented by a numerical symbol or code (e.g. 1,2,3,4,5, . . , . .) depending on the range of variation. Thus, if we again look at the colour of the rose flower, we can give different codes to different colours such as white = 1 , pink = 2, red = 3. yellow = 4, and so on. Besides qualitative characters such as colour of flower . type of placentation, etc., multistate coding is also useful of quantitative characters such as plant height, leaf length, leaf breadth, and other characters involving measurements. A code is prepared for the range of variation and appropriate symbols are allotted to each unit in the. range.

The data obtained by scoring the characters in the OTU's ?re then presented in a table as a .data matrix giving the OTU's on one side of the table and the codes for different' characters against each OTU. Thus, if one has studied 25 OTU's and has scored 75 characters from each, the data matrix will contain 25 x 75 = 1875 units of information. his kind of large it of information in the data matrix necessitates the use' of computers to help the taxgnomists. to digest the knowledge quickly. It is also important to. remember that computer programmes are based on mathematical equations and- computer language and the data matrix is essential for this purpose. In addition, 'the . next step in

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

154

numerical taxonomy is entirely dependent ;on the data matrix.

Example

Character Matrix	Taxon	1	2	3	4	5	6	7	8	9	10
	A	0	1	1	0	0	0	1	1	1	0
	B	0	0	0	1	1	1	0	1	1	1
	C	0	0	1	0	0	1	0	0	0	1
	D	1	1	0	0	0	1	1	1	1	0

S-value Matrix	Taxon	A	B	C	D
	A	--	0.3	0.4	0.7
	B		--	0.5	0.4
	C			--	0.3
	D				--

REAL-LIFE APPLICATIONS

The Urge to Classify

One might ask what all the fuss is about. Why is classification so important? We attempt to answer that question from a few angles, including a brief look at the lengthy historical quest to develop a workable taxonomic system. But what was the original impulse that motivated that quest? One clue can be found in the Greek roots of the word *taxonomy: taxis,* or "arrangement," and *nomos,* or "law." The search for a taxonomic system represents humankind's desire to make order out of the complexities with which nature presents us. When it comes to the organization of ideas (including ideas about the varieties of life-forms), this desire for order is more than a mere preference. It is a necessity.

THE LURE OF A NEW SPECIES.

When a zoologist or botanist discovers what he or she believes to be a new species, the taxonomic system provides a standard against which to check it—rather as you would do if you thought you had discovered a book that was not in the library. If the "new" species matches an established one, that may be the end of the story—unless the scientist has discovered a new aspect of the species or a new subspecies. And if there is no match in the taxonomic "library," the scientist has discovered an entirely new life-form, with all the grand and terrifying ramifications that may ensue.

The new species might be an herb from which a cure can be synthesized for a devastating disease, or it could be a parasite that carries a new and previously unknown malady. Whatever it is, it is better to know about it than not to know, and though the vast majority of "new" species are not nearly as exciting as the preceding paragraph would imply, each has its part to play in the overall balance of life. Discovery of new species is particularly important when those species are endangered or might be in the process of disappearing even as they are identified.

Simple Example of Phenetic Method

In this example will use the following descriptions of five different species of the genus *Charizaria* to evaluate species variation and develop a matrix of homologous character states within each character. These will be coded in a binary method and entered into a data matrix for this quasi-phenetic analysis. Use the following steps.

Step 1: Examine taxa to be evelated.

Step 2: Code homologous character states within a character as having the same binary code. For

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

155

example, for the character "eye color" if you observe two different states in the group, red and yellow, then all red-eyed taxa would be coded as "0" and all yellow-eyed taxa would be coded as In order to accomplish this step one examines character variation in the target taxa through specimen analysis, descriptions, etc. In this group we have character descriptions of 5 species of *Charizaria*.

Table 1. Character descriptions for five species of *Charizaria*.

Taxa	Character Descriptions
C. aus	Humeralis straight; humeral claw absent; pterium yellow; pterial support present; abdomen green; patellar shield absent; gape width smaller than jaw length; snout rounded; nostrils fused; parietal spines with a bifurcated tip; caudal pyrium absent; body coloration green.
C. bus	Humeralis straight; humeral claw present; pterium yellow; pterial support present; abdomen green; patellar shield absent; gape width smaller than jaw length; snout pointed; nostrils separate; parietal spines with a single tip; caudal pyrium present; body coloration orange.
C. cus	Humeralis straight; humeral claw present; pterium yellow; pterial support present; abdomen green; patellar shield absent; snout pointed; nostrils separate; parietal spines with a single tip; caudal pyrium present; body coloration orange.
C. dus	Humeralis bowed; humeral claw present; pterium yellow; pterial support present; abdomen yellow; patellar shield present; gape width smaller than jaw length; snout pointed; nostrils separate; parietal spines with a single tip; caudal pyrium present; body coloration orange.

Table 2. Character descriptions and binary codes for five species of *Charizaria*. Homologies, characters, and character states derived from information in Table 1.

Character Number	Character	Character States	Character States
1	Humeralis shape:	0=bowed	1=straight
2	Humeral claw:	0=absent	1=present
3	Color of pterium:	0=blue	1=yellow
4	Pterial support:	0=absent	1=present
5	Color of abdomen:	0=yellow	1=green
6	Patellar shield:	0=absent	1=present
7	Gape width relative to jaw length:	0=equal	1=smaller
8	Snout shape:	0=rounded	1=pointed
9	Nostrils:	0=separate	1=fused
10	Parietal spines:	0=one tip	1=bifurcated
11	Caudal pyrium:	0=present	1=absent
12	Body coloration:	0=orange	1=green

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

156

Table 3. Binary Coded Characters and Character States in *Charizaria.* Information derived from homologies, characters, and character states hypothesized in Table 2.

Characters and Character States

Taxa	1	2	3	4	5	6	7	8	9	10	11	12
C. aus	1	0	1	1	1	0	1	0	1	1	1	1
C. bus	1	1	1	1	1	0	1	1	0	0	0	0
C. cus	1	1	1	1	1	0	0	1	0	0	0	0
C. dus	0	1	1	1	0	1	1	1	0	0	0	0
C. eus	0	1	0	0	0	1	1	1	0	0	0	0

Step 3. Construct a matrix of taxa x taxa showing the total number of differences in binary-coded character states for all pairwise comparisons. This represents a distance measure that we will use in our phenetic algorithm.

Table 4. Distance matrix for five species of *Charizaria*; distances derived from comparisons of taxa for homologous characters and states outlined in Table 3..

Taxa	*C. aus*	*C. bus*	*C. cus*	*C. dus*	*C. eus*
C. aus	0	6	7	9	11
C. bus	----	0	1	3	5
C. cus	----	----	0	4	6
C. dus	----	----	----	0	2
C. eus	----	----	----	----	0

 Step 3. Rearrange distance matrix provided in Table 4 such that species most similar are adjacent to one another. The information in the matrix has not changed; however, your calculations outlined in Step 4 will be easier by using a rearranged distance matrix.

Table 5. Rearranged distance matrix for five species of *Charizaria;* data from Table 4.

Taxa	*C. dus*	*C. eus*	*C. bus*	*C. cus*	*C. aus*
C. dus	0	2	3	4	9
C. eus	----	0	5	6	11
C. bus	----	----	0	1	6
C. cus	----	----	----	0	7
C. aus	----	----	----	----	0

Step 3. Now we can generate our phenetic dendrogram by calculating the distances between the most similar taxa and ploting these grahically.

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

157

In this instance *C. cus* and *C. bus* are the most similar taxa in having only one difference between them. Taxa *C. dus* and *C. eus* are second most similar in having only two differences. *Charizaria aus* is the least similar to all remaining taxa with 6 to 11 differences with each. Thus, we group *C. cus + C. bus* (distance = 1) and we group *C. dus + C. eus* (distance = 2).

Having done this we must now calculate the distance between the *C. cus + C. bus* group, the *C. dus + C. eus group,* and *C. aus* to determine similarities of these groupings. The following calculations apply:

C. cus + C. bus group vs *C. dus + C. eus* group	*3+4+5+6=18*	*18/4 = 4.5*
C. cus + C. bus group vs *C. aus*	*6+7=13*	*13/2 = 6.5*
C. dus + C. eus group vs *C. aus*	*9+11=20*	*20/2 - 10*
C. dus + C. eus + C. cus + C. bus group vs C. aus	*9+11+6+7=33*	*33/4=8.25*

In this case study it is clear that the *C. cus + C. bus + C. dus + C. eus* group is more similar to each other than either of the earlier groupings of two taxa are to *C. aus.* Furthermore, the *C. cus + C. bus + C. dus + C. eus* group is least similar to *C. aus.*

We may now construct our phenetic dendrogram showing general similarities among the five taxa based on the series of characters and character states used. This is illustrated below.
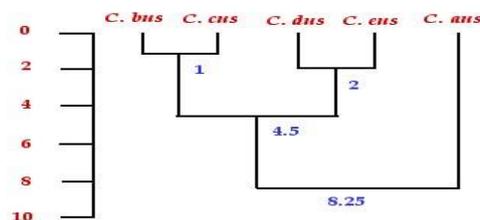
Phenogram showing relative similarity of the five species of *Charizaria* based on characters and character states provided in Tables 1-5

EXAMPLE : 2 PLANTS

Zygophyllaceae is a widespread family of some 27 genera and 285 species subdivided into five subfamilies (Sheahan & Chase 1996; 2000). It consists of herbs, shrubs and trees growing in arid and semi-arid areas in the tropics and subtropics. Earlier studies place the Zygophyllaceae in different orders, e.g. Sapindales, Rutales, Polygalales, Linales, and Geraniales (Cronquist 1968; Takhtajan 1969; 1980; 1983; 1986; Thorne 1992). Soltis *et al*. (2000) put the Zygophyllaceae and Krameriaceae together in their own order Zygophyllales within the eurosid I group, and this position is changed in APG III (2009). They put it within Fabids group. Delimitation of taxa within the family has repeatedly changed over time, because of their diversity in structural detail, particularly in *Balanites, Nitraria*, *Peganum* and *Tetradiclis*. For example, Engler (1896a; 1931) divided the family into seven subfamilies, 8 tribes and 4 subtribes: Peganoideae, Tetradiclidoideae, Chitonioideae, Augeoideae, Zygophylloideae, Nitrarioideae and Balanitoideae. He considered that Zygophylloideae (including the tribuloid genera) formed the central typical group, with Augeoideae based on the characters of the leaves and fruits.

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

158

UPGMA and PCO analysis can be used to study the morphological variation within the family to determine the discontinuities among genera and subfamilies. My results indicate that Zygophyllaceae are heterogeneous, including *Peganum harmala* which has been proposed to form a separate family. There are many splits between subfamilies Zygophylloideae, Tribuloideae, Tetradiclidoideae and Seetzenioideae which seem to be distinct groups. However, there is also some degree of similarity among certain taxa of the subfamilies Balanitoideae, Peganoideae and Zygophylloideae. I consider *Zygophyllum* the most heterogeneous of the Zygophylloideae because I found that the taxa from this genus interspersed with taxa from Peganoideae (*Peganum harmala*), and this is congruent with the results of Sheahan & Chase (2000). Although this study has contributed new conclusions to literature, it is limited to the known genera in Egypt. A comprehensive study covering all *genera* would be necessary to make a more thorough classification and it would be very useful for the further studies to use molecular data.

REFERENCES

1. http://bio.slu.edu/mayden/systematics/bsc420520lect9.html
2. http://www.encyclopedia.com/science-and-technology/biology-and-genetics/biology-general/taxonomy
3. **http://www.citeulike.org/group/15400/article/13078534**
4. http://onlinelibrary.wiley.com/doi/10.1111/j.1095-8339.1973.tb01734.x/full
5. http://onlinelibrary.wiley.com/doi/10.1002/jsfa.2740240305/full
6. http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-33062012000100017
7. *Classification—The Dinosaur FAQ* (Web site). <http://www.miketaylor.org.uk/dino/faq/index.html#4>.
8. *The Germplasm Resources Information Network (GRIN), Agricultural Research Service* (Web site). <http://www.ars-grin.gov/npgs/tax/>.
9. Goto, H. E. *Animal Taxonomy.* London: Arnold, 1982.
10. Lacey, Elizabeth A., and Robert Shetterley. *What's the Difference?: A Guide to Some Familiar Animal Look-Alikes.* New York: Clarion Books, 1993.
11. Margulis, Lynn, and Karlene V. Schwartz. *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth.* New York: W. H. Freeman, 1988.
12. *National Geographic News* (Web site). <http://news.nationalgeographic.com/news/>.
13. O'Neil, Dennis. *Classification of Living Things/Palomar College* (Web site). <http://anthro.palomar.edu/animal/>.
14. Parker, Steve. *Eyewitness Natural World.* New York: Dorling Kindersley, 1994.
15. Simpson, George Gaylord. *Principles of Animal Taxono my.* New York: Columbia University Press, 1961.
16. *Taxonomy Browser,* National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health (Web site). <http://www.ncbi.nlm.nih.gov/Taxonomy/>.
17. *The Tree of Life Web Project* (Web site). <http://beta.tolweb.org/tree/>.
18. Tudge, Colin. *The Variety of Life: A Survey and a Celebration of All the Creatures That Have Ever Lived.* London: Oxford University Press, 2000.

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

159

19. Whyman, Kathryn. *The Animal Kingdom: A Guide to Vertebrate Classification and Biodiversity.* Austin, TX: Raintree Steck-Vaughn, 1999.
20. Willis, Paul. "Dinosaurs and Birds: The Story." Australian Broadcasting Corporation (Web site). <http://www.abc.net.au/science/slab/dinobird/story.htm>.

Proceedings of UGC Sponsored Two Day National Conference on
"RECENT ADVANCES IN MATHEMATICS AND ITS APPLICATIONS" (RADMAS– 2016) 17th&18th November, 2016,
Department of Mathematics, St. Joseph's College for Women (Autonomous) , Visakhapatnam

160