Biometrics, numerical taxonomy and cladistics

NUMERICAL TAXONOMY

Taxonomy deals with description, identification, nomenclature, and classification of living organisms; while as, systematic aims at determining the evolutionary relationships among these organisms.

Classification was largely developed by taxonomists on the basis of a few ―important‖ characters largely determined by the experience of taxonomists gained over decades of research work.

The concept of numerical taxonomy was first proposed by French naturalist **Michel Adanson** in **1763.**

History of numerical taxonomy

Adanson published his Familles des Plantes in 1763 which was the beginning of coming into being of a new and highly original system of classification which later took the shape of numerical taxonomy. This work contained 65 different classifications systems, each based on a single character such as ovary position or inflorescence type.

Numerical taxonomy is also known as **_taximetry**‗. The word evolved to its present form through stages, like **_taxonometrics**‗ and **_taxometrics**‗ (Rogers, 1963). _Taximetrics‗ or _numerical taxonomy‗ uses mathematical methods to evaluate observable differences and similarities between taxonomic groups. It is a method of evaluating data and doesn‗t produce any new data of its own. Numerical taxonomy is often treated as a synonym of phenetics. Phenetics is an attempt to classify organisms based on overall similarity, usually in morphology or other observable traits, regardless of their evolutionary relationship. Numerical taxonomy involves use of multivariant numerical techniques in taxonomy, since itis entirely predicted on the bases of phenetics, the two are often treated as synonyms.

Principles of numerical taxonomy

Sneath and Sokal (1963) outlined seven tenets for numerical taxonomy which are as under

1. Taxonomy should be practiced as empirical science.

2. Classifications should be based on morphological similarities.

3. Classifications based on large number of characters will be more reliable.

4. Every character should be of equal weightage in creating natural taxa.

5. Overall similarity between any two entities is a function of their individual similarity in each of the many characters for which they are being compared.

6. Distinct taxa can be recognized because correlation of the characters differs in the groups of organisms under investigation.

7. Phylogenetic inferences can be drawn from the taxonomic structure of the group and from character correlations given certain assumptions about evolutionary pathways and mechanisms.

Methodology of numerical taxonomy

Classification using numerical taxonomic methods involves the following steps:

1. Selection of taxa: this is the first step in classification; organisms under investigation are usually called operational taxonomic units (OTUs) which may refer to individual organisms, populations, species or genera.
2. Selection of characters: The minimum numbers of characters is 60 for numerically based classification whereas 100 characters are considered ideal. This is followed by description or measurement of these characters. Use of large number of characters necessitates use of computers which in turn makes coding of characters important.
3. Construction of character-taxon matrix: the coding of characters and character-state is presented in a tabular form which is known as character-taxon matrix. In character-taxon matrix, rows represent the character and columns the taxa. Consider, for example, 5 plant species which are to be classified. Suppose we have the following data in which _1' represents presence of a character and _0' represents its absence.
4. Measurement of phenetic similarity: the character states (presence or absence of a character) are compared to know the measure of overall similarity (S) or the phenetic relationship between each pair of OTUs and to determine the taxonomic structure, i.e., the detection of possible groups and sub-groups among all the OTUs. Each character is compared with every other character and results are expressed in percent terms and the data are presented in tabular form wherein 100% _S' represents identical taxa and 0% _S' indicates no similarity at all.

Application of numerical taxonomy

1. Numerical taxonomy is quite useful for taxonomists interested in delimiting similarities and differences between taxa; as such it has been used in many studies involving different classes of bacteria and in many animal groups.

2. Numerical taxonomy has been used in many angiosperms, including Oryza (Morishima, 1969), Solanum (Soria and Heiser, 1961) and Sarcostema (Johnson and Holm, 1968). This method has also been ap-plied to Farinosae of Engler (Hamann, 1961).

3. The results obtained from numerical taxonomic methods have been largely found to be in conformity with the earlier classifications. However, in case of Farinosae the earlier assemblages were shown to be unnatural.

4. Numerical taxonomic studies in case of Ononis from Papilionaceae were found in conformity with the classification provided by Sirjaev (1932).

5. Applying numerical taxonomy to refine classification of grasses, it was concluded that grasses may have evolved from palms. Ovary in both grasses as well as in palms has an ecarpellate origin and seems to have undergone greater reduction in grasses than palms.

Merits of numerical taxonomy

1. The data in numerical taxonomy are collected from more diverse sources such as morphology, physiology, chemistry, amino acid sequences and proteins as compared to the conventional taxonomy. 2. Numerical taxonomic work can be done by less skilled taxonomists.

3. The numerical methods being more sensitive in delimiting taxa, they provide better keys and classifications in comparison to the conventional taxonomic systems.

4. Numerically coded data may be used easily for creation of keys, maps, descriptions, catalogues etc. with the help of existing electronic data processing systems in herbaria and other taxonomic institutions.

5. The quality of conventional taxonomy is improved by numerical taxonomy as it uses higher number of better described characters.

6. Numerical taxonomy has suggested several fundamental changes in conventional principles of taxonomy in general and classification systems in particular.

7. A number of biological concepts have been re-interpreted in the light of numerical taxonomy.

Demerits of numerical taxonomy

1. The method is based on machine algorithms and may not give useful results if the characters chosen for comparison are inadequate.

2. The number of characters needed for satisfactory results which is 100- 400 may pose a difficulty for taxonomists who normally manage classification with far less number of characters.

3. It has been seen that results achieved by mechanical means are not superior to those achieved by practicing taxonomists. A survey conducted by Stearn (1968) indicated that the capacity of computer-aided taxonomic methods to build from an assemblage of characters, a grouping of species, is comparable in validity to one made by a conscious taxonomic effort.

4. Furthermore, numerical taxonomy ignores phylogeny of the organisms being classified.

Cladistic Characters

Any variation between individuals and taxa may be considered as characters to be used in reconstructing phylogeny. Such variation may be morphological, physiological, behavioral, ecological or molecular. For present, palaeontological purposes I will stress morphological variation as characters. Some palaeontologists, who are also **cladists**, additionally use stratigraphic variation of taxa as characters in order to reconstruct phylogeny: this is a more contentious issue to which I will return in another article. Stratigraphy has also been used to choose between equally parsimonious trees, as well as to root the tree, but since these are activities that we do after cladogram/tree construction I will leave these issues until later.

Cladistic Characters:

The characters have developed from the cladistic approach to classification attempting to determine branching sequence of evolution and base a classification upon them. Only derived character states are regarded as significant cladistically.

Characters are Primitive vs. derived character states; or as synonyms, general vs. unique; generalized vs. Specialised, Primitive vs. advanced, Plesiomorphic vs. apomorphic and Pfesiotypic vs. apotypic etc. Plesiotypic and Apotypic terms were used by cladists like Wiley (1981), Wagner (1983) etc. Shared derived character states between and among the taxa are called synapomorphies (or synapotypies) and shared primitive states are symplesiomorphies (or symplesiotypies).

Dendrogram:

It is branching diagram in the form of tree and depicts degree of relationship.

Phenogram:

It is representation of phenetic relationship.

Cladogram:

It is the depiction of cladistic relationship.

Construction of Taxonomic Groups:

The following points highlight the four major steps involved in the construction of taxonomic groups. The steps are:

1. Operational Taxonomic Units

2. Unit Characters

3. Estimation of Resemblances

4. Cluster Analysis.

1. Operational Taxonomic Units (OTUs):

Operational taxonomic unit is the basic unit in numerical taxonomy. It can be an individual, species, genus, family, order or class. Since the taxonomic units employed in numerical methods are not always comparable to formal taxonomic units, they are termed as operational taxonomic units. Comparison of OTUs of equal rank is always made in numerical taxonomy.

In case of OTUs above the level of individual, adequate representation of various polymorphic forms is essential. For example, when genera are compared, they should be represented by different species. Similarly, when families are compared, they should be represented by different genera and so on.

In traditional numerical taxonomy (Sokal and Sneath, 1963; Sneath and Sokal, 1973), an Operational Taxonomic Unit (OTU) is a term that means "the thing(s) being studied". The definition is intentionally vague. The "thing(s)" could be an individual organism, a named taxonomic group such as a species or genus, or a group with undetermined evolutionary relationships that share a given set of observed characters. It is up to a scientist to specify and justify his or her definition of OTUs in the context of a particular study.

2. Unit Characters:

Unit characters are the characters used in numerical taxonomy. According to Sokal and Sneath (1963), unit character is defined as a taxonomic character of two or more states, which within the study at hand cannot be subdivided logically, except subdivision brought about by changes in the method of coding. Only the phenotypic characters are used as unit characters, e.g. presence or absence of an awn in a grass spikelet.

Types of unit characters:

i. Binary characters:

Those unit characters, which exist in two states are called binary characters, e.g. presence or absence of trichomes. They can be represented by the simplest form of coding, where characters are divided into + and – or as 1 and. 0.

In such cases, the positive characters are recorded as + or 1 and negative characters are recorded as – or 0. In case the organ possessing a given character is missing in an organism, the character is scored NC, which means **—no comparison—**.

ii. Multistate characters:

Those unit characters, which exist in more than two states are called multistate characters. Such characters can be coded into number of states (1,2,3…) corresponding to their range of variation. They may be further of two types:

Qualitative multistate characters:

These characters contain three or more contrasting forms and each form is ranked on equal footing, e.g., flower colour, which can be in any number of states – red, white, purple, yellow, etc.

Quantitative multistate characters:

These characters represent measures of the size on a continuous scale, e.g., length of the leaf, which can be 2cm, 3cm, 4cm, etc., or height of the plant, or amount of pubescence on a leaf, etc.

3. Estimation of Resemblances:

Most phenetic methods involve taxon-to-taxon distance, similarity or dissimilarity measures. Distance and dissimilarity are sometimes treated as the same thing, though a distinction can be made between them. As the name implies, distance and dissimilarity measure increase with dissimilarity between taxa, while similarity measures decrease with dissimilarity.

Thus the resemblance between two OTUs is estimated or measured either:

a. In terms of similarity i.e., percentage of characters in which they agree, or

b. In terms of dissimilarity i.e., percentage of characters in which they do not agree.

4. Cluster Analysis:

Cluster analysis or clustering is a type of multivariate statistical analysis. It is used to group organisms into separate clusters based on their statistical behaviour. The main objective of clustering is to find similarities between organisms, and then group similar organisms together to assist in understanding relationships that might exist among them.

Thus, different OTUs are grouped together on the basis of degree of similarity and these groups of OTUs are called clusters.

(i) Cluster analysis characteristics:

Cluster analysis is based on a mathematical formulation of a measure of similarity.

There are a number of characteristics that can distinguish different approaches to cluster analysis, which include:

a. Numerical, statistical, and conceptual clusters.

b. Agglomerative vs. divisive – Agglomerative methods start with individual taxa and seek to connect them into pairs, etc. in such a way that similarities between pairs or group members are maximized at each level. In contrast, divisive clustering splits are sought in the group of taxa that will in some way maximize the collective phenotype disparity between the two groups formed by the split.

c. Overlapping vs. disjoint clusters.

d. Incremental vs. non-incremental.

e. Flat vs. hierarchical representatives.

(ii) Clustering methods:

Clustering can be achieved in two ways:

a. Monothetic system — This system employs the attributes one at a time. The monothetic method obviously leads to artificial clustering.

b. Considering all their attributes simultaneously. This method gives a natural grouping.

c. Phenetic clustering methods — Phenetic clustering methods have largely been overtaken by cladistic methods in order to relate data to evolutionary history, and they are now rarely employed in taxonomy as they often lead to substantially different clusters when applied to real data.

Cluster analysis:

A large number of numerical techniques can be used to analyze the groups of related OTUs based on high similarity coefficients. These techniques include elementary cluster analysis, clustering by single, complete or average linkage, central or nodal clustering, etc. The groups of similar organisms organized in this manner are termed phenons, which are arbitrary and relative groups.

Taxonomy Codes

Taxonomy codes are administrative codes set for identifying the provider type and area of specialization for health care providers. Each taxonomy code is a unique ten character alphanumeric code that enables providers to identify their specialty at the claim level. Taxonomy codes are assigned at both the individual provider and organizational provider level.

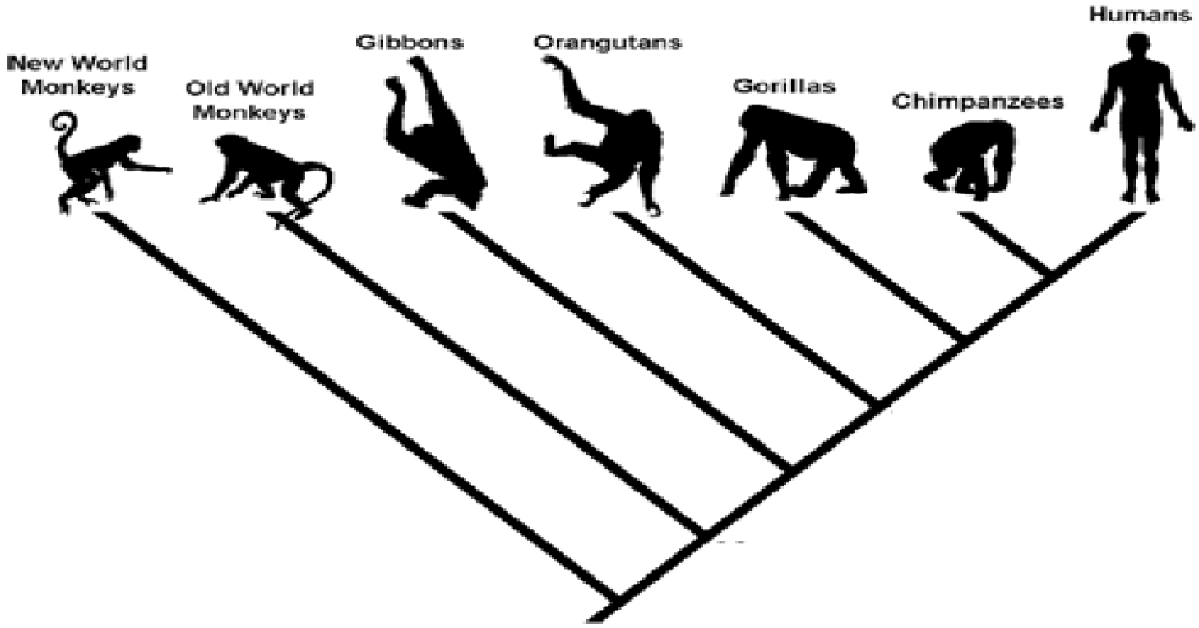
Taxonomy codes have three distinct levels: Level I is the provider type, Level II is Classification, and Level III is the Area of Specialization. A complete list of taxonomy codes can be found within the Health Insurance Portability and Accountability Act (HIPAA) related code list section of the Washington Publishing Company (WPC) web site, at http://www.wpc-edi.com/products/codelists/alert service. If you do not have internet access, you may contact the WPC at 1-425-562-2245 to find out how to purchase a printed code list. Taxonomy codes are self-reported, both by registering with the National Plan and Provider Enumeration System (NPPES) and by electronic and paper claims submission. Taxonomy Codes registered with NPPES at the time of NPI application are reflected on the confirmation notice document received from NPPES with the provider‘s assigned NPI number. Current taxonomy codes registered, including any subsequent changes, may be obtained on an inquiry basis by visiting the NPI Registry Website at https://nppes.cms.hhs.gov/NPPES/NPIRegistry.

Phenograms, cladograms (definitions and differences)

Cladogram

A cladogran is a diagrammatic representation which shows the relationship of the closely related organisms. It is a type of a phylogenetic tree. But it only shows the relationships between clades with the common ancestor. As an example, a cladogram shows human are more loosely related with chimpanzees than gorilla, but it does not show the evolutionary time and the exact distance from the common ancestor.

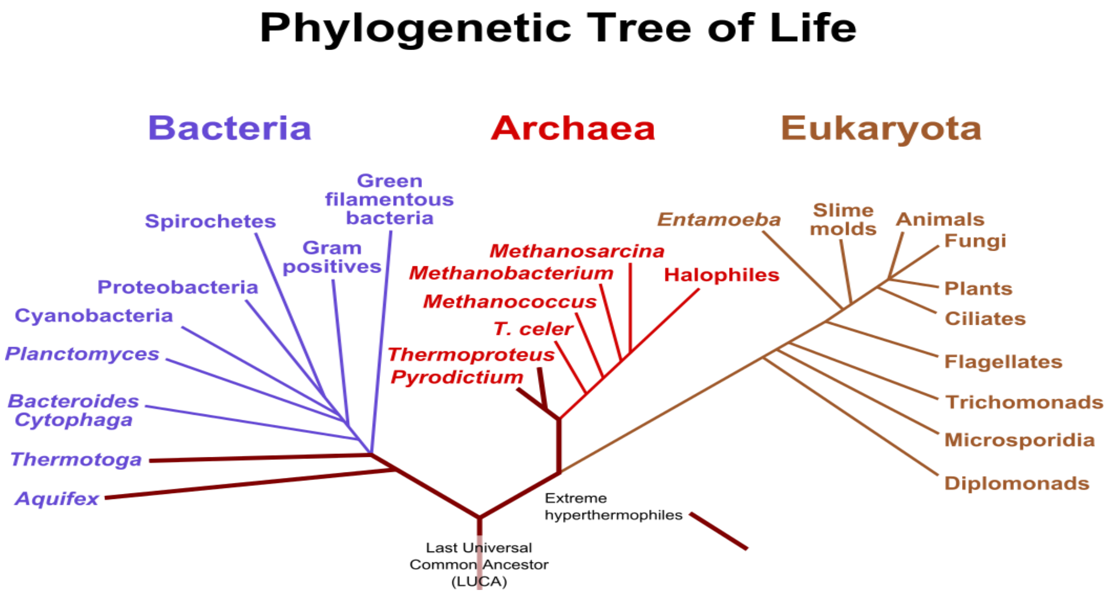
Cladogram is a tree-like diagram which is drawn using lines. The nodes of a cladogram represent the splitting of two groups from a common ancestor. Clades are summarized at the ends of the lines and the members of a particular clade share similar characteristics. Clades are built using molecular differences instead of morphological characteristics. However, cladograms can be constructed using the correct morphological and behavioral data as well.



Phylogenetic Tree

Phylogenetic studies are useful for finding answers to different problems in evolutionary biology such as the relationship between species and their origin, spread of viral infections, migration patterns of species, etc. Advanced molecular biological techniques have helped biologists to evaluate phylogenetic relationships between organisms in relation to the evolutionary changes of the organisms. A phylogenetic tree is a diagram which shows the relationship between organisms based on their characteristics, genetic background, and evolutionary relationships. Compared to a cladogram, phylogenetic tree has more value when discussing the relationships of organisms in a meaningful way with respect to their ancestors and evolution. Phylogenetic tree is drawn like a branching tree diagram in which branch length is proportional to the evolutionary distance, unlike a cladogram.

Biologists analyze different characteristics of organisms using different analytical tools such as parsimony, distance, likelihood and bayesian methods, etc. They consider many characteristics of organisms including morphological, anatomical, behavioral, biochemical, molecular and fossil characteristics to construct phylogenetic trees.



Cladogram	**Phenograms**
Cladogram is not an evolutionary tree. Therefore, it doesn‘t show evolutionary relationships.	Phylogenetic tree is an evolutionary tree. It shows evolutionary relationships.
Cladogram represents a hypothesis about the actual evolutionary history of a group.	Phylogenetic tree represents the true evolutionary history of organisms.
Cladogram is drawn with equal-length. The length of the branch does not represent an evolutionary distance.	Branch length of a phylogenetic tree indicates the evolutionary distance.
Cladogram does not indicate the amount of evolutionary time when separating the organisms‘ taxa.	Phylogenetic tree indicates the amount of evolutionary time when separating the organisms‘ taxa.