

NUMERICAL TAXONOMY

Numerical Taxonomy is an attempt to assign numerical values to the degree of similarity between taxonomic units and to build a classification on this basis. Adanson, a French botanist was the first to give numerical values to the similarity between organisms. Later it was termed as Taximetrics by Sneath & Sokal 1973.

Numerical taxonomy is based on phenetic evidences - similarity shown by observed and recorded characters - not on phylogenetic probabilities. Not only it is empirical but it is operational. This means that it is based on statements and hypotheses so formulated that they can be tested by observation and experiments. Since numerical taxonomy is operational in this sense, it is divided into a series of repeatable steps, allowing its results to be checked back step by step. Sneath & Sokal 1982,

Seven main benefits of Numerical Taxonomy

1. A classification is better if based on more characters and the information content in the taxa is greater.
2. Every character has equal weighting in creating natural taxa.
3. Over all similarity between any two entities is a function of their similarities in each of the many characters upon which they are compared.

4. Each character correlations differ in the group of organisms under study, therefore distinct taxa can be recognised.
5. Phylogenetic inferences can be drawn from the taxonomic structure of a group and from characters' co-rrrelation, given certain assumptions about evolutionary pathways and mechanisms.
6. Taxonomy is practised as an empirical science.
7. Classifications are based on phenetic similarity.

Most of these principles bear resemblance to the aims and methods of Adanson and are therefore known as neo-Adansonian-principles.

The logical steps involved in numerical taxonomy are as follows:

1. **Choice of units to be studied**

The first step is to decide what kinds of units to study - these may be individuals, lines or strains, species etc. It is important that the units should be as representative as possible of the organisms being considered.

The entities of lowest rank in any particular study used are called operational Taxonomic units (OTU's).

2. Character selection

As wide a selection as possible is made of the characters of the OTU's, at least 50 and preferably 100 or more are needed to produce a fairly stable and reliable classification.

The characters have to be broken down into unit character.

A unit character is a character of two or more states which logically cannot be further sub-divided.

The characters selected are then coded. All-or-none is the simplest form of coding whereby a character is divided into two states (+ or -). When no data is available the symbol 0 or NC (No comparison) is used.

Multistate coding is when characters are divided into several states 1, 2, 3, 4, 5 etc. each state representing an equal division of a continuous variable.

All characters selected are treated as equal value and not given special or a priori weighting.

The information is then presented in a t x n table or data matrix consisting of t OTU's scored for n characters.

4.

| Characters | Taxa OTU'S | | | |
|------------|------------|---|----|----|
| | A | B | C | D |
| 1 | + | + | - | NC |
| 2 | + | + | + | + |
| 3 | + | + | + | - |
| 4 | - | + | NC | NC |
| 5 | + | + | + | + |

Coded data table (t x n) table.

3. Measurement of similarity

Overall similarity S is calculated by comparing each OTU with every other and is usually expressed as a percentage, '100' percent S for identity and '0' percent for no resemblance. Thus, similarity S for each pair of organisms examined can be calculated as follows:

$$S = \frac{N_s}{N_s + N_d}$$

where, N_s = No. of positive features shared by any two OTU'S.

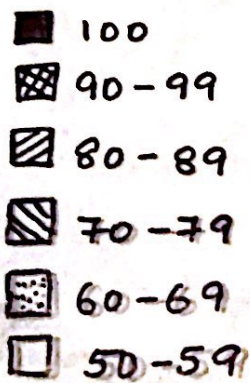
N_d = No. of features positive in one OTU'S and negative in the other.

This can also be termed as co-efficient of association.

A similarity table or matrix is then calculated / constructed tabulating the S co-efficient, one for each OTU.



Fig I



% Similarity.

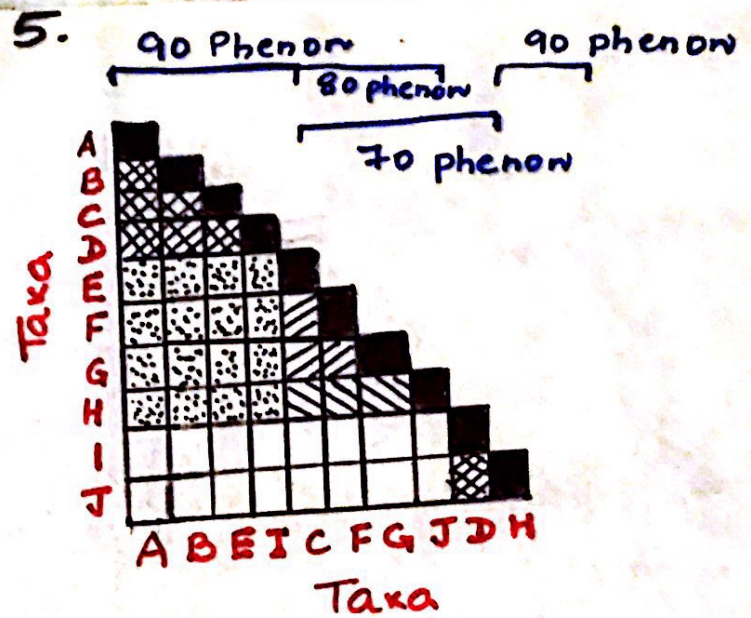


Fig II

4. Cluster Analysis

The similarity matrix is now rearranged so as to bring together into cluster those OTUs whose members have the highest mutual similarity. This allows related taxa or groups to be recognised.. Fig II.

These clusters are called phenons.

5. Discrimination

Having made our classification we can now go back and re-examine the characters to find out which are most constant and therefore most valuable for keying and diagnosis. In other words we have let the computer show us which characters have a posteriori weighing.

Majority of the published applications have been in micro-organisms such as bacteria and viruses. However a hundred numerical classifications of groups of higher plants have been produced.

There is much criticism of Numerical Taxonomy - by those who feel that classifications produced by a computer are limited in value since they rely upon a machine to make automatic calculations - instead of the sensitive judgements of the experienced taxonomists. This seems scarcely justified since the computer performs the time-consuming arithmetic i.e. analysis the similarities and differences of organisms, after the taxonomist using all his skill and preception has decided what characters are to be used.

Numerical Taxonomy are increasingly applied to -

- evaluate phylogenetic trend
- comparative evolutionary rates of organisms.
- sets of characters of organisms etc.