Taxa-Operational Units (OTUs)

- 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.
- The first step in data analysis involves the selection of Taxa for data collection, often called Operational Taxonomic Units (OTUs) in Taxometrics, Operational Evolutionary Units (OEUs) in cladistics, referring to the sample from which the data is collected.
- Although it would be ideal to select different individuals of a population, practical considerations make it necessary to select the members of the next lower rank.
- Thus, for the analysis of a species would need selection of various populations, for the study of a genus they would be different species, and for a family they would be different genera.
- It is not advisable, however, to use genera and higher ranks, as the majority of characters would show variation from one species to another and thus would not be suitable for comparison. The practical solution would be to use one representative of each taxon.
- Thus, if a family is to be analysed and its genera to be compared, the data from one representative species of each genus can be used for analysis.
- Once the taxa are selected, a list of such taxa is prepared. A unique feature of cladistic studies, however, is that the list of taxa generally includes a hypothetical ancestor, the comparison with which reveals crucial phylogenetic information, and is used for rooting of the tree.

Characters

- A conventional definition of a taxonomic character is a characteristic that distinguishes one taxon from another.
- A more practical definition espoused by numerical taxonomists defines character (Michener and Sokal, 1957) as a feature, which varies from one organism to another.

- When selecting a character for numerical analysis, it is important to select a unit character, which may be defined as a taxonomic character of two or more states, which within the study at hand cannot be subdivided logically, except for the subdivision brought about by the method of coding.
- The first step in the handling of characters is to make a list of unit characters.
- A preliminary step involves character compatibility study in which each character is examined to determine the proper sequence of character-state changes that take place as the evolution progresses (morphoclines - A series of morphological transformations that occurs during the evolution of a species., or transformation series).
- The list should include all such characters concerning which information is available.
- A priori, all characters should be weighted equally (no weighting to be given to characters). Although some authors advocate that some characters should subsequently be assigned more weightage than others (a posterior weighting), such considerations generally get nullified when a large number of characters is used.
- It is generally opined that numerical studies should involve not less than 60 characters, but more than 80 are desirable.

For detailed study, please refer to: Singh, G. *Plant Systematics:* Theory and Practice. Oxford & 1131-I Pvt. Ltd., New Delhi.

Binary and multistate characters

A character ("transformation series") is a collection of mutually exclusive states (attributes; features; "characters," "character states," or "stages of expression") which

- a) have a fixed order of evolution such that
- b) each state is derived directly from just one other state, and
- c) there is a unique state from which every other state is eventually derived.

Pimentel and Riggins (1987) defined a character more simply as "a feature of organisms that can be evaluated as a variable with two or more mutually exclusive and ordered states."

- The characters most suitable for computer handling are two-state (binary or presence absence) characters (habit woody or herbaceous).
- However, all characters may not be two-state. They may be qualitative multistate (flowers white, red, blue) or quantitative multistate (leaves two, three, four, five at each node).
- Such multistate characters can be converted into two-state (flowers white or coloured; leaves four or more vs leaves less than four).
- Or else the characters may be split (flowers white vs not white, red vs not red, blue vs not blue; leaves two vs not two, three vs not three and so on).
- Such a splitting may, however, give more weightage to one original character (flower colour or number of leaves).
- It is essential that different character states identified are **discrete** or discontinuous from one another.
- Discreteness of character states can be evaluated by comparing the means, ranges, and standard deviations of each character for all taxa in analysis.
- Additionally t-tests and multivariate analysis may also be used for evaluating character state disreteness.