**Course: Advance Bio Informatics**

**Module Title: Phylogenetic Analysis Stage 3**

**Module No: 57**

**Stage 3: Models of DNA and Amino Acid Substitution**

Phylogenetic analyses rely on models of DNA or amino acid substitution. These models may be implicit or explicit. For distance-based methods, statistical models are employed to estimate the number of DNA or amino acid changes that occurred in a series of pairwise comparisons of sequences. For maximum likelihood and Bayesian approaches, statistical models are applied to individual characters (residues) in order to assess the most likely topology as well as other features such as substitution rates along individual branches. For maximum parsimony, the criterion for ﬁnding the best tree is based on the shortest branch lengths, and while individual characters are also evaluated, many of these statistical models are not applicable.

**Model of Substitution**

A substitution model describes the process from which a sequence of characters changes into another set of traits

**Example**

In cladistics, each position in sequence might correspond to a property of a species

00110 could mean species

0 does not have feathers

0 does not lay eggs,

1 does have fur,

1 is warm-blooded, and

0 no breathing underwater.

11010 would mean

1 species has feathers, 1 lays eggs, 0 does not have fur,

1 is warm-blooded, 0 cannot breathe underwater.

**Usefulness:** Constructing evolutionary trees in phylogenetics or

cladistics. These models can also be used for simulating sequences to test other methods and algorithms.

**Types of Substitution Models:**

**Neutral:** Selection does not operate on the substitutions, and so they are unconstrained.

**Independent:** Changes in one site do not affect probability of changes in another site.

**Finite Sites**:  There are finitely many sites, and so over evolution, a single site can be changed multiple times. Character value 0 at t 0 time t, 0,1, back to 1 or 0

**Jukes Cantor Model:** In circumstances, where substitutions are common; there is no guarantee that a particular site has not been subjected to multiple changes.

This model assumes each nucleotide has the same probability of being replaced by any other

For mathematical formulation of Jukes Cantor Model, please consult the course books.