

# Algorithms as Models of Evolution

## Assessing Phylogenetic Signal:

- ! DNA sequences must be largely free of homoplasy (parallel fixations and reversals).
- ! Transition/Transversion ratios are greater than for saturated sequences.
- ! Closely related taxa will have similar Transition/Transversion ratios.
- ! Saturation of Transitions occurs faster.

## Choosing a method requires preliminary assumptions:

- ! Models of evolutionary process must reflect biological reality.
- ! NOT all sequences or positions in a sequence will reflect phylogenetic history.
- ! Ockham's Razor: Simplicity is always preferred (i.e. most parsimonious).
- ! Errors: Random (or stochastic) and systematic. (More data vs. better assumptions.)

## Most methods *share* common assumptions that:

- ! Nucleotides are evolving **independently** (and are neutrally selected).
- ! Comparison involves **orthologous genes** (i.e. divergent evolution).
- ! **Positional homology** has been inferred correctly.

## Estimating Divergence and weighting (before and after the fact):

- ! Multiple parameter models of nucleotide substitution.
- ! Must justify weighting rationale:           A priori - preconceived notion of evolution?  
  A posteriori - estimate observed homoplasy?

## The major algorithms to be examined: Remember each is a process!

- ! Distance Matrix Methods
- ! Maximum Parsimony
- ! Maximum Likelihood
- ! Method of Invariants

Remember: Optimality vs. Single tree issue from last time.

## Distance Matrix Methods

Distance-based data analysis (as opposed to characters).

### UPGMA & WPGMA:

- ! Gives single tree, not necessarily optimal.
- ! Groups taxa in order of decreasing similarity.
- ! Assumes data are Ultrametric (i.e. constant clock). Weakness!
- ! Both are forms of cluster analysis.

### DeSoete Method:

- ! Yields a single “optimal” tree.
- ! Does not require ultrametric data.
- ! Uses a least squares algorithm while satisfying the 4-point condition with a penalty function.

### Neighbor-joining:

- ! Heuristic approach to estimate a single tree with a minimal overall tree length.
- ! Does not require ultrametric data.
- ! Does require additivity assumption, where the distance between two taxa is equal to the sum of the branches that join them (i.e. 4-point condition).
- ! Keeps track of nodes not taxa.

## Character-Based Methods

### Maximum Parsimony:

- ! Selects the phylogeny with the minimum number of evolutionary changes (i.e., Ockham's razor).
- ! Approach relies heavily upon phylogenetically informative characters (i.e., those with two or more states shared by two or more taxa).
- ! Popularity is due to logical simplicity.
- ! Permits unequal rates and assumes homoplasy is minimal.
- ! Minimizes total tree length, or the number of steps required to explain a given data set.

### Maximum Likelihood:

- ! Calculates probability of data set, given a particular model of evolutionary change.
- ! Independently calculates probabilities at each site, with joint least probability for tree.
- ! Frequently the method least affected by sampling error.
- ! More robust to systematic errors.

### Method of Invariants:

- ! Counts the number of transversion events supporting phylogeny after adjusting for homoplastic change.
- ! Designed for 4-taxon problems where homoplasy is expected to be high (e.g., distantly related taxa with unequal rates of evolution).
- ! Requires a balance between specific classes of transversions.
- ! Relatively long sequences are required, therefore inefficient.
- ! No assumption about rates!

### Resampling techniques – Primarily through Bootstrapping:

- ! Characters randomly drawn with replacement, leading to new data set of original size.

# The Chemical Aspects of The Origin of Life

Life is the cumulative product of interactions among the many kinds of chemical substances that make up the cells of an organism.

! The first life on Earth originated from abiotic surroundings.

! Organic molecules have been successfully generated from abiotic elements.

! The abiotic chemical evolution of life follows 4 major steps:

1. The abiotic synthesis and accumulation of small organic molecules, or monomers, such as amino acids and nucleotides.
2. The joining of these monomers into polymers, including proteins and nucleic acids.
3. The aggregation of abiotically produced molecules into droplets, e.g., protobionts, that had chemical characteristics different from their surroundings.
4. The origin of heredity or information transference.

! To understand how this creation of life from abiotic material occurred, we have to consider 2 critical concepts:

1. The extension of the idea of natural selection to chemical level.

2. The realization that the condition of the early Earth when life first arose must have been vastly different from present:

a) non-oxidizing atmosphere: present level of oxygen, which began to accumulate around 2.1 billion years ago with the presence of cyanobacteria, would have been lethal to primitive organisms

b) abundant resources produced non-biologically

c) long time scale without competition

## The Molecular Clues to the Origin of Life on Earth

! Molecules of living organisms are rich in **hydrogen-containing carbon** compounds. This suggests that there were little or no free molecular oxygen on primitive Earth.

! All **amino acids** exist in both the right-handed state and the left-handed state. However, only 20 amino acids of the left-handed variety are used by living organisms in proteins. Therefore, suggesting that there was one single origin of life.

! **DNA** and **RNA** are the universal informational basis of all life forms on Earth.

! **ATP** is the universal energy currency of all living organisms; suggesting a common origin of metabolism.

! In any cell, first steps of carbohydrate metabolism involve **fermentation**, with the last steps in aerobic organisms the usage of oxygen via **respiration** – suggesting that aerobic organisms evolved from anaerobic ones.