

# PHYLOGENETIC METHODS

# Definition

- . A phylogeny is a hypothetical relationship between groups of organisms being compared. A phylogeny is often depicted using a *phylogenetic tree*, such as the simple one below describing the evolutionary relationships between the great apes.

# Introduction

- **Phylogeny** is discovered using molecular sequencing data and morphological data matrices.
- Greek Origin Phylon-Tribe, Genetico- Relative to birth
- The evolutionary development and history of a species or higher taxonomic grouping of organisms

# Maximum likelihood

- Maximum likelihood is the third method used to build trees.
- Likelihood provides probabilities of the sequences given a model of their evolution on a particular tree.
- The more probable the sequences given the tree, the more the tree is preferred. Likelihood is defined to be a quantity proportional to the probability of observing the data given the model

## Simple Coin Flip example

- The likelihood for heads probability  $p$  for a series of 11 tosses assumed to be independent

HHTTHTHHTTT 5 heads ( $p$ ), 6 tails ( $1-p$ ) Assuming a fair coin what is the likelihood of this series results?

$$L = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = 0.22559$$

- Where  $n$  is the number of tosses,  $k$  is the number coming up heads,  $p$  is the probability of heads (0.5 used above).  $L$  is maximized (0.23609) when  $p=0.45454$ , intuitively  $5/11$ . Thinking of  $L$  as a function dependent on the number of tosses ( $n$ ), number of heads observed ( $k$ ) and the true probability ( $p$ ) of getting heads in a single toss. Since  $n$  and  $k$  are observed you can try various values for  $p$  to find the

# Advantages and disadvantages of maximum likelihood methods

- Supposed advantages.
- Appropriate for simple data like DNA sequences, where we can reasonably model the largely stochastic processes, i.e. a statistical description of the stochastic processes.
- lower variance than other methods

# Disadvantages

- very computationally intensive and so slow
- -Apparently susceptible to asymmetrical presence of data in partitions.
- the result is dependent on the model used and information is derived from sites that are uninformative under parsimony is only due to the model used.



- Neighbor joining is a bottom-up (agglomerative) clustering method for the creation of phylogenetic trees. Neighbor-joining (Saitou and Nei, 1987) is a method that is related to the cluster method but does not require the data to be ultrametric.