Phylogeny and sequence comparison

Basic principles:

- Degree of sequence difference is proportional to length of independent sequence evolution
- Only use positions where alignment is pretty certain – avoid areas with (too many) gaps

Methods to reconstruct phylogenies

1. Distance Methods
   - The genetic distance between each of the sequences is given as the estimated number of substitutions (DNA seq changes) since their last common ancestor. Sequences with smaller genetic distances are considered to be more recently diverged than sequences with greater distances.

Example

Admissible with Traceback

Still optimal, but inadmissible with Traceback

Parsimony

- One of the most popular methods

Idea: Find the tree that explains the observed sequences with a minimal number of substitutions

Two computational subproblems:

1. Find the parsimony cost of a given tree (easy)
2. Search through all tree topologies (hard)
Maximum Likelihood
- Incorporating sophisticated evolutionary models based on what we know about molecular evolution. Statistical analysis. Finds which tree, of all possible trees, has the highest likelihood of giving rise to the observed sequences.
- The maximum likelihood method is the best, because it includes what we know about how molecular evolution happens. However, it is slow (takes a lot of computation time), and becomes too computationally intensive for large datasets (very long, and many, sequences).
- Maximum parsimony is the method most likely to give an incorrect result, especially when there are differences in the rates of evolution between lineages.
- Distance-based methods are fast, and can handle lots of data and are pretty reliable.

Molecular Phylogenetics: Case Studies
The relationships among humans and apes
- Apes: Gibbons (southeast Asia) and great apes (orangutan [southeast Asia]; gorilla, chimpanzee and bonobo [Africa]).
- Humans were given a separate taxonomic group: Homo
- However, this classification is anthropocentric. All evidence (morphological and molecular) indicates that humans belong in the same clade as the African great apes.

Early molecular studies were unable to resolve the relationship of humans to chimp and gorilla.
Possible phylogenies:

The evidence
Phylogenetic tree based on the DNA sequence of a mitochondrial gene.

Many other analyses based on other genes produced the same tree topology indicating that humans and chimp are more closely related than either is to gorilla.
2. The relationship of whales to other mammals

Morphologically, whales, dolphins and porpoises (Cetaceans) are quite distinct from other mammals – made classical phylogenetic analysis difficult.

Traditionally, were classified as close relatives of ungulates (hoofed mammals), specifically as relatives of the artiodactyls (below)

What do the molecular data say?

DNA analysis of genes from different mammals indicate that not only are cetaceans related to artiodactyls, they are artiodactyls.

3. Where did HIV come from?

The genome of the HIV virus is a record of its evolutionary past.

Closely related viruses are found in monkeys (called Simian Immunodeficiency viruses [SIV], though don’t make them sick)

- Have the viruses been around ever since the common ancestor of humans and monkeys, and only recently became a problem because of some change in behaviour? Or is there another explanation for their origin?

- If the virus was present in humans since the common ancestor with monkeys, then we would expect that the topology of the tree of the viruses is the same as that of the host.
The tree of HIV and relatives shows:

1. There is more than one kind of HIV (HIV-1 and HIV-2)
2. Within each kind of HIV, all the viruses do not group together
3. The tree topology does not mirror the known topology of primates.
4. The human viruses are very closely related to monkey and ape viruses

The molecular data indicate that HIV came from a zoonotic transmission from other primates.

HIV1 came from chimp
HIV2 came from Sooty mangabeys.

This is plausible because people living in the area that is the epicentre for each of these infections regularly hunt and eat these primates.