

BIOL2007 - EVOLUTIONARY TREES AND THEIR USES

Reproductive isolation between two species allows them to accumulate different alleles and to diverge. Over evolutionary time, a species can “branch” into 2 species evolving modifications of their phenotypes. Process of **branching and modification** repeated many times has yielded the TREE of life: the pedigree of EXTANT (and EXTINCT) organisms

Figure 2.1 shows relationships among major branches estimated from variation in DNA sequences, especially of the genes encoding ribosomal RNA. Note the 3 empires of life – **Bacteria Archaea Eucarya** (Archaea & Eucarya) pair appears to have the most recent common ancestor.

Today see how evolutionary biologists “**reconstruct**” the tree of life by estimating phylogenetic (genealogical) relationships among organisms. Key questions; Which species share a recent common ancestor? Which species share more distant ancestors?

Phylogenetic analysis is closely linked to classification & naming. Linnaeus grouped species in **HIERARCHICAL** classification of groups **NESTED** within larger groups. This **reflects the real history** of evolution that has produced pedigrees of species, close or distant to varying degree.

How can we use phylogenies in taxonomy?

If species steadily diverge over time, then infer history of branching that gave rise to different taxa by measuring their **degree of similarity** (or **difference**). Analyse patterns of variation in the characteristics of organisms = **CHARACTERS**.

Each **character** can have several **states**. Potentially use character states to determine recent and ancient ancestors – giving phylogenetic tree. Ask where in tree have evolutionary transitions between character states occurred?

Example: **10** variable characters in each of **4** species. **Task:** which species are derived from recent, and which from ancient, common ancestors? Let each character have 2 states, labelled 0 and 1; where **0 = ancestral** state – in common ancestor of group & **1 = derived** state – a change from ancestral state
See diagrams on handout. Figure 2.4A shows **hypothetical** phylogeny of 4 species descended from common ancestor. Each state change shown by tick on branch. A set of species derived from any one common ancestor = **monophyletic group**. So picture shows 3 monophyletic groups of species: (1 + 2 + 3 + 4), (1 + 2 + 3) & (2 + 3).

Imagine that Fig. 2.4A does, indeed, represent, the true phylogeny of the 4 species; can we infer, or estimate, this phylogeny from underlying data? Need to calculate similarity of **species pairs**. How many character states are **shared**?

Method a) **count**, irrespective of whether ancestral or derived state is shared. Eg Species 1 & 2 = score of 4 etc. Find that Species 2 & 3 closest, Species 1 closer to 2 & 3 than to species 4....

Method b) here we know which state is ancestral and which derived. So **can count** only the **shared derived** character states (those that change during ancestry of any species' pair).

Here methods a) and b) yield same outcome matrix.

BUT inferring phylogeny isn't straightforward. Why not?

One difficulty is inequality of evolutionary rates. In Fig. 2.4A **number** of character state changes **similar** from ancestor of group (**Ancestor 1**) to each descendant species. But need not be true. Look at **Figure 2.4B** (handout).

Rate of evolution between ancestor 2 and species 2 greater than in rest of phylogeny. In matrix of shared character states. Species 1 & 3 = most similarities – so most closely related? If restrict comparison to derived character states then species 2 & 3 most similar – most closely related? Why is latter the true indicator of the real phylogeny?

Species 1 & 3: Share **1 derived** character state & **6 ancestral** character states. **Species 2** underwent faster evolution with changes to 4 characters. So species 2 is less similar to species 3 than species 1 is, even though species 2 more closely related to 3

Solution to inequality of evolutionary rates is to use rule:

use derived character states that are shared in order to infer phylogeny successfully

Another difficulty. Thus far, each character changed only once across phylogeny. All species sharing a character state inherited it without change from a common ancestor. But often not true. **Figure 2.4C** (handout) shows **consequences of independent evolution** at 2 or more sites in phylogeny:

Examples of **CONVERGENT EVOLUTION** -

State g1 evolved in species 1 and 3 & State h1 evolved in species 1 and 2.

Example of **EVOLUTIONARY REVERSAL** -

State j1 evolved in ancestor 2 but reverted to j0 in species 2.

Both types of event can provide misleading evidence about phylogeny

Characters **g** and **j** erroneously suggest that Species 1 & 3 = closest relatives: but in fact shared **derived** character states show Species 2 & 3 most similar = most closely related

Character **h** erroneously suggests that Species 1 & 2 = monophyletic group

HENNIG (1966) – established modern inference of phylogenetic relationships. He emphasised that Taxa can be similar because they share **1) uniquely derived character states, 2) ancestral character states, 3) converged character states. ONLY 1) ARE USEFUL FOR TREE-BUILDING**

In the real world how do we **1) identify** which state of a character is derived? **2) determine** whether a character state is uniquely derived or reflects convergent evolution?

METHOD OF MAXIMUM PARSIMONY

One way to handle problems exploits principle of **parsimony**. In phylogenetic analysis: of all possible trees for group of taxa, the **best *parsimonious*** estimate of **true** phylogeny is the one requiring **fewest** evolutionary changes.

EXAMPLE: whales (see handout). Whales and fish share a character, the **dorsal fin** (denoted character 1 on phylogeny) – so **Hypothetical phylogeny A)** is Whales & fish = monophyletic group & Rest of mammals = another monophyletic group.

However whales uniquely share several characters with mammals and with tetrapods so **Conventional phylogeny B)** is that Whales share common ancestor with mammals.

For hypothetical A) to be true then

Features shared by **whales & other tetrapods** would have to have evolved twice: pectoral girdle (**2**), limb skeleton (**3**), lungs (**4**), cervical vertebrae (**5**)

Also, features shared by **whales & mammals** would have to have evolved twice: single aortic arch (6), dentary jawbone (7), milk (8), 4-chambered heart (9)

For conventional B) to be true: only require dorsal fin (1) to have evolved twice

Parsimony favours B)

ANOTHER EXAMPLE: use DNA variation to determine relationships among **3 species of primate**. Assume we have extensive prior evidence they are monophyletic group relative to **2 outgroups** which are more distantly related (e.g. rodent & marsupial). There are **3 possible relationships** between the 3 primate species – which do we choose? (see handout)

Resolve using data matrix of nucleotide bases at 7 sites in homologous DNA sequence from each species. **Which tree needs fewest evolutionary changes?**

Choose a tree. Plot onto it, using parsimony, positions at which each character must have changed.

e.g. **Tree 1, character a:** state in Species 1 & 2 & 3 = A, state in 4 & 5 = C. We can infer there was a change from C to A in the common ancestor of 1, 2, 3. We reject possibility that direction of change was from A to C because requires 2 changes in evolution of species 4 and 5 – **not** parsimonious

Repeat for all characters to complete tree 1.

Repeat process (usually via computer) **for every possible tree**. Somewhat tedious.

Finally, **count number of state changes (“length”) of each tree.**

Here we find that Tree 1 is shortest (8 changes) = our best estimate of the true phylogeny.

This **maximum parsimony** method is logical and relatively easy to use. But it is only one of many methods for inferring phylogenetic relationships. Remember that for many real phylogenies the number of possible trees to be evaluated can be enormous. In turn the outcome may be that the most parsimonious tree is only very slightly different from a group of near-optimal trees. Many researchers argue that it is necessary to examine closely the topology of these to reach an informed judgement.

MORE PHYLOGENETIC METHODS

Parsimony methods have been criticised for failing to use all of the information available – often only small proportions of datasets are used. Some researchers seek to exploit information more efficiently and (often using sophisticated statistical tests) seek to choose best tree among all possible trees.

1) tree-searching programmes – allow assessment of reliability of shortest tree compared to other trees that are almost as short (“near-parsimonious” trees). May use “**bootstrapping**”: run repeated phylogenetic analyses with random subsamples of the character data to find consistent support for a given grouping of species.

Other, complex, tree-searching methods make specific assumptions about how characters evolve and how to join similar taxa together in the search for the best phylogeny - e.g. all nucleotide substitutions are equally likely or e.g. different kinds of substitution occur at different rates.

Maximum likelihood techniques: if have a mathematical model of the probability of different nucleotide substitutions and a possible tree with known branch lengths, then how likely is the observed

data (e.g. particular set of DNA sequences)? Here best estimate of phylogeny, the best tree, is one that maximises this likelihood.

Bayesian methods – calculates probability of observing a specific tree, given the model and the observed data. Here best estimate of phylogeny is one that maximises this probability.

2) Distance-based methods – radically different approach, using computerised “algorithmic” methods, to convert discrete character data into a distance value between species within a phylogeny. General strategy is to cluster taxa with most similar forms close to each other on the tree. This hierarchical grouping method also called a **phenetic** approach.

Several types which vary in their underlying model of character evolution. Example: **Neighbour-joining** method – doesn’t assume equal rates of DNA sequence evolution among lineages. Another example: **Unweighted-pair-group-mean** method – assumes equality of evolutionary rates across tree. See simple example on handout.

Evaluating phylogenetic hypotheses

How to assess validity of our inferred phylogeny?

Does it agree with independent data? If obtain a phylogeny from DNA sequence data, compare it with phylogeny obtained using morphological data.

Example: phylogenetic relationships among higher taxa of vertebrates from DNA sequences are same, with a few exceptions, as those inferred from morphology.

Difficulties in phylogenetic analysis

1) Scoring characters can be problematic.

Example: some mammals have on each side of each jaw, 2 incisors, 1 canine, 3 premolars & 4 molars but other mammals (eg anteaters) have no teeth at all.

1 character difference – loss of teeth?

4 character differences – loss of 4 kinds of teeth?

10 character differences?

2) Convergent evolution and evolutionary reversals are common.

May cause several trees to be equally plausible. If see this pattern, try to get more data to resolve.

3) Rapid divergence of some lineages can allow little opportunity for ancestors of such groups to evolve derived character states.

Example: Bursts of divergence during evolutionary or adaptive radiations where lineages acquire different adaptations. Relationships between families of songbirds are poorly understood.