Evolution (PCB 4674).

Chapter 4. Reconstructing evolutionary trees

Main topics of lecture:
I: Parsimony and phylogeny:

II: The phylogeny of whales:
1.- Are whales closely related to hippopotamuses?
2.- Choosing characters: Morphology and molecules
3.- Alternatives to parsimony: Maximum likelihood and genetic distances

III: Using phylogenies to answer questions:
4.- Classifying the diversity of life
5.- Coevolution
6.- The spread of AIDS

I: Parsimony and phylogeny:

I.1.- The evolutionary history of a group is called its phylogeny. A phylogenetic tree is a graphical summary of this history. The tree describes the pattern, and in some cases the timing of branching events. It records the sequence of speciation and documents which taxa are more closely or distantly related.

I.2.- Because we do not have direct knowledge of evolutionary history, a phylogeny must be inferred indirectly from data !!!!. As its most basic level, the logic of estimating evolutionary relationships is simple: The most closely related taxa should have the most traits in common. We can say that the traits that we can use to reconstruct a phylogeny should be:

i.- Independent of one another
ii.- Heritable
iii.- Variable among the taxa

Examples of characters are DNA sequences, presence or absence of certain skeletal elements or flower parts.

I.3.- In choosing characters to study, however, IT IS ESSENTIAL !!!
to realize that evolutionary relationships are ONLY!!! revealed by traits that are similar because they were derived from a common ancestor. Traits that are shared by a group of taxa because they were modified in a common ancestor are called synapomorphies. Mammals for example share fur and lactation as traits that were derived from their common ancestor (a population from a long-extinct group called the synapsids)

I.4.- Two ideas are key to understand the value of synapomorphies in phylogenetic reconstruction:

i.- Synapomorphies identify evolutionary branch points because after a species splits into two lineages that begin evolving independently, some of their homologous traits undergo changes due to mutation, selection, and drift. These changed traits are synapomorphies that identify populations in the two independent lineages

ii.- Synapomorphies are nested. That is, as you move back in time and trace an evolutionary tree backward from its tips to its root, each branching event adds one or more synapomorphies

I.5.- Figure 13.1 shows how a series of these shared derived character states (that we have defined as synapomorphies) defines phylogenetic relationships among groups of birds that lived during the Mesozoic era. These synapomorphies identify trends in changes in several bones that undoubtedly are adaptations for flight.

**Figure 13.1.: Phylogeny of bird groups from the Mesozoic era.**

**Phylogeny based on skeletal characters.** All the characters depicted in this tree are synapomorphies
I.6.- If similarities among organisms occurred only because of descent from a common ancestor, then inferring a phylogeny for any group of species would be a simple problem of combining progressively larger sets of species together based on their synapomorphies, as in Figure 13.1.

I.7.- Unfortunately, things are not nearly this simple in practice. For example, species can have similar traits because those traits evolved independently in each of the lineages leading to those species. A good example are flippers, this is a derived trait found in both penguins and seals, but these two species did not evolve through modification in the common ancestor of penguins and seals. Instead, the similarity we observe in the forelimbs of penguins and seals results from natural selection that favored the same type of structure in ancestors from very different lineages (birds and mammals). This is called convergent evolution. We can be fairly sure that this similarity is due to convergence because many other traits (other than the shape of the forelimbs) unite penguins with other birds and seals with other mammals. From this point of view, the phylogenetic information from the forelimb characteristics of penguins and seals disagrees with information from many other characters.

I.8.- In addition, derived traits can also revert to the ancestral form due to mutation or selection. Events like these remove similarity caused by descent from a common ancestor. A reversal wipes out the phylogenetic signal and restricts our ability to estimate evolutionary relationships.

I.9.- Convergence and reversal are lumped under the term homoplasy. Homoplasy represents noise in the data sets used to reconstruct phylogenies, and it will arise whenever some characters give conflicting information about relationships among a group of species. Phylogeneticists try to minimize the confusing influence of homoplasy by choosing characters that evolve slowly relative to the age of the groups involved and that shows few instances of apparent convergence or reversal.

I.10.- Parsimony provides one approach for identifying which branching patterns among the many that are possible, most accurately reflects evolutionary history. The number of possible trees that we can obtain grows exponentially:

- For 2 taxa we can have only 1 tree
- For 3 taxa we can have 3 trees
- For 4 taxa we can have 15 trees
- For 5 taxa we can have 105 trees
- For 6 taxa we can have 954 trees
- For 7 taxa we can have 10,395 trees
- For 15 taxa we can have 213,458,046,676,875 trees
Under parsimony, the preferred tree is the one that minimizes the total amount of evolutionary change that has occurred (tree "e" in Fig. 13.A). The reason for invoking parsimony is simple and compelling. In many instances, we can assume that convergence and reversal will be rare relative to similarity that is due to modification from a common ancestor.

**Figure 13.A.: Explanation of parsimony in terms of analysis of character distribution.** (a) Data matrix of six characters (1-6) distributed among four taxa (A-D). Presence of the character is indicated by solid boxes. (b) The three possible resolutions of taxa B-D relative to taxon A. The most parsimonious and optimal solution is depicted in (e)
I.11. The most parsimonious tree will minimize the amount of homoplasy inferred for the data, and this will be the best estimate of the true phylogenetic relationships among the species being studied. In the following sections we use recent analyses of whale phylogeny to illustrate the process of inferring this best phylogeny by parsimony

II: The phylogeny of whales:

1.- Are whales closely related to hippopotamuses?
1.1.- The whales belong to the order Cetacea in the class Mammalia. Taxa of this order share a set of features that are unusual among the mammals, such as loss of the posterior limbs. The relationships between the cetaceans and other mammal groups are difficult to discern because whales are highly evolved for aquatic life. Several studies, however, have placed cetaceans as a relative of the ungulates (superorder Ungulata which includes horses, rhinoceroses, deer, pigs, antelope, and camels) as in Figure 13.2.a. Proponents of this hypothesis suggest that cetaceans are sister to the order Artiodactyla. This is known as the Artiodactyla hypothesis (Fig. 13.2.a)

Figure 13.2.a.: Phylogenetic hypotheses for whales and other mammals. The tree shows the Artiodactyla hypothesis: Whales and dolphins are related to the ungulates, possibly as the sister group of the order Artiodactylis
1.2.- Recent analyses of DNA sequences and other molecular characters have challenged this view of whale relationships. These molecular studies strongly suggest that whales are not merely related to the ungulates but are, in fact, close relatives of one particular group of ungulates: the hippopotamuses (Fig. 13.2.b). This conclusion would upset our traditional understanding of mammalian evolution and classification. For example, it would suggest that some traits of hippos and whales, thought to be *convergent adaptations* for aquatic life, are perhaps shared derived characters.

**Figure 13.2.b.: Phylogenetic hypotheses for whales and other mammals. The tree shows the Whale + Hippo hypothesis: Whales and dolphins are related to the ungulates, possibly as sister group to the hippos**
2. Choosing characters: Morphology and molecules

2.1. The first task of any phylogenetic analysis is choosing which characters to use as data. The phylogeny of whales has been studied using two very different types of characters

   i. Morphological traits (especially those of the skeleton)
   ii. Molecular traits (including DNA restriction sites, and DNA sequences)

2.2. Morphological traits are essential in the case of extinct species found only in museum collections or as fossils. A disadvantage of morphological characters is that scoring a single morphological trait for a group of species may require slow, painstaking work by a highly trained taxonomic expert

2.3. Molecular characters (especially DNA sequences) may be scored rapidly in nearly limitless numbers in different genes, and molecular biologists have develop sophisticated models to predict how sequences change through time. However,
homoplasy in molecular similarities can be difficult to identify, and DNA sequence characters are limited to just four character states (A, C, G, and T). These different features often lead phylogeneticists to use both morphological and molecular characters whenever possible to analyze relationships.

**PARSIMONY WITH A SINGLE MORPHOLOGICAL CHARACTER**

2.4.- The ungulates are traditionally divided into two monophyletic taxa: hippos, cows, deer, pigs, giraffes, antelope, and camels (order **Artiodactyla**), and the horses and rhinoceroses (order **Perissodactyla**). The Artiodactyla are grouped together on the basis of some skeletal and dental characteristics, but most notably by the form of a bone in the ankle called the astragalus. In all artiodactyls, the astragalus has an unusual shape. Both ends of the bone are smooth and pulley-shaped.

2.5.- This shared derived character state is one reason some morphologists have rejected the idea that hippos and whales could be sister groups. If hippos and other Artiodactyla form a monophyletic group, then pulley-shaped astragalus evolved just once without subsequent changes. This trait will be a synapomorphy for the order Artiodactyla.

2.6.- On the other hand, if whales are a sister group to hippos, the origin of the pulley-shaped astragalus was followed by the loss of this synapomorphy in the lineage leading to whales. Based on this character the whale = hippo hypothesis is less parsimonious than the Artiodactyla hypothesis, because it implies one extra step in ankle bone evolution.

**PARSIMONY WITH MULTIPLE MOLECULAR CHARACTERS**

2.7.- John Gatesy and his colleagues (1999) summarized the available molecular evidence on whale relationships to other mammals. They assembled all of the DNA sequence data relevant to this question into a single data set. Their analyses of this data strongly supported the whale + hippo hypothesis.

2.8.- Sixty of those characters for a subset on the original data set are shown in Fig. 13.5. They correspond to nucleotide sites 141-200. We use these eight taxa and 60 characters to illustrate the way that multiple characters are used to infer phylogenies.

*Figure 13.5.: Sequence data for parsimony analysis of whales and members of the order Artiodactylis*
2.9.- Of the 60 characters, 15 group 2 or more taxa. As a result these sites contain **phylogenetic information**. The others are invariant (like site 142, for which all taxa have G) or variable but **uninformative** (like site 192, for which all taxa, have C, except the camel which has G). Site 162 provides a synapomorphy for a clade including the cow, the deer, whale, and hippo sequences (all with T). Site 166 provides a sinapomorphy for a clade consisting of hippos and whales. However **not all informative characters** support the same groupings. Site 177 provides a synapomorphy for a clade of whales, hippos, pigs, and peccaries, excluding the cow and deer. This synapomorphy **conflicts** with site 162, and indicates that reversal or convergence has resulted in homoplasy at one of these sites.

2.10.- We use the information from these characters to choose between alternative phylogenies. To obtain the tree of the Artiodactyla hypothesis (Fig. 13.2.a), 47 nucleotide changes are needed. For the same data the whale + hippo hypothesis implies only 41 changes. Therefore the whale + hippo hypothesis is more parsimonious than the Artiodactyla hypothesis and is, therefore, the preferred tree.

3.- **Alternatives to parsimony: Maximum likelihood and genetic distances**

3.1.- There are other methods to obtain phylogenetic trees. Two of these methods are called **maximum likelihood** and **genetic distances**. The essence
of the maximum likelihood approach for molecular phylogenetics: Given a mathematical formula that describes the probability that different types of nucleotide substitutions will occur, and given a particular phylogenetic tree with known branch lengths, how likely am I to obtain this particular set of DNA sequences?

3.2.- To implement this strategy, a computer program evaluates each tree topology and computes the probability of producing the observed data, given the specified model of character change. This probability is reported as the tree's likelihood. The criterion for accepting or rejecting competing tree topologies, then, is to choose the one with the highest likelihood. Unfortunately, likelihood methods are computationally slow, and very large data sets cannot be analyzed as completely by this technique as they can using much faster parsimony methods.

3.3.- A radically different approach is to convert discrete character data into a distance value. For example, the percentage of nucleotide sites that differ between two taxa (a 10% difference means that an average of 10 nucleotides have changed per 100 bases). As in likelihood methods, distance analyses require the investigator to assume a model of character evolution in order to convert information from the original data set into distance between taxa.

3.4.- To estimate a phylogeny from distance data, computer programs are used to cluster taxa, so that the most similar forms are found close to one another on the resulting tree. The preferred tree is the one that minimizes the total distance among taxa.

3.5.- These two methods (maximum likelihood and genetic distances) use different methods for joining similar taxa together in the search for the best phylogeny. When these very different methods agree on the same best phylogeny, this gives some degree of confidence that the best tree has actually been found.

---

**III: Using phylogenies to answer questions:**

**4.- Classifying the diversity of life**

4.1.- Traditional classifications group organisms according to similar features. This is called a phenetic approach. Phylogenetic systematics argue that classification systems should be tree-based, with names and categories that reflect the actual sequence of branching events.

4.2.- We can distinguish three kinds of taxonomic groups in a phylogenetic tree (Fig 13B)

   i.- A **monophyletic** group contains the common ancestor and all of its descendants

   ii.- A **paraphyletic** group contains the common ancestor and some - but not all - of its descendants
iii.- A **polyphyletic** group contains some of the descendants of a common ancestor but not the common ancestor itself.

**Figure 13.B.: Monophyletic, paraphyletic, and polyphyletic groups.**

Ancestors are represented by the interior nodes of the tree

4.2.- According to the method of phylogenetic systematics only monophyletic groups, which include all descendants of a common ancestor are named. Phenetic and phylogenetic naming can produce conflicts. The relationship between whales and ungulates is one of many recent examples. A phylogenetic classification would nest whales as a subgroup within the order Artiodactyla. This means a reduction of the Cetacea from a high-level taxon (an order) to a low level group (perhaps a subfamily related to hippos!!!).

4.3.- Phylogenetic systematics represent an important break from classical approaches to taxonomy. An increasing number of taxonomists are calling for a complete overhaul of the traditional phenetic scheme, the goal being to create a phylogenetic classification.

5.- **Coevolution**

5.1.- Coevolution is an umbrella term for interactions among species that result in reciprocal adaptation. Predation, mutualism, and parasitism are cases of coevolution. Here, we introduce how phylogenetic thinking is used in coevolutionary studies. Our focal system is ants that farm fungi.

5.2.- The 200 species of ants in the tribe Attini are the dominant herbivores in the New World tropics. The group includes the leaf-cutting ants, which dissect pieces from leaves and carry them to their nests. There, they use the leaf material as a substrate for growing fungi in underground gardens.
The fungi are harvested and serve as the primary food source for the ant colony. This symbiosis, between ants and fungi is thought to be 50 million years old. The relationship is also obligate: To the best of our knowledge, none of the symbiotic and or fungal species can live without the other.

5.3. We would like to answer the following question about this association: If the symbiosis originated just once, did the ants and fungi subsequently evolve in tandem? That is, did the two groups "cospeciate"? If so, we would expect the phylogenies of the two groups to be congruent (match up branch for branch).

5.4. To answer this question Ignacio Chapela et al. (1994) sequenced the 28s rRNA gene in 37 species of ant-associated and free-living fungi and matched the branching pattern inferred from these data to an ant phylogeny previously obtained by other researchers. The two phylogenies are shown in Fig. 13.12.

Figure 13.12: Comparing phylogenies of fungi and their ant symbionts. Names refer to fungi.

5.4. Their analysis show that the first lineage of ants that branch off lead to Myrmicocrypta squamosa. If coespeciation were occurring, the next branch should have lead to Leucoagaricus or Leucoagaricus. Instead it leads to Apterostigma. This means that ants "domesticated" free-living fungal species more than once. The tree shows that early in the evolution of ant-fungi
symbiosis, some ants that farmed fungi switched species, picking up new, free-living fungi to "domesticate". Cospeciation has occurred **ONLY** in the most recently evolved forms

---

**6.- The spread of AIDS**

6.1.- Phylogeny inference methods are being used to understand how AIDS and other diseases spread. In the early 1990s, and HIV-positive dentist in Florida was suspected of transmitting the virus to one of his patients. In response, many of his patients were tested for the presence of the virus. Several tested positive. Did they get the virus from the dentist?

6.2.- To answer the question, investigators sequenced the gp120 gene from viruses in infected patients and infected no patients. The resulting tree clearly demarcated a "dental clade". This defined a group of individuals in whom the gp120 sequences were closely related to gp120 sequences of the dentist, leading to the conclusion that these patients (but not others) had acquired the AIDS virus from the dentist.

*Figure 13.13: Using a phylogeny to determine the history of HIV transmission*