

Introduction to characters and parsimony analysis

Genetic Relationships

- **Genetic relationships exist between individuals within populations**
- **These include ancestor-descendent relationships and more indirect relationships based on common ancestry**
- **Within sexually reducing populations there is a network of relationships**
- **Genetic relations within populations can be measured with a coefficient of genetic relatedness**

Phylogenetic Relationships

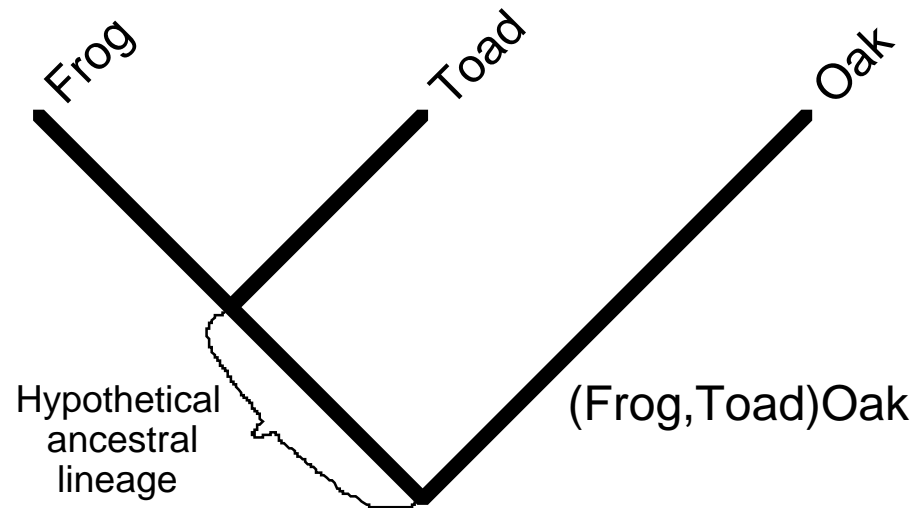
- **Phylogenetic relationships exist between lineages (e.g. species, genes)**
- **These include ancestor-descendent relationships and more indirect relationships based on common ancestry**
- **Phylogenetic relationships between species or lineages are (expected to be) tree-like**
- **Phylogenetic relationships are not measured with a simple coefficient**

Phylogenetic Relationships

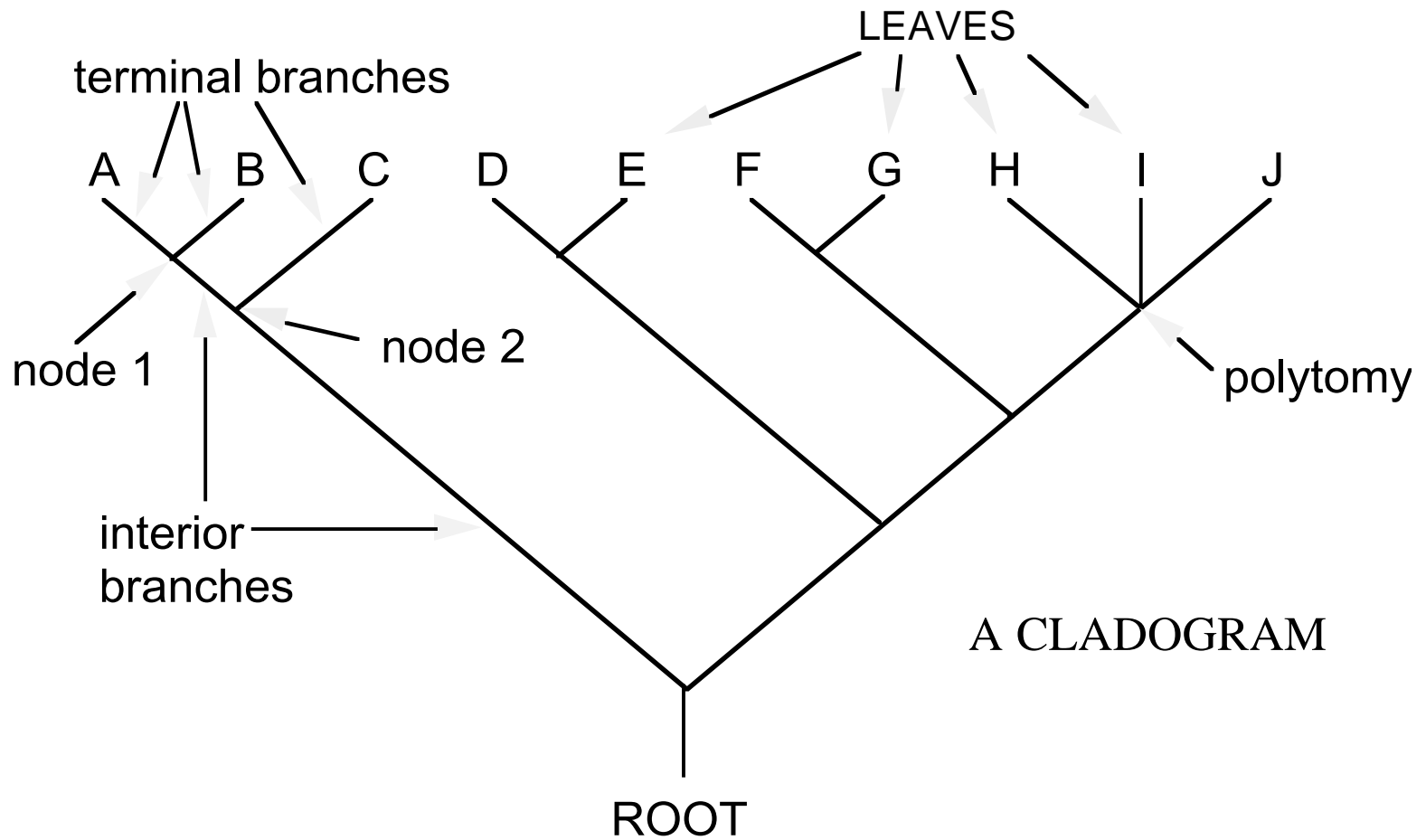
- **Traditionally phylogeny reconstruction was dominated by the search for ancestors, and ancestor-descendant relationships**
- **In modern phylogenetics there is an emphasis on indirect relationships**
- **Given that all lineages are related, closeness of phylogenetic relationships is a relative concept.**

Phylogenetic relationships

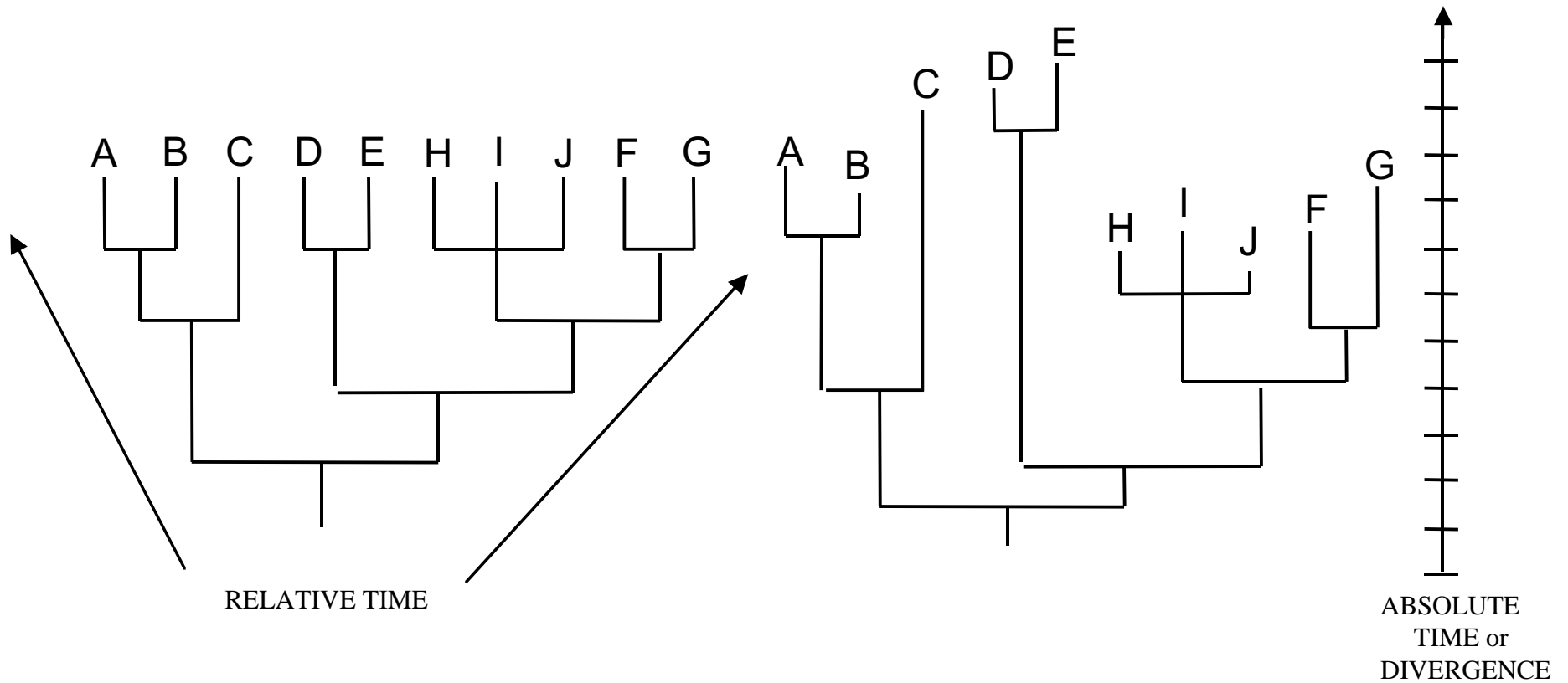
- **Two lineages are more closely related to each other than to some other lineage if they share a more recent common ancestor - this is the cladistic concept of relationships**
- **Phylogenetic hypotheses are hypotheses of common ancestry**



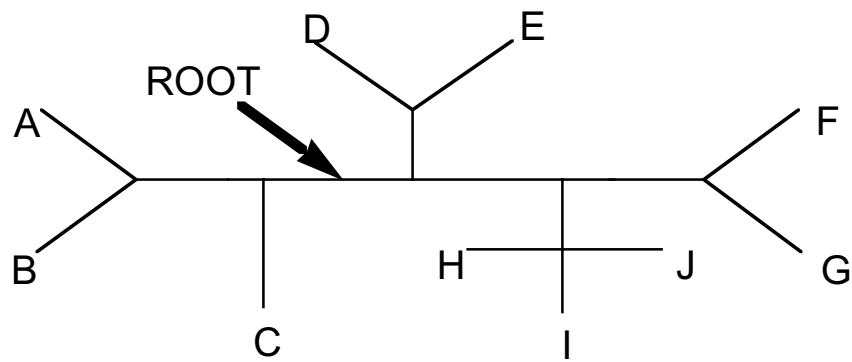
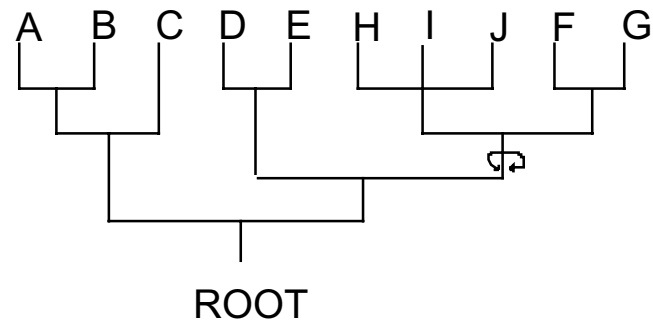
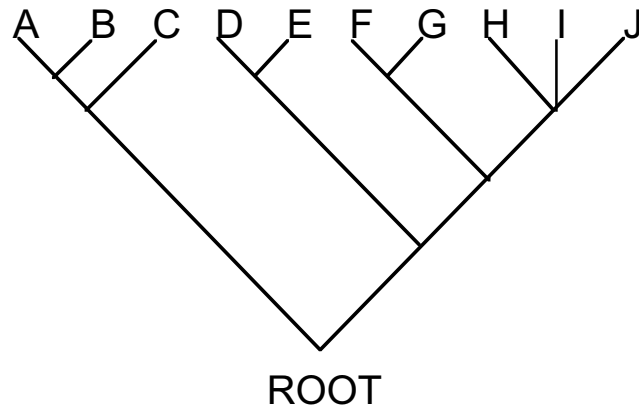
Phylogenetic Trees



CLADOGRAMS AND PHYLOGRAMS



Trees - Rooted and Unrooted



Characters and Character States

- **Organisms comprise sets of features**
- **When organisms/taxa differ with respect to a feature (e.g. its presence or absence or different nucleotide bases at specific sites in a sequence) the different conditions are called *character states***
- **The collection of character states with respect to a feature constitute a *character***

Character evolution

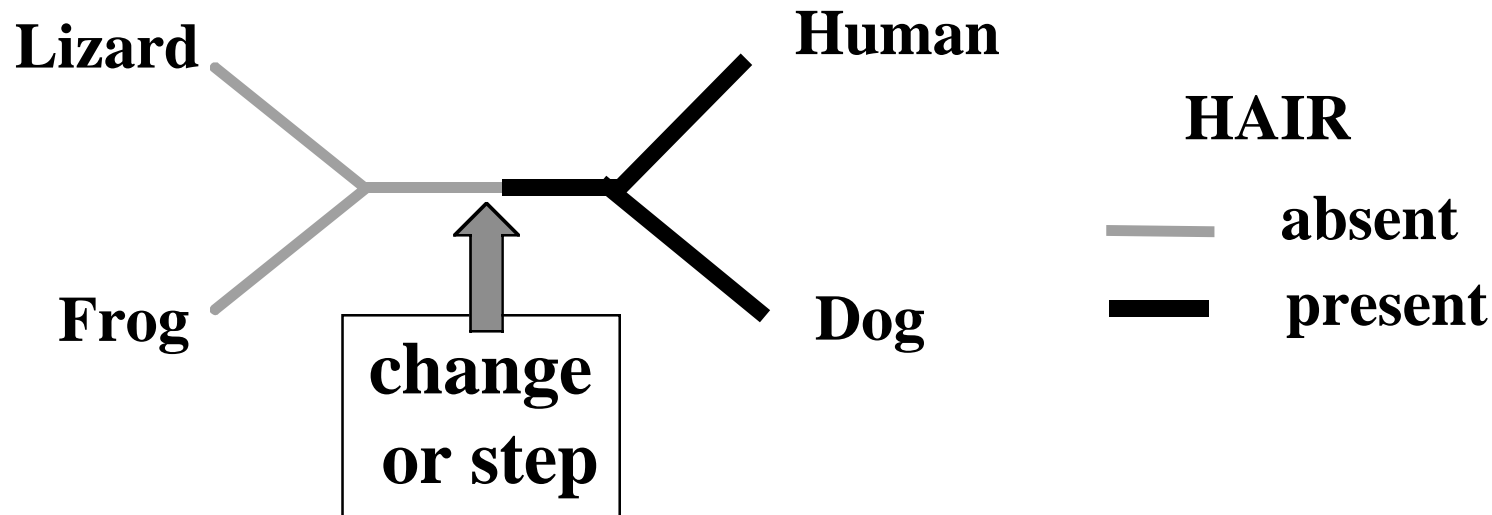
- **Heritable changes (in morphology, gene sequences, etc.) produce different character states**
- **Similarities and differences in character states provide the basis for inferring phylogeny (i.e. provide evidence of relationships)**
- **The utility of this evidence depends on how often the evolutionary changes that produce the different character states occur independently**

Unique and unreversed characters

- **Given a heritable evolutionary change that is unique and unreversed (e.g. the origin of hair) in an ancestral species, the presence of the novel character state in any taxa must be due to inheritance from the ancestor**
- **Similarly, absence in any taxa must be because the taxa are not descendants of that ancestor**
- **The novelty is a *homology* acting as badge or marker for the descendants of the ancestor**
- **The taxa with the novelty are a clade (e.g. Mammalia)**

Unique and unreversed characters

- Because hair evolved only once and is unreversed (not subsequently lost) it is *homologous* and provides unambiguous evidence for relationships

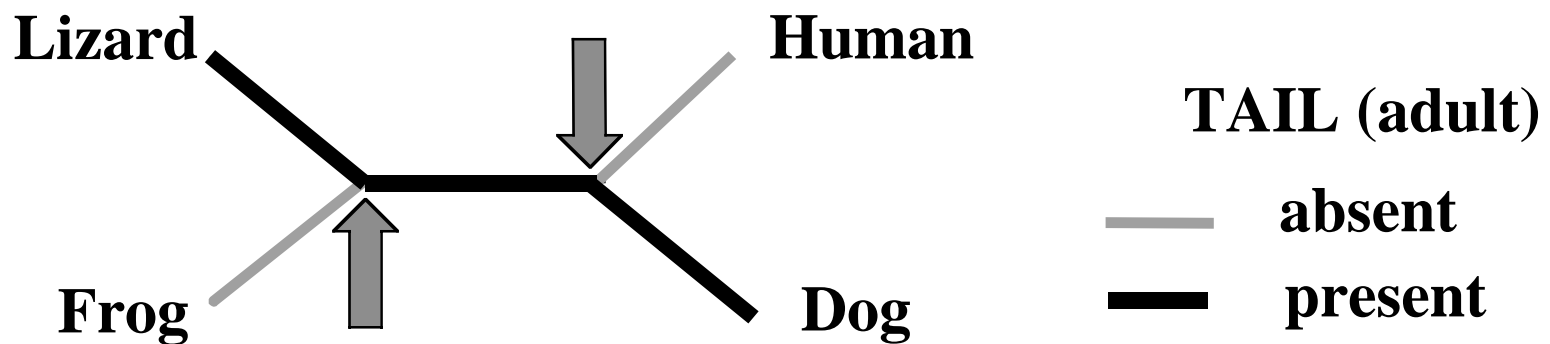


Homoplasy - Independent evolution

- **Homoplasy is similarity that is not homologous (not due to common ancestry)**
- **It is the result of independent evolution (convergence, parallelism, reversal)**
- **Homoplasy can provide misleading evidence of phylogenetic relationships (if mistakenly interpreted as homology)**

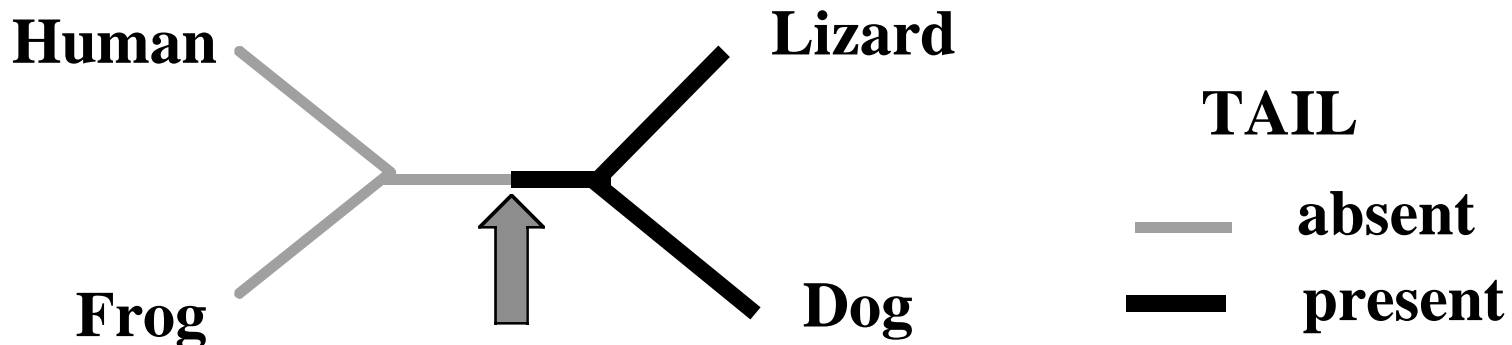
Homoplasy - independent evolution

- **Loss of tails evolved independently in humans and frogs - there are two steps on the true tree**



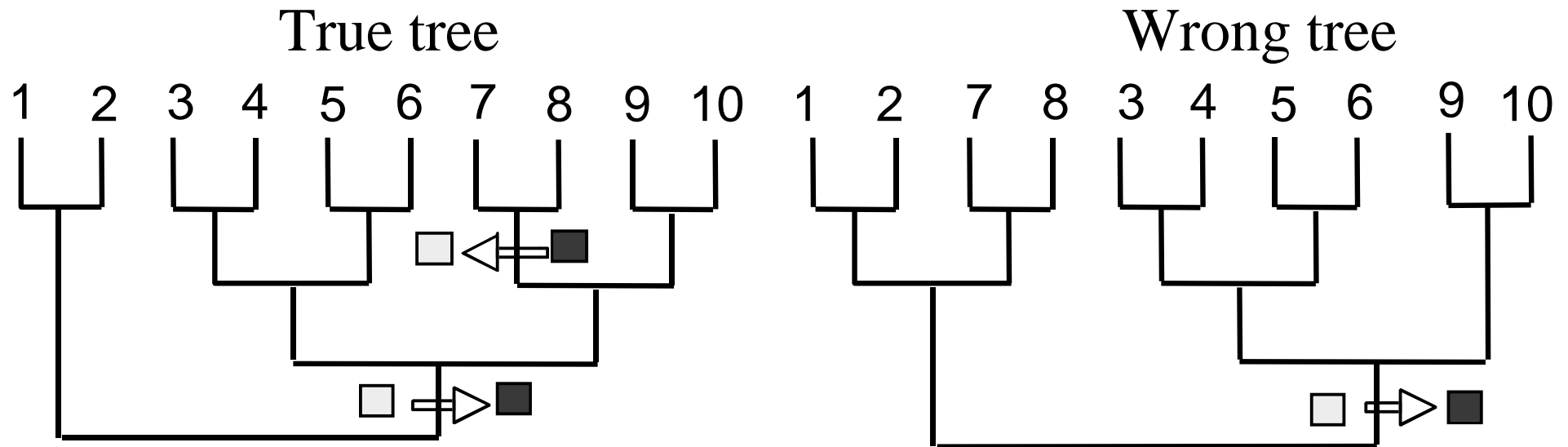
Homoplasy - misleading evidence of phylogeny

- If misinterpreted as homology, the absence of tails would be evidence for a wrong tree: grouping humans with frogs and lizards with dogs



Homoplasy - reversal

- **Reversals are evolutionary changes back to an ancestral condition**
- **As with any homoplasy, reversals can provide misleading evidence of relationships**



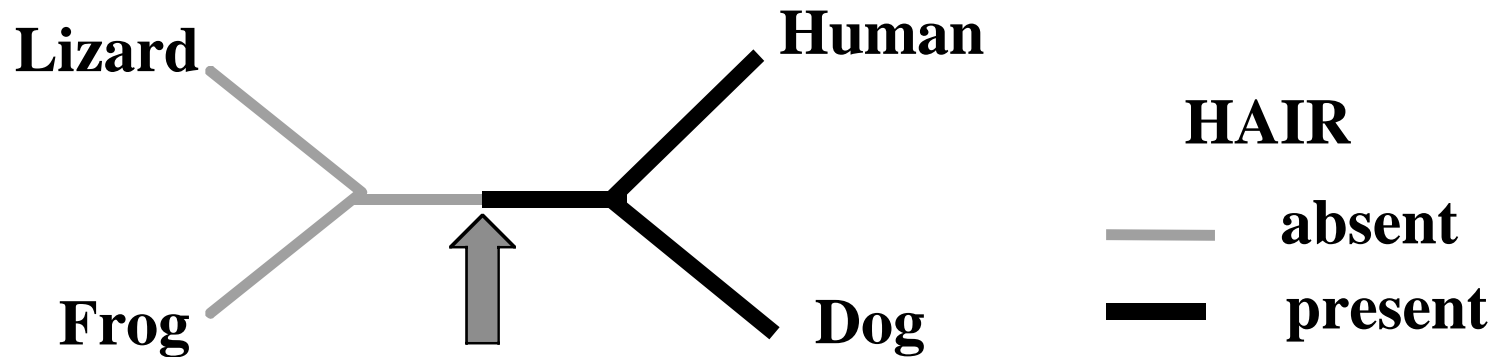
Homoplasy - a fundamental problem of phylogenetic inference

- If there were no homoplastic similarities inferring phylogeny would be easy - all the pieces of the jig-saw would fit together neatly**
- Distinguishing the misleading evidence of homoplasy from the reliable evidence of homology is a fundamental problem of phylogenetic inference**

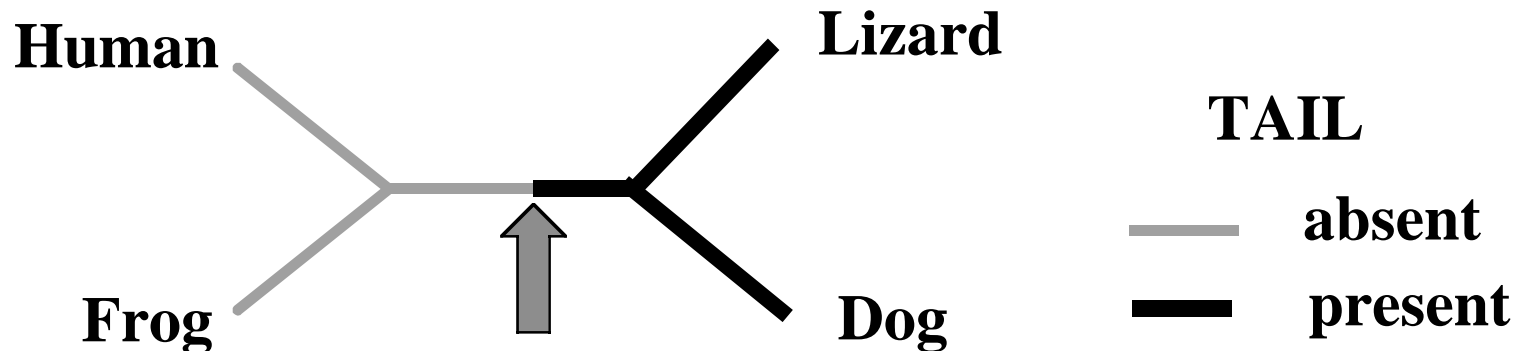
Homoplasy and Incongruence

- If we assume that there is a single correct phylogenetic tree then:**
- When characters support conflicting phylogenetic trees we know that there must be some misleading evidence of relationships among the incongruent or incompatible characters**
- Incongruence between two characters implies that at least one of the characters is homoplastic and that at least one of the trees the character supports is wrong**

Incongruence or Incompatibility



- These trees and characters are incongruent - both trees cannot be correct, at least one is wrong and at least one character must be homoplastic



Distinguishing homology and homoplasy

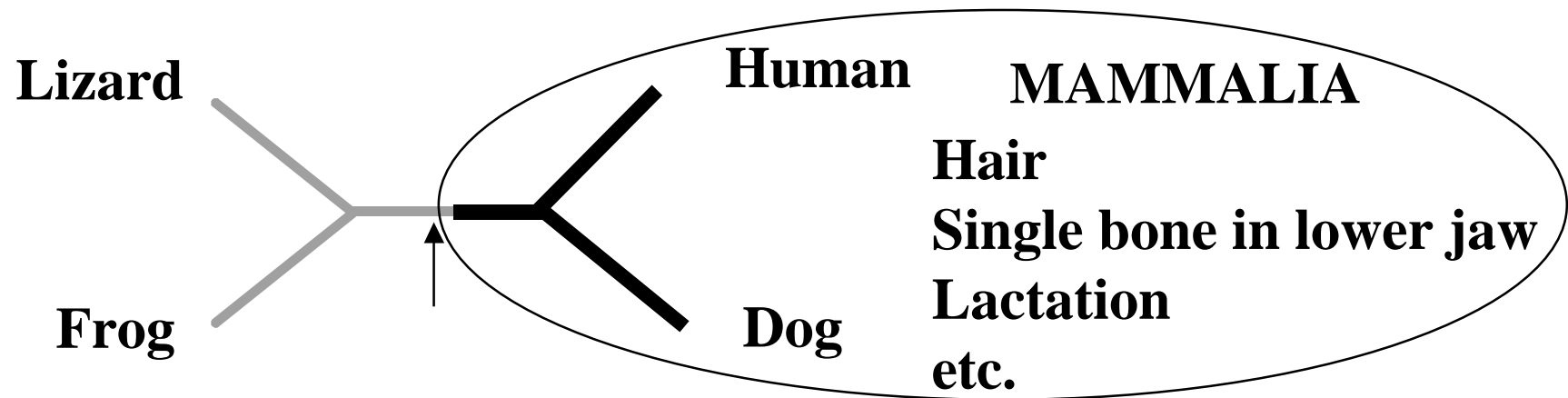
- **Morphologists use a variety of techniques to distinguish homoplasy and homology**
- **Homologous features are expected to display detailed similarity (in position, structure, development) whereas homoplastic similarities are more likely to be superficial**
- **As recognised by Charles Darwin congruence with other characters provides the most compelling evidence for homology**

The importance of congruence

- **“The importance, for classification, of trifling characters, mainly depends on their being correlated with several other characters of more or less importance. The value indeed of an aggregate of characters is very evident a classification founded on any single character, however important that may be, has always failed.”**
- **Charles Darwin: Origin of Species, Ch. 13**

Congruence

- We prefer the 'true' tree because it is supported by multiple congruent characters



Homoplasy in molecular data

Incongruence and therefore homoplasy can be common in molecular sequence data

- There are a limited number of alternative character states (e.g. Only A, G, C and T in DNA)**
- Rates of evolution are sometimes high**

Character states are chemically identical

- homology and homoplasy are equally similar**
- cannot be distinguished by detailed study of similarity and differences**

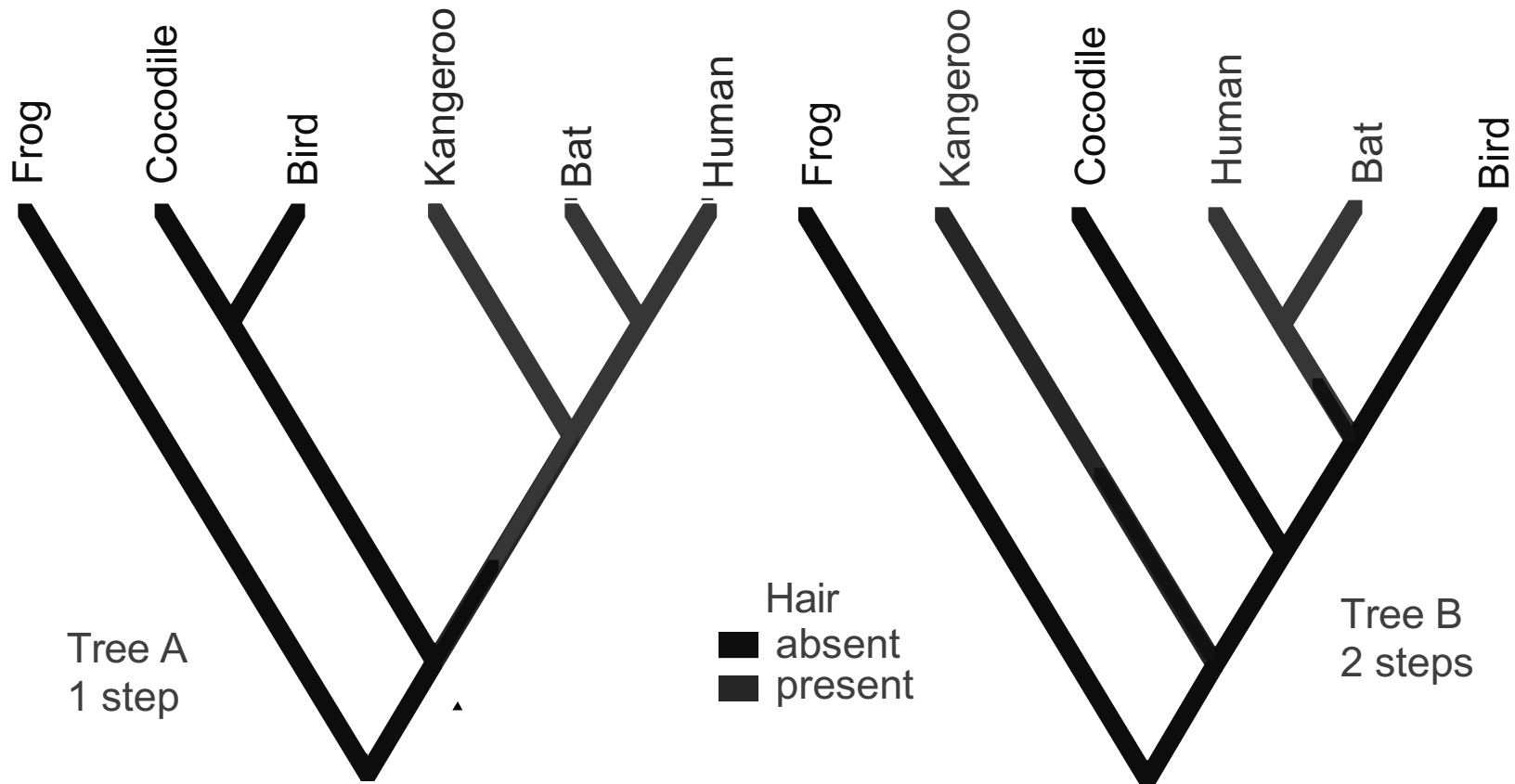
Parsimony analysis

- **Parsimony methods provide one way of choosing among alternative phylogenetic hypotheses**
- **The parsimony criterion favours hypotheses that maximise congruence and minimise homoplasy**
- **It depends on the idea of the fit of a character to a tree**

Character Fit

- **Initially, we can define the fit of a character to a tree as the minimum number of steps required to explain the observed distribution of character states among taxa**
- **This is determined by parsimonious character optimization**
- **Characters differ in their fit to different trees**

Character Fit

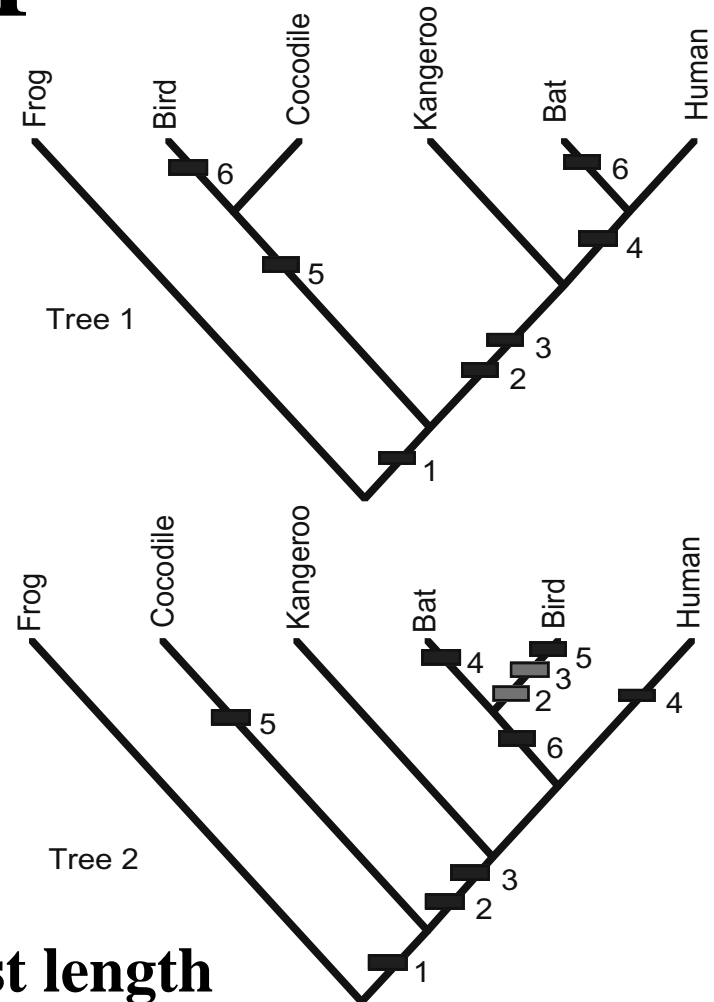


Parsimony Analysis

- **Given a set of characters, such as aligned sequences, parsimony analysis works by determining the fit (number of steps) of each character on a given tree**
- **The sum over all characters is called Tree Length**
- **Most parsimonious trees (MPTs) have the minimum tree length needed to explain the observed distributions of all the characters**

Parsimony in practice

		CHARACTERS						
		1	2	3	4	5	6	
		amnion	hair	lactation	placenta	antorbital fenestra	wings	
TAXA	Frog	-	-	-	-	-	-	
	Bird	+	-	-	-	+	+	
	Crocodile	+	-	-	-	+	-	
	Kangaroo	+	+	+	-	-	-	
	Bat	+	+	+	+	-	+	
	Human	+	+	+	+	-	-	
								TREE LENGTH
FIT	Tree 1	1	1	1	1	1	2	7
	Tree 2	1	2	2	2	2	1	10



Of these two trees, Tree 1 has the shortest length and is the most parsimonious
Both trees require some homoplasy (extra steps)

Results of parsimony analysis

- **One or more most parsimonious trees**
- **Hypotheses of character evolution associated with each tree (where and how changes have occurred)**
- **Branch lengths (amounts of change associated with branches)**
- **Various tree and character statistics describing the fit between tree and data**
- **Suboptimal trees - optional**