Lecture Handout 23: MOLECULAR EVOLUTION AND PHYLOGENETICS
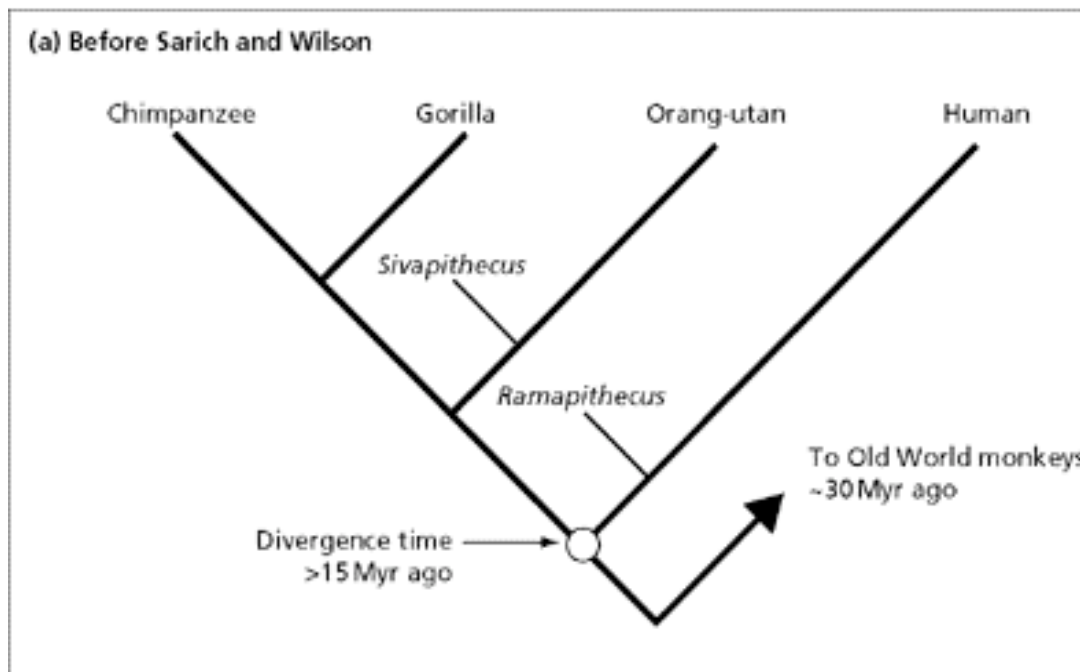
Advantages of molecular data

a) Molecular data is easily converted into a numerical form and can be statistically analyzed.

b) Large molecular datasets can be generated quickly.

c) Character states are unambiguous and easily recognizable (A, G, T, & C – four character states).

d) One sequence has many polymorphic sites.

Milestones to Molecular Evolution:
Human-primate phylogeny

(a) Before Sarich and Wilson

Chimpanzee   Gorilla   Orang-utan   Human

Sivapithecus

Ramapithecus

To Old World monkeys ~30 Myr ago

Divergence time >15 Myr ago

- Human phylogenetics held a common view that humans were on a branch separate from all great apes
- Human split was placed at 15 Mya
  - based on Ramapithecus thought to be the first representative of the human lineage

Milestones to Molecular Evolution:
Human-primate phylogeny

(b) After Sarich and Wilson

Chimpanzee   Gorilla   Human   Orang-utan

Divergence time
~5 Myr ago

Sivapithecus and
Ramapithecus

To Old World monkeys
~30 Myr ago

b Page, Holmes
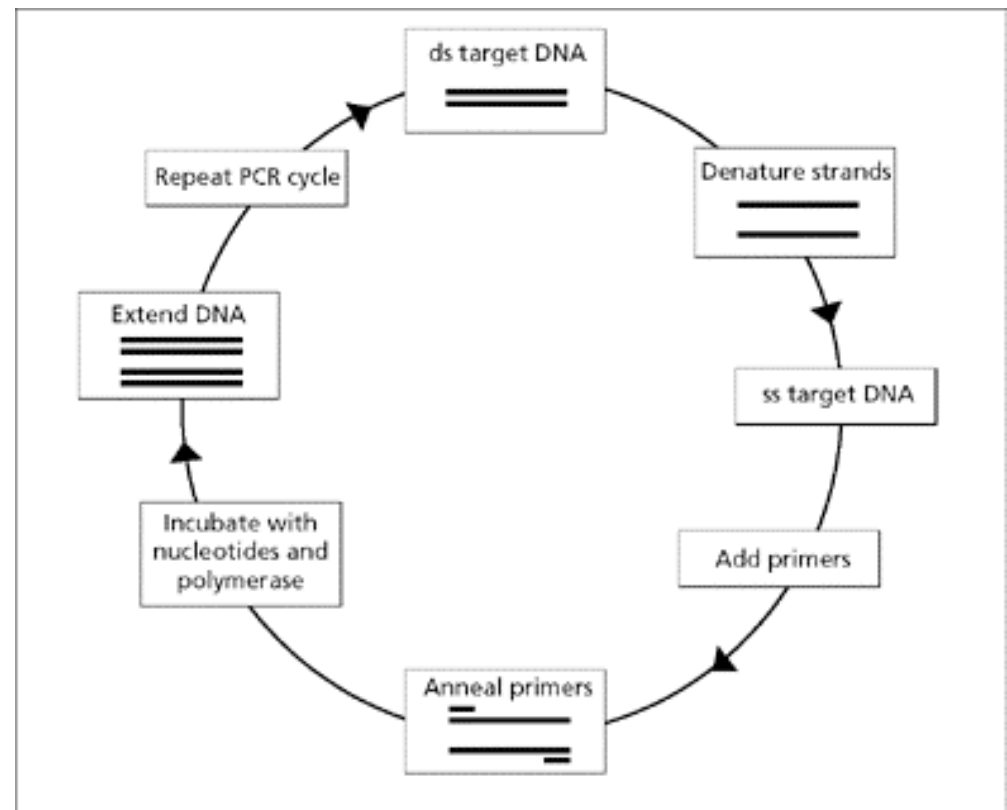Molecular Evolution
Blackwell
Science

- The work of Sarich and Wilson (1967) with updated immunological data showed that chimp gorilla and human were equidistant from orangutan
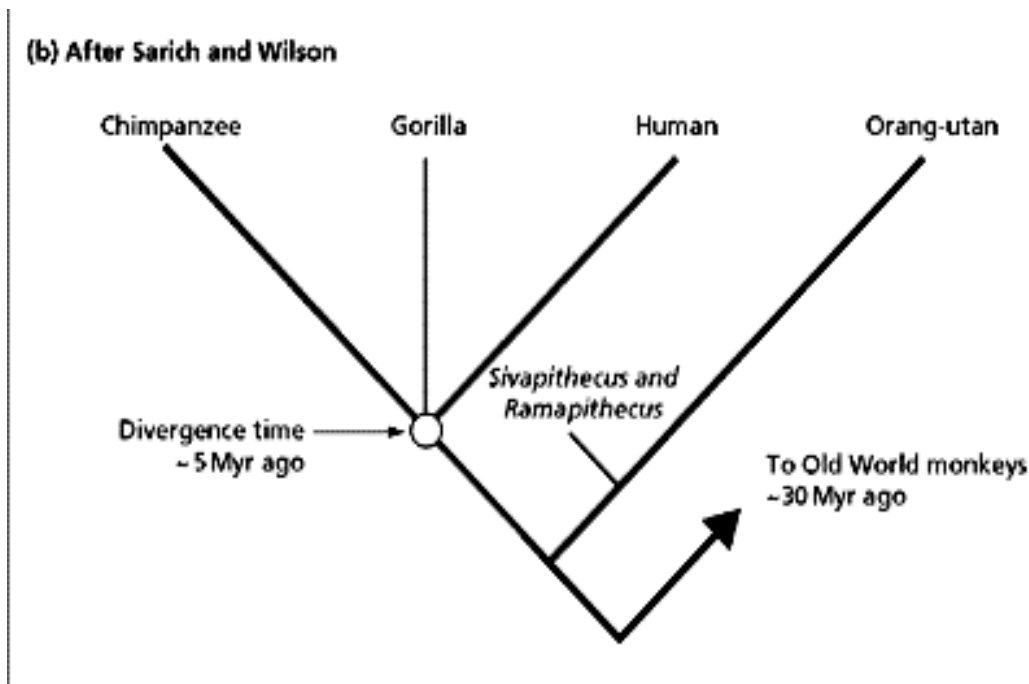
- **Split was placed at 5 Mya**

Milestones to Molecular Evolution: Polymerase Chain Reaction (PCR)

- DNA primer is usually a short oligo made by chemical synthesis
- The sequence of this oligo determines the position at which it attaches to the template DNA

Page, Holmes
Molecular Evolution
Blackwell Science

Milestones to Molecular Evolution:
Polymerase Chain Reaction (PCR)

- Allowed segments of DNA to be generated quickly and from small amounts of material

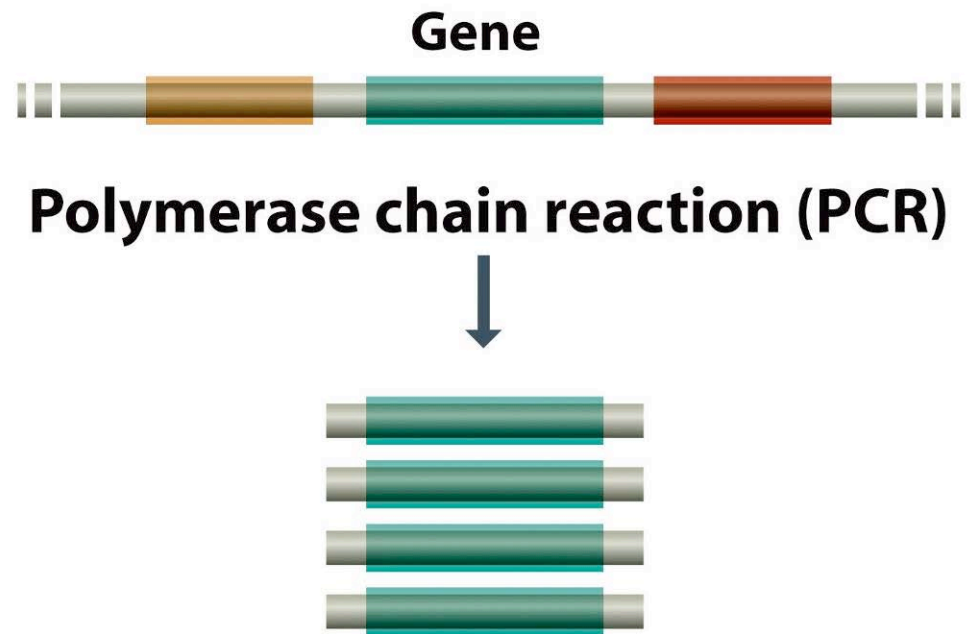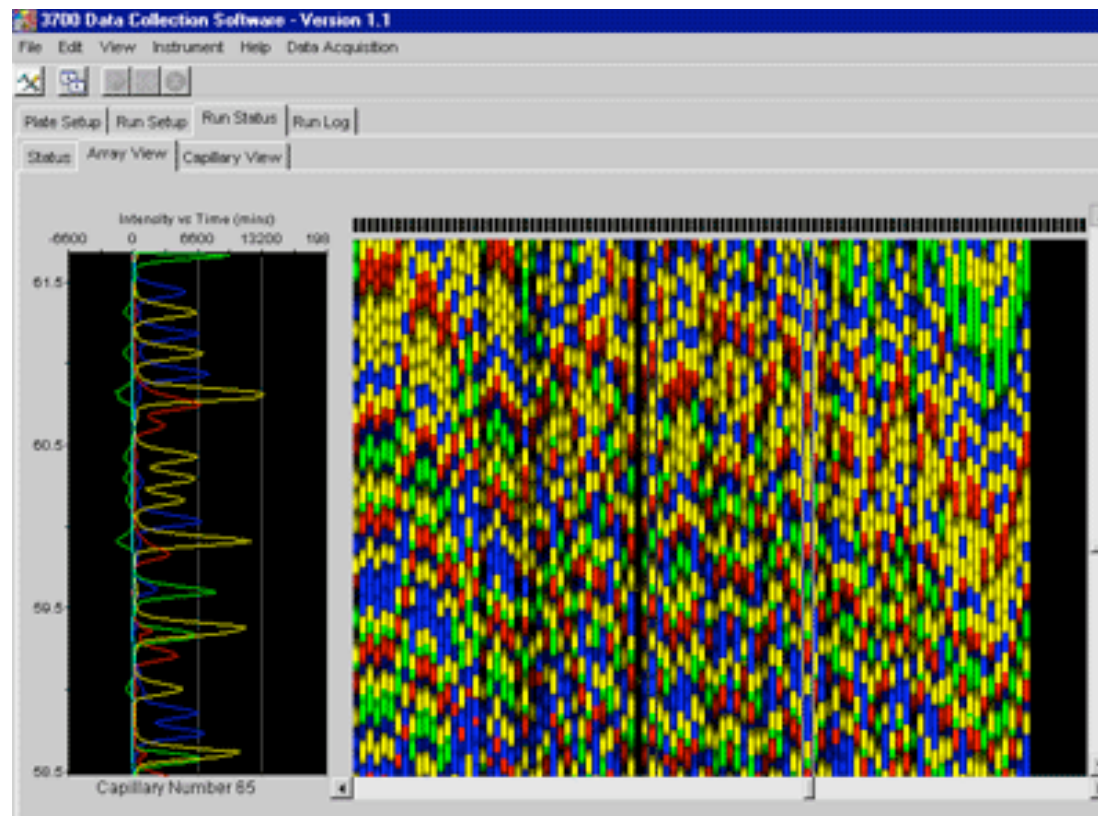- Sequencing DNA – reading evolutionary signatures in the genes

Gene

Polymerase chain reaction (PCR)

Figure 2-3 Genomes 3 (© Garland Science 2007)

Automated sequencing

Two main objectives in phylogenetic analysis

- The first one is finding the true tree topology: **determining the true evolutionary trajectory**.

- The second is establishing timing of speciation events using degree of difference between two sequences to assign the divergence time: **timing with a molecular clock**.

The steps in a phylogenetic analysis

- First, align DNA sequences and obtain the comparative data that will be used in reconstructing the tree.

- Then, convert the comparative data into the reconstructed tree.

- Assess the accuracy of the tree using resampling approaches.

- A tree obtained by phylogenetic analysis is an **inferred (reconstructed) tree** since it is inferred from the data analyzed.

Sequence alignment

Comparing DNA sequences

Species A   ATTGCAACTGGTATCGAGGTTCTAC

Close relative ATTGCCACTGGAATCGAGGTTCTAC

Distant relative ATTGCCACTGGAATCGTGGTTCGAC

2 differences in 25 nucleotides
2/25 = 8% or 92% similarity

4 differences in 25 nucleotides
4/25 = 16% or 84% similarity

Figure 9-17a Biology: Science for Life, 2/e
© 2007 Pearson Prentice Hall, Inc.

Trees based on sequence data

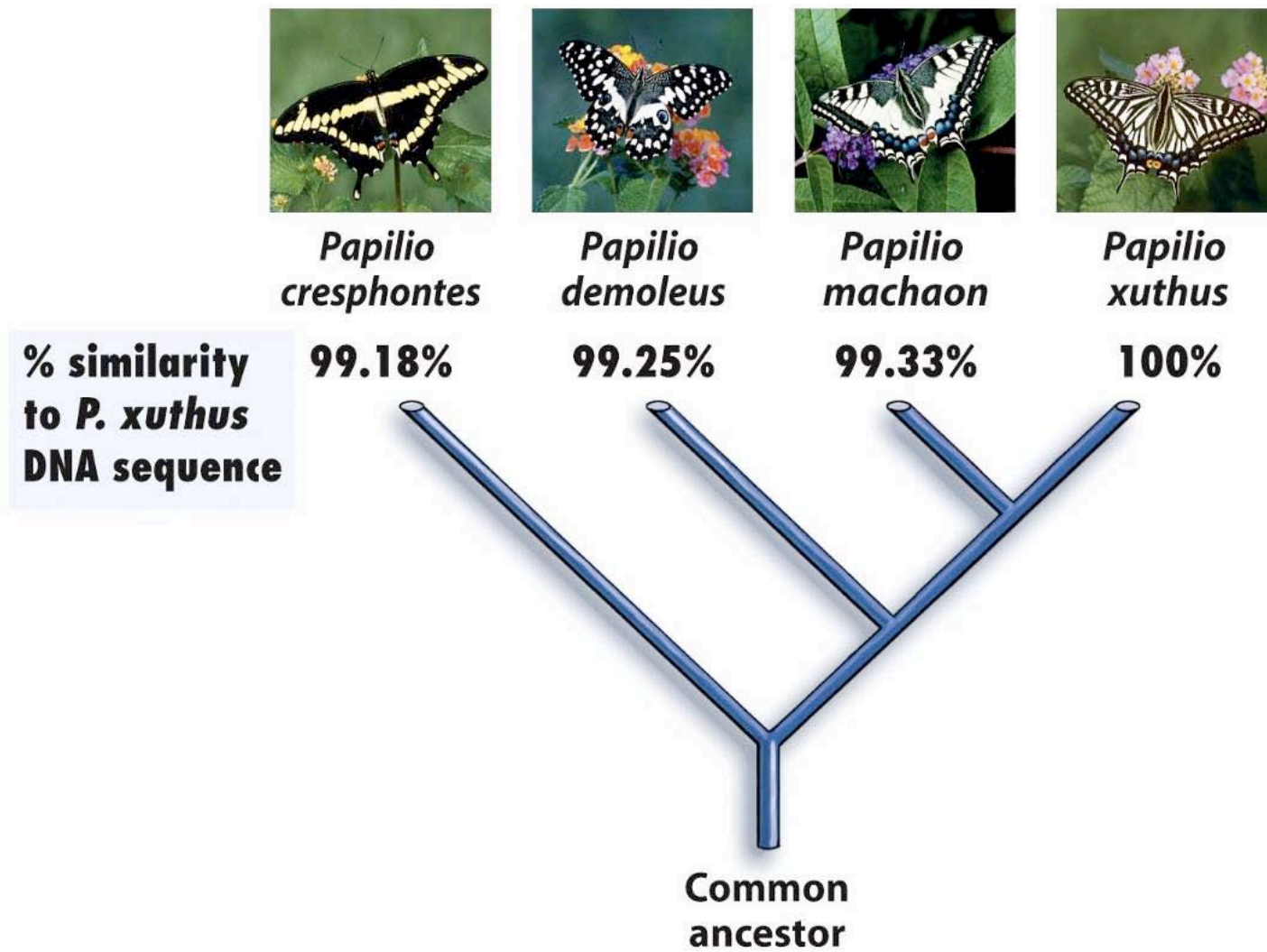|	Papilio cresphontes	Papilio demoleus	Papilio machaon	Papilio xuthus
% similarity to *P. xuthus* DNA sequence	99.18%	99.25%	99.33%	100%

Common ancestor

Figure 9-18 Biology: Science for Life, 2/e
© 2007 Pearson Prentice Hall, Inc.

True tree vs. an inferred tree

- A tree obtained by phylogenetic analysis is an **inferred (reconstructed) tree**
  - It is inferred from the data analyzed
- **True tree** – depicts the actual series of events that occurred:
  - Sometimes we can be very confident that the inferred tree is the true tree, but there is usually some uncertainty

Milestones to Molecular Evolution:
Molecular Phylogenetics

Page, Holmes
Molecular Evolution
Blackwell Science

- Until the rise of molecular biology, Archaea was completely unknown!
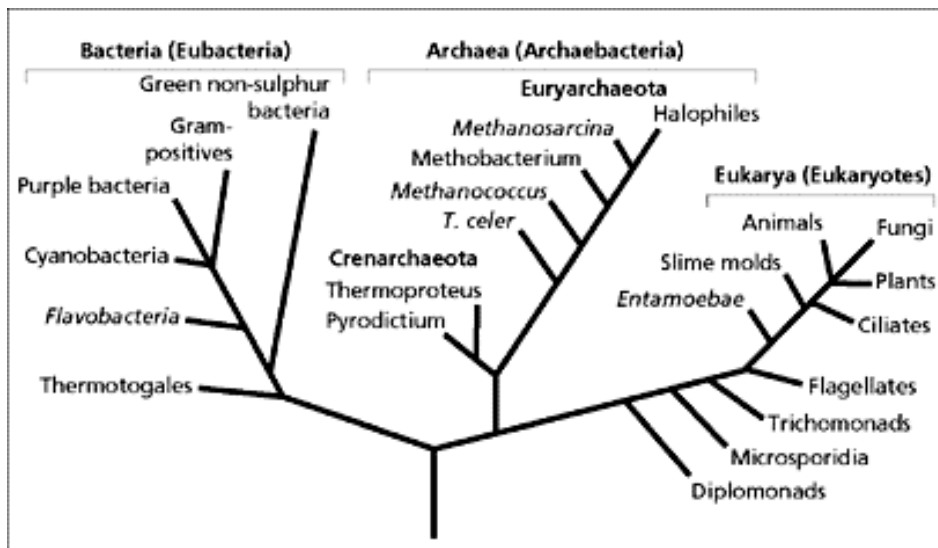
- Only using molecular phylogenetics allowed us to see that there was a whole separate domain of life

The origins of AIDS

African green monkey

Sykes monkey

Mandrill

Sooty mangabey

HIV-2

HIV-1O

Chimpanzee

HIV-1M

ZR59

- Immunodeficiency viruses are present in chimpanzees and Old World monkeys

- SIVs are not pathogenic in their normal hosts, but if transferred to humans they may cause diseases
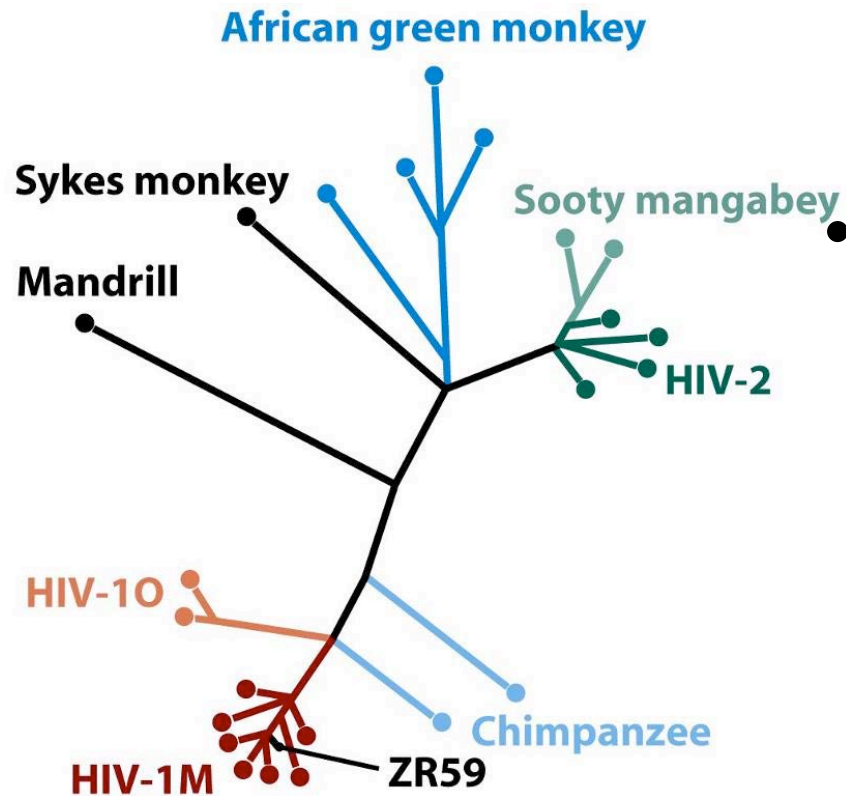
Figure 19.20 *Genomes 3* (© Garland Science 2007)

The origins of AIDS

African green monkey

Sykes monkey

Mandrill

Sooty mangabey

HIV-2

HIV-1O

Chimpanzee

HIV-1M   **ZR59**

- Retroviruses accumulate mutations quickly because reverse transcriptase has no efficient proofreading activity
  - Molecular clock is fast
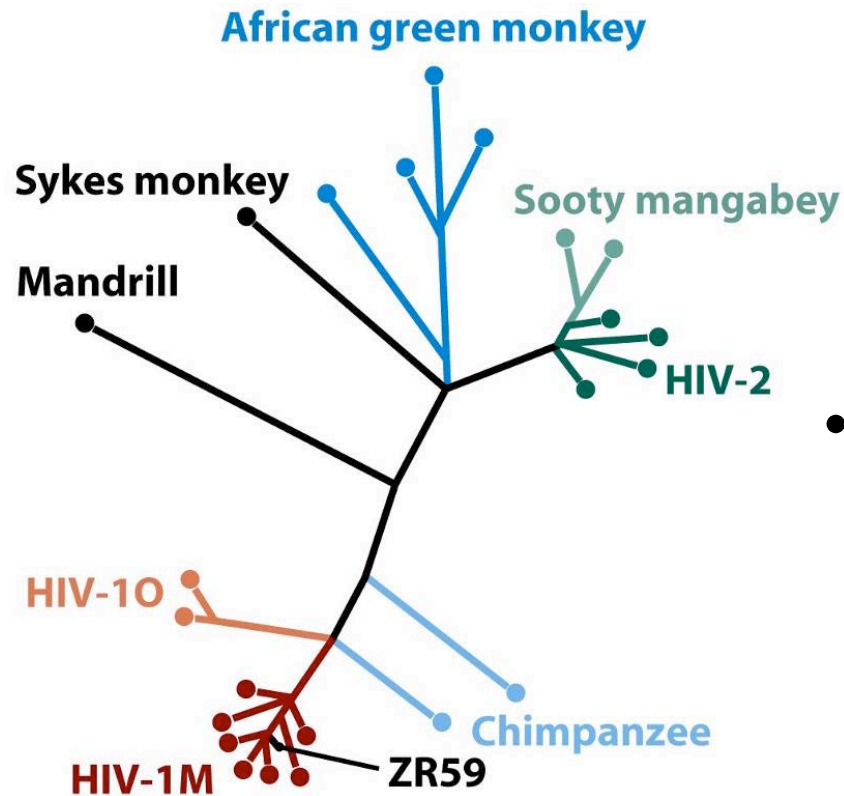- Suggested that HIV spread begun ~1931 (1915-1941)

Figure 19.20 *Genomes 3* (© Garland Science 2007)

Human origins

- Origin in Africa – most of the early hominid fossils are there

- 2 Mya – *Homo erectus*
  - First hominid to disperse around the world
  - This specimen "Nariokotome boy" is about 1.6 My old, found near lake Turkana, Kenya
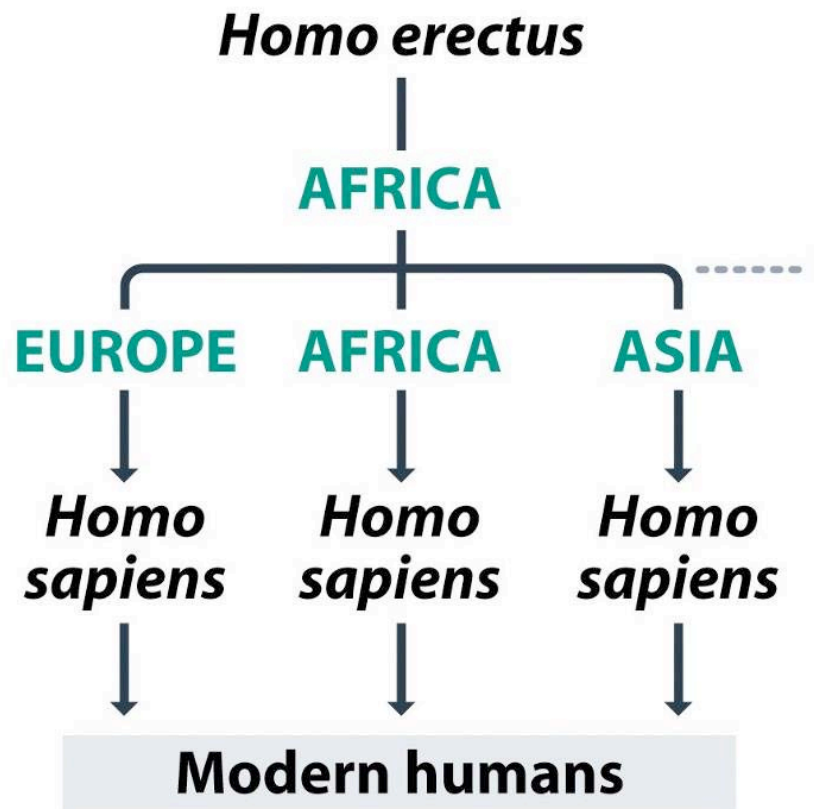
- What happened then?

Figure 19.22 *Genomes 3* (© Garland Science 2007)

Origins of modern humans:
Multiregional evolution

Multiregional evolution

Homo erectus

AFRICA

EUROPE   AFRICA   ASIA

Homo sapiens   Homo sapiens   Homo sapiens

Modern humans

- Because *H. erectus* was found on so many continents
- *H. sapiens* arose on those continents independently with some limited contact between populations

Figure 19.23a *Genomes 3* (© Garland Science 2007)

Origins of modern humans: "Out of Africa" hypothesis

Homo erectus

AFRICA

EUROPE   AFRICA   ASIA

✗   Homo sapiens   ✗

Modern humans

- Rather than evolving in parallel throughout the world,
  - *H. sapiens* originated in Africa 200 kya
  - Spread around the world 100-50 kya
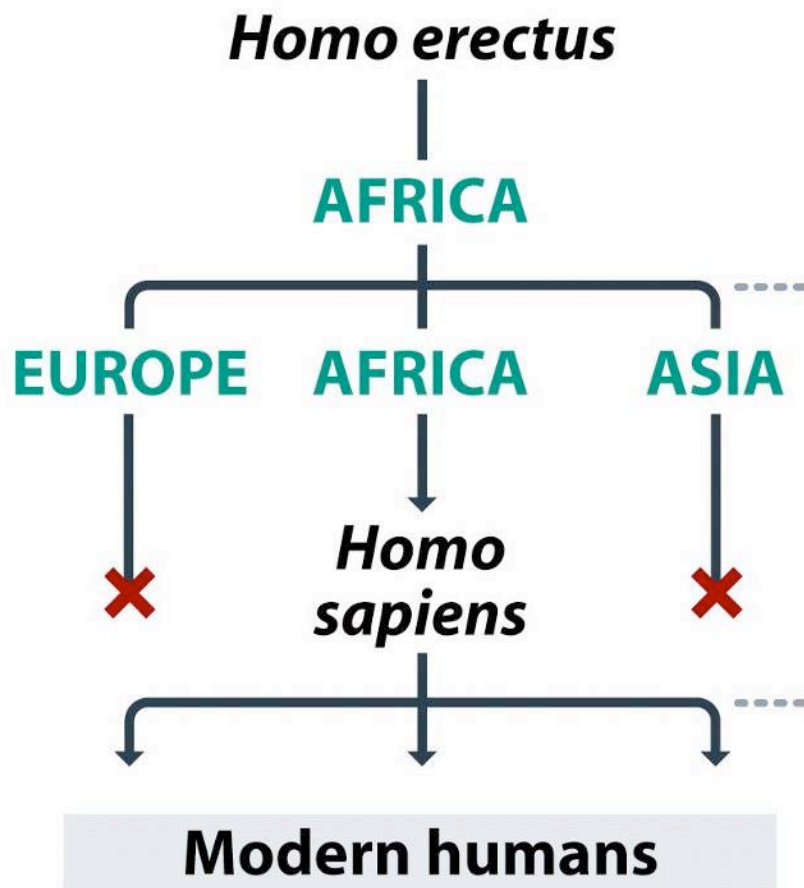
Figure 19.23b *Genomes 3* (© Garland Science 2007)

Origins of modern humans:
"Out of Africa" theory

- RFLP MtDNA tree supports the hypothesis:
  - 7 African mtDNA genomes - on one side and all the rest on the other side of the split shown

- This analysis has since been criticized, but other studies came forward to confirm the theory:
  - mtDNA
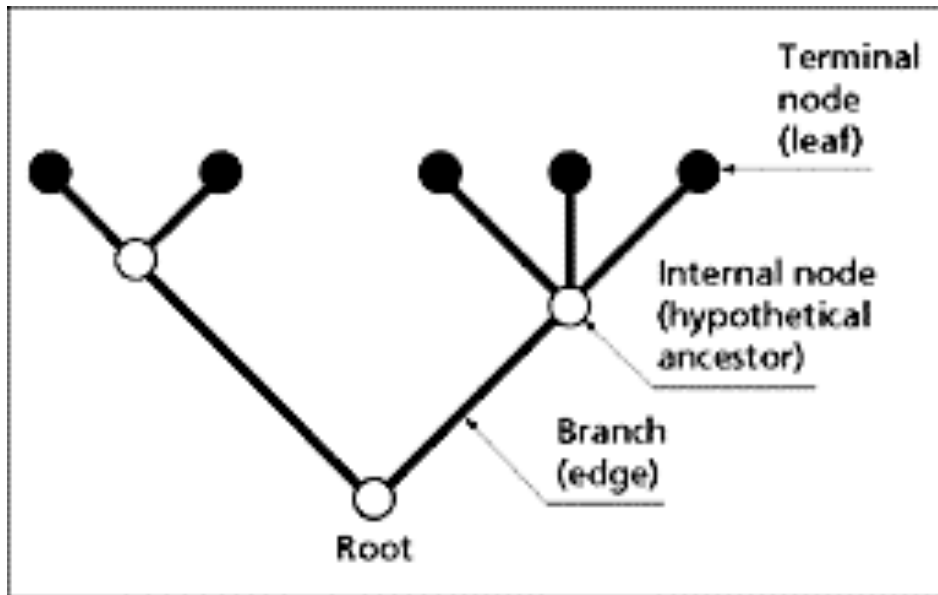 - Y-chromosome
    - but a different date 800 kya
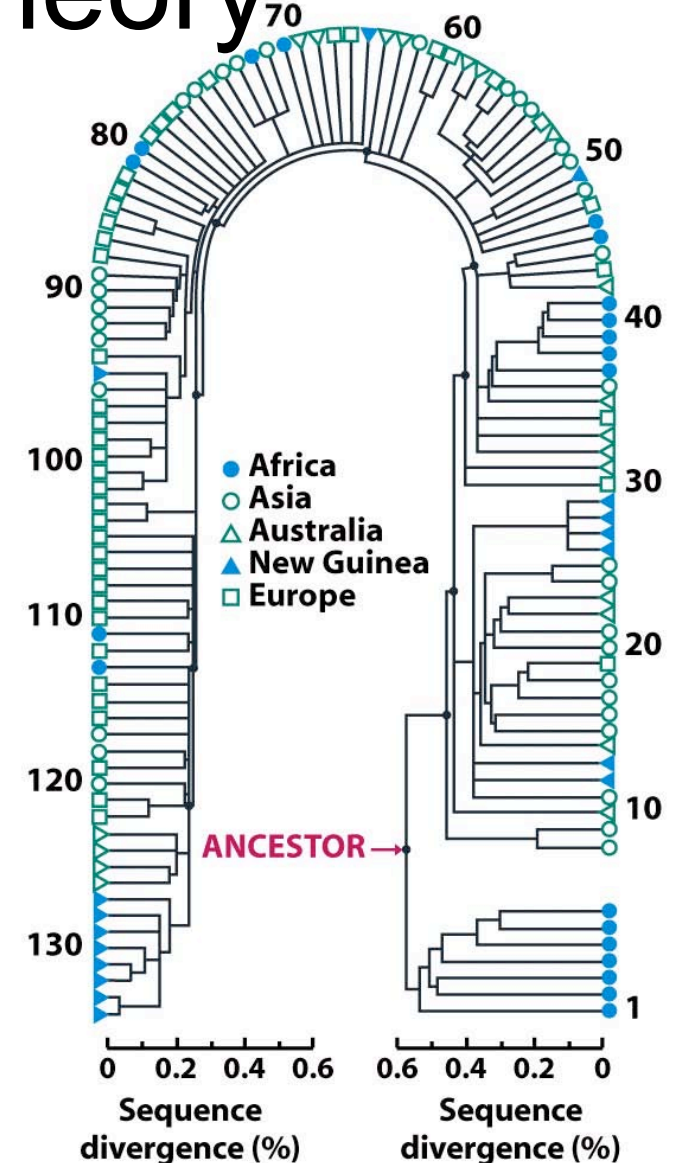
Figure 19.24 *Genomes 3* (© Garland Science 2007)
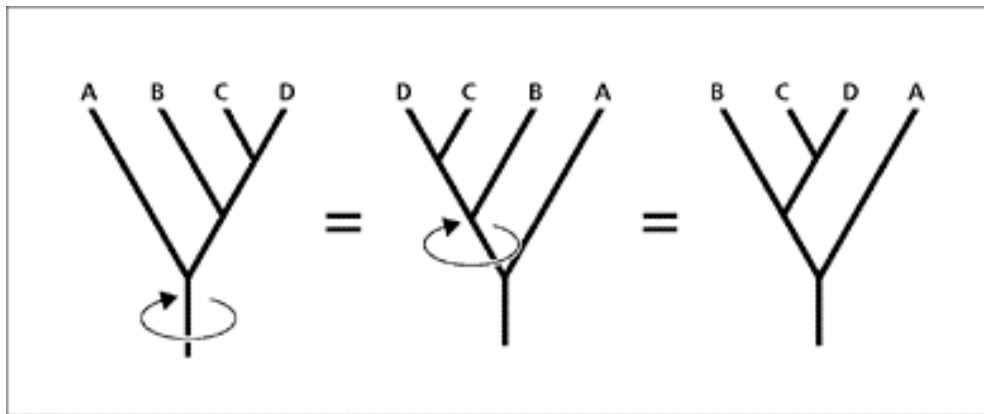
A tree

Page, Holmes
Molecular Evolution

- **Terminal nodes** (OTUs, operational taxonomic units) – represent sequences of organisms for which we have data:
  – Either extant or extinct
- **Internal nodes** – represent hypothetical ancestors
- **Root of the tree** – the hypothetical ancestor to all of the sequences in the tree

Branch lengths

- Most methods estimate the amount of evolution that takes place between each node of the tree

- This can be represented as **branch lengths**
- **Weighted trees** – trees that indicate branch lengths

Different ways of drawing a tree

Page, Holmes
Molecular Evolution
Blackwell
Science

- The order in which the labels on this tree are drawn can be different without changing the meaning of the tree

Resolution

Star tree   Partially resolved   Fully resolved
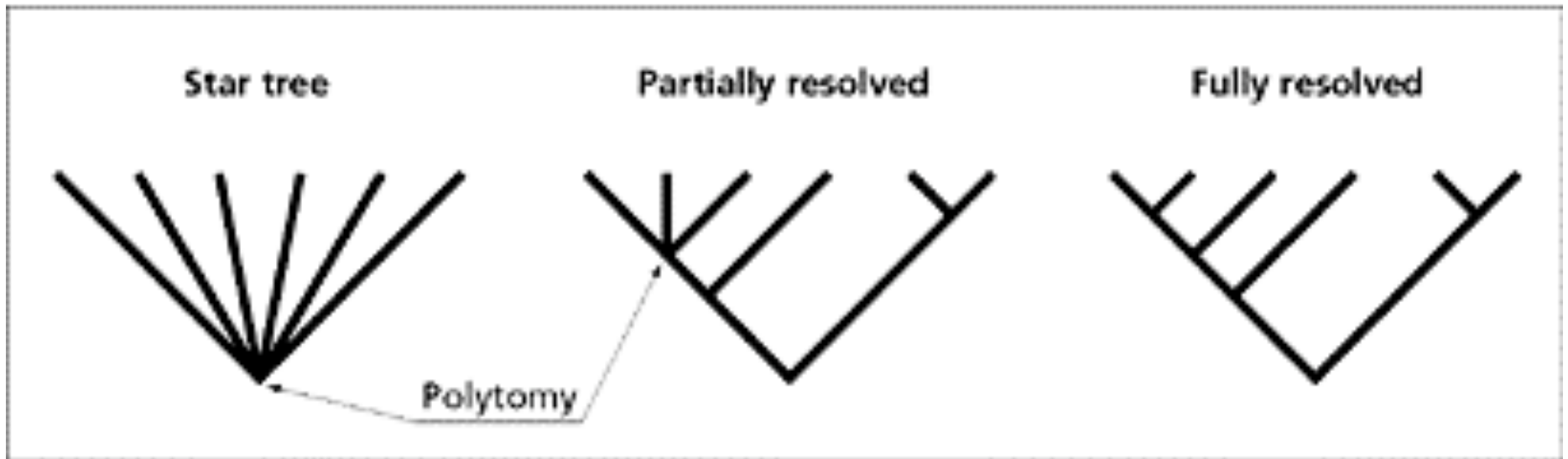
Polytomy

- **"Degree" of a node** – the number of adjacent branches possessed by an internal node
- **Polytomy** – if a node has a degree higher than three
- A tree that has no polytomies is **fully resolved**

Polytomy

'Hard' polytomy (simultaneous divergence)   'Soft' polytomy (uncertainty)

or   or...

b Page, Holmes
Molecular Evolution

1. **Hard polytomy:** represents simultaneous divergence – all the descendants evolved at the same time

2. **Soft polytomy** – represents our uncertainty about phylogenetic relationships: not sure about the order
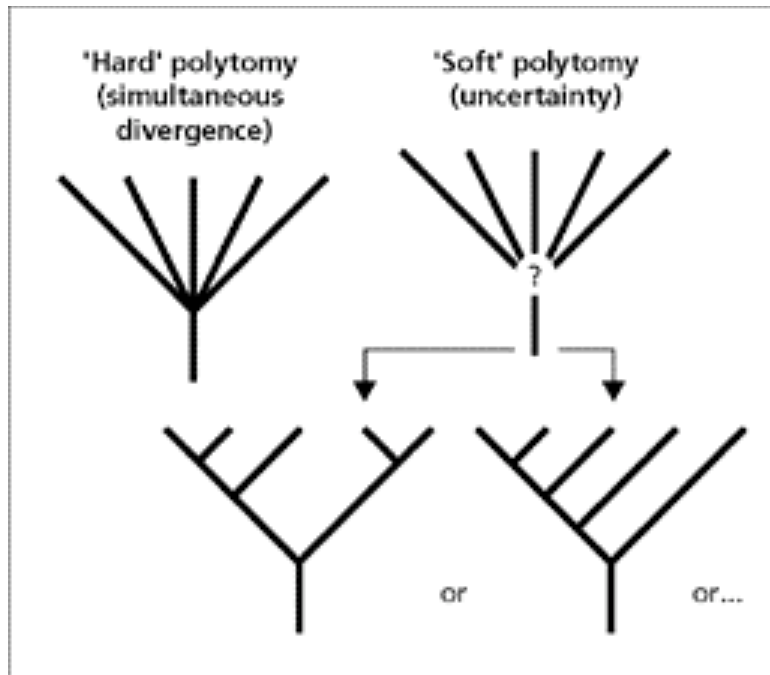
- Typically all polytomies are treated as soft

Different types of trees

- Dendrogram (any tree-like diagram used to illustrate clustering)
- Cladograms
- Phylograms (Additive trees)
- Chronograms (Ultrametric trees)

Different types of trees:
# The shorthand for trees

Page, Holmes
Molecular Evolution

- Trees can be represented using parentheses (((A,B),C),(D,E))

- We can write trees in text without using pictures:
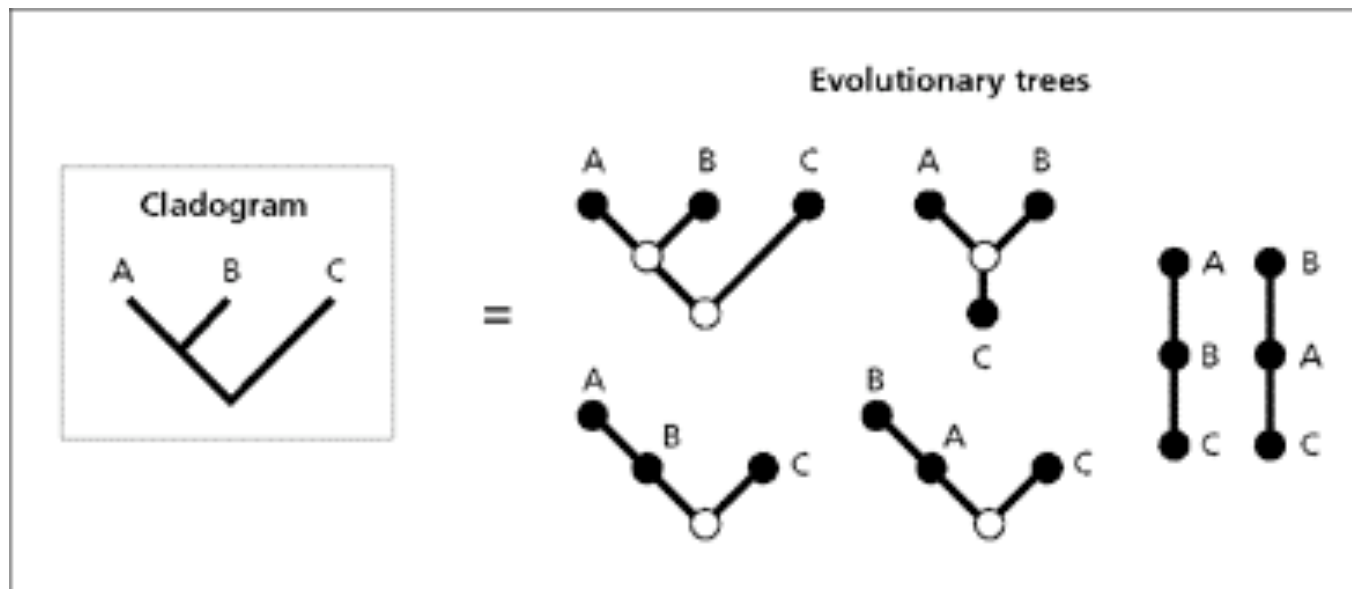
- Same format is used by software to store trees

NEWICK'S
Lobster House
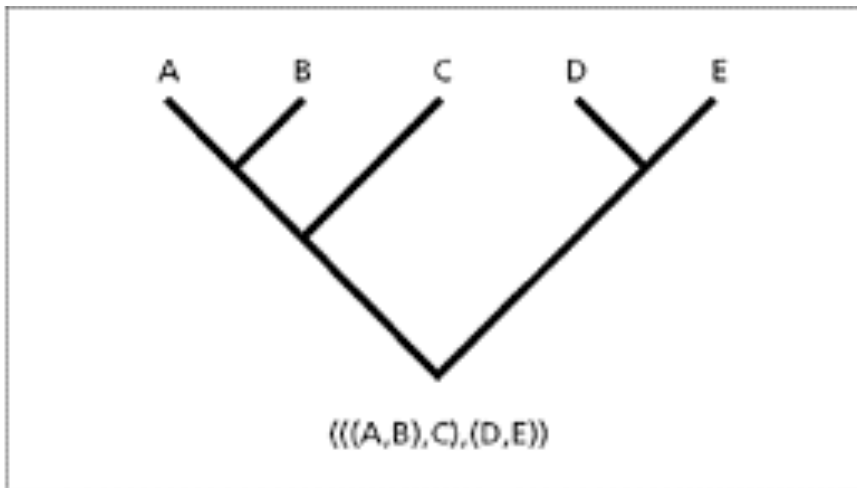
Different types of trees: Cladograms

- A **cladogram** simply shows the relative recency of common ancestry:

  - A and B share a common ancestor more recently than A & C or B & C

  - Sometimes are called
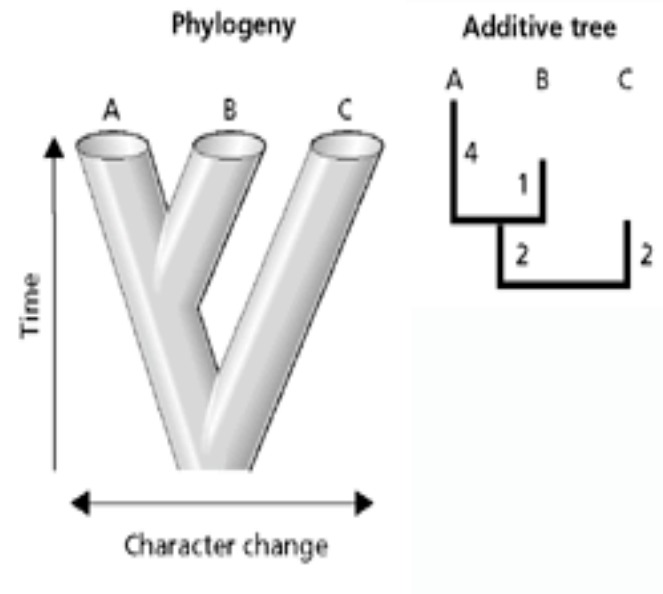   **n-trees**
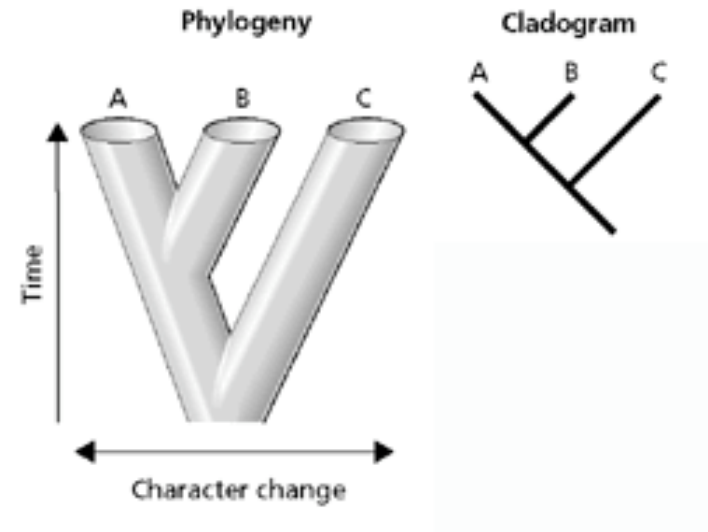
Cladogram vs. evolutionary tree

- There are six evolutionary trees consistent with the cladogram

Different types of trees:
Phylograms

- **Phylograms** contain additional information such as branch lengths

- This can represents the amount of evolutionary change

  – A acquired 4 substitutions since it shared a common ancestor with B

- Sometimes these are called **metric trees** or **additive trees**

Different types of trees:
Chronograms

- **Chronograms** are a special type of additive trees where all the tips of the trees are equidistant from the root

- Thus they can be used to depict evolutionary time either in years of in the amount of sequence divergence using molecular clock

- Sometimes these are called **ultrametric trees**

Phylogeny   Ultrametric tree

Time

Character change
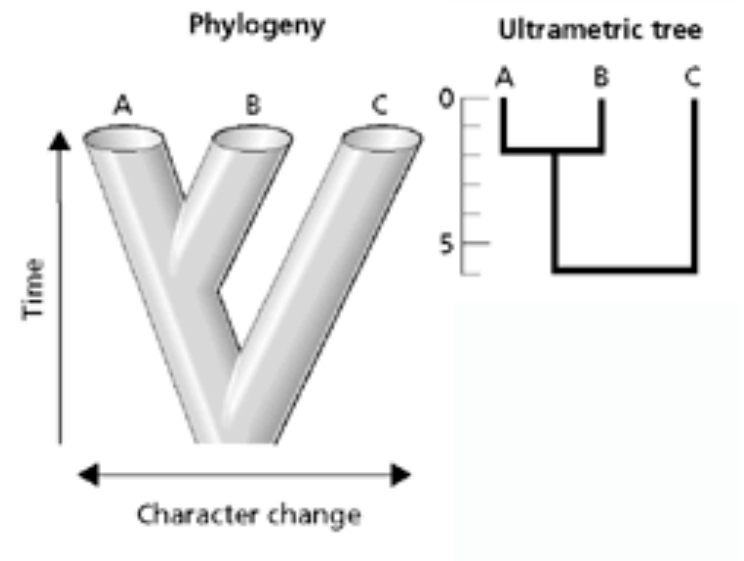
Axes of the trees
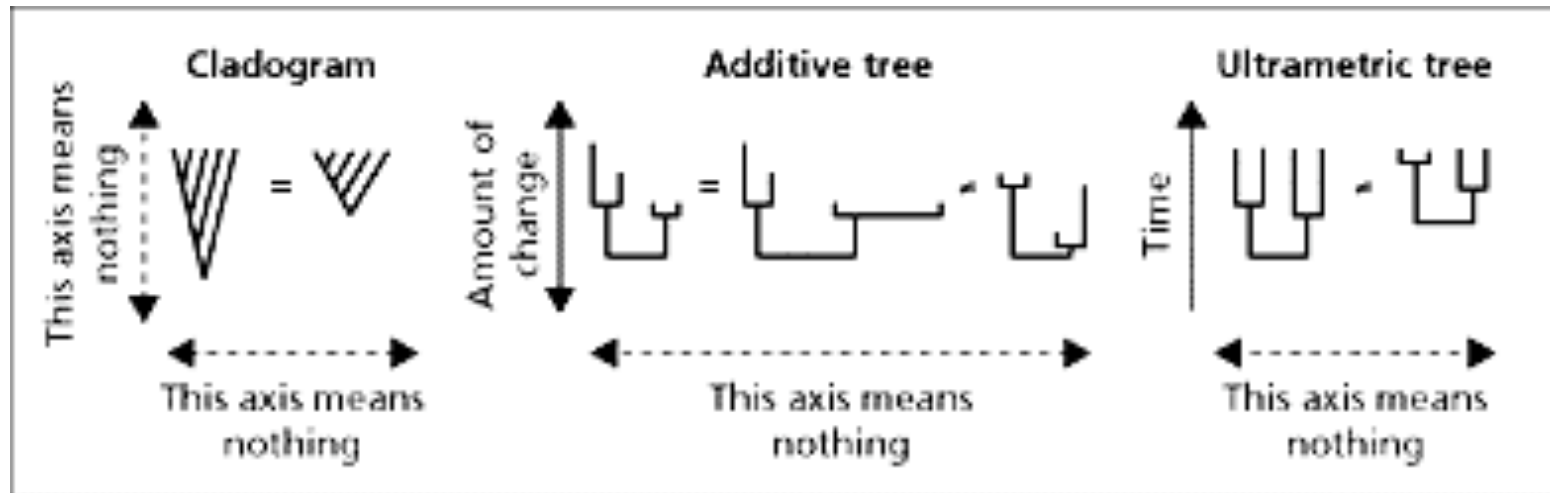
- Fro **cladograms**, both axes mean nothing
- For **phylograms (additive trees)**, one axis mean the amount of evolutionary change
- For **chronograms (ultrametric trees),** one axis signifies time

Which tree do we need?

- Sometimes a cladogram is enough, but when determining the rate of evolution, need a metric tree

- Metric trees already contain all the information from the cladograms

Phylogeny   Cladogram

Additive tree

Ultrametric tree

*b* Page, Holmes
Molecular Evolution
Blackwell
Science
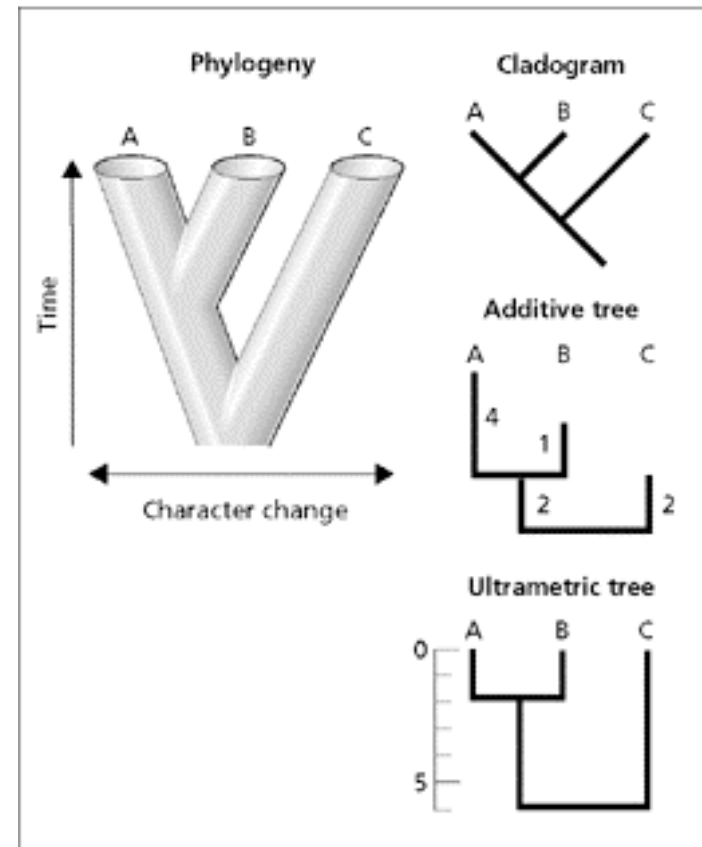
Rooted & unrooted trees

- **A rooted tree** has a node identified as a root from which ultimately all nodes descend
  - Hence it has directionality (polarity)
  - Corresponds to evolutionary time – the closer the node is to the root, the older it is in time
  - Node closest to the root is the ancestor of all the nodes away from the root
- **An unrooted tree** lacks a root, and hence does not specify any evolutionary directionality

Unrooted trees

- Typical **unrooted tree**
  - Does not tell anything about the sequence of the evolutionary events
- Topology:
  - 4 **external nodes**
  - 2 **internal nodes**
  - Length of branches indicates the **degree of difference (if phylogram)**
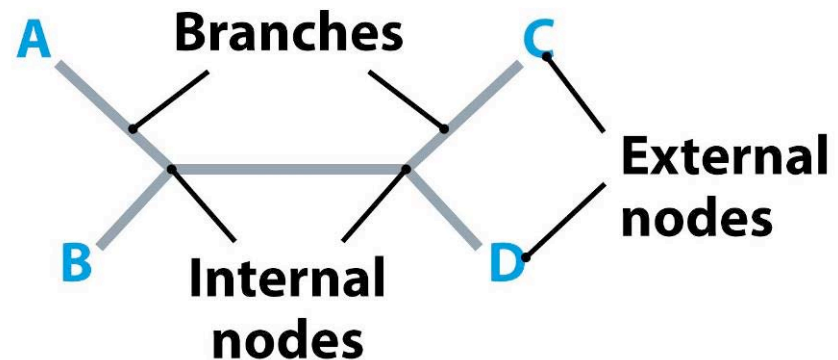
Rooted trees

- Examples of typical rooted trees

- **An outgroup** – a homologous sequence that is less closely related to A,B,C & D than they are to each other

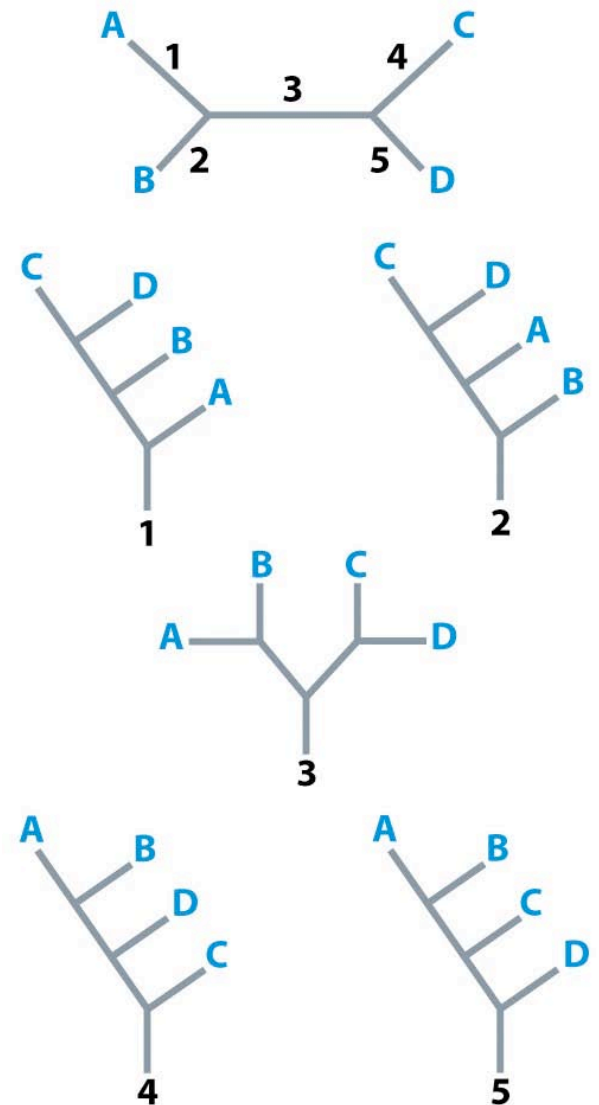 – It enables the root of the tree to be located and the correct evolutionary path identified

Figure 19.4b *Genomes 3* (© Garland Science 2007)
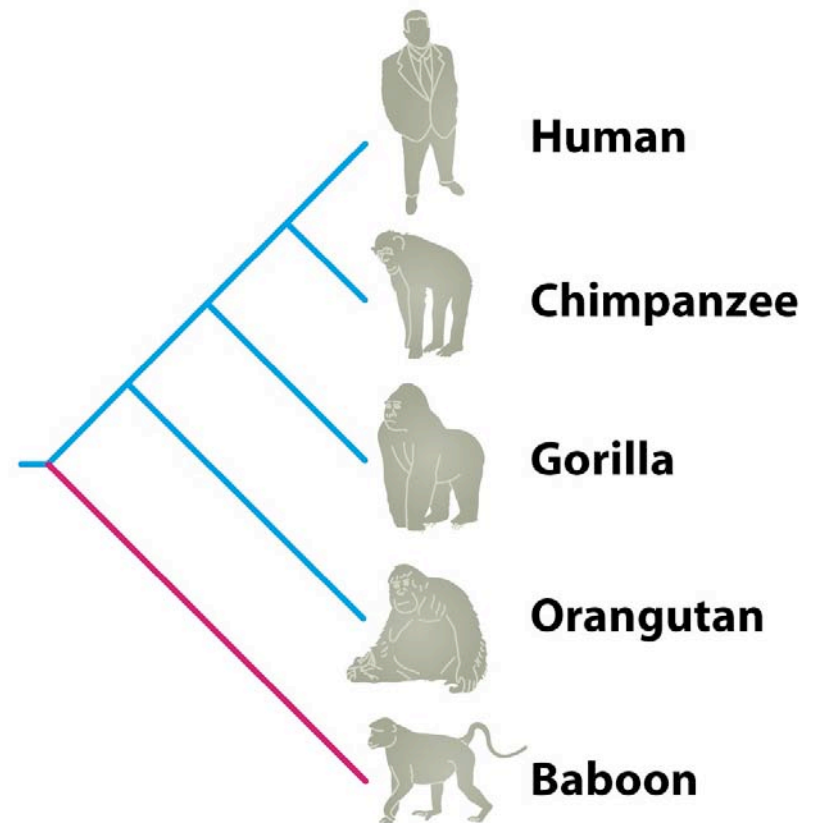
Rooted & unrooted trees:
The outgroup

- We could place the root on any of the seven branches of this tree

- Hence this unrooted tree can correspond to seven different rooted trees!

- So we can't root trees arbitrarily, and need additional information to place a root
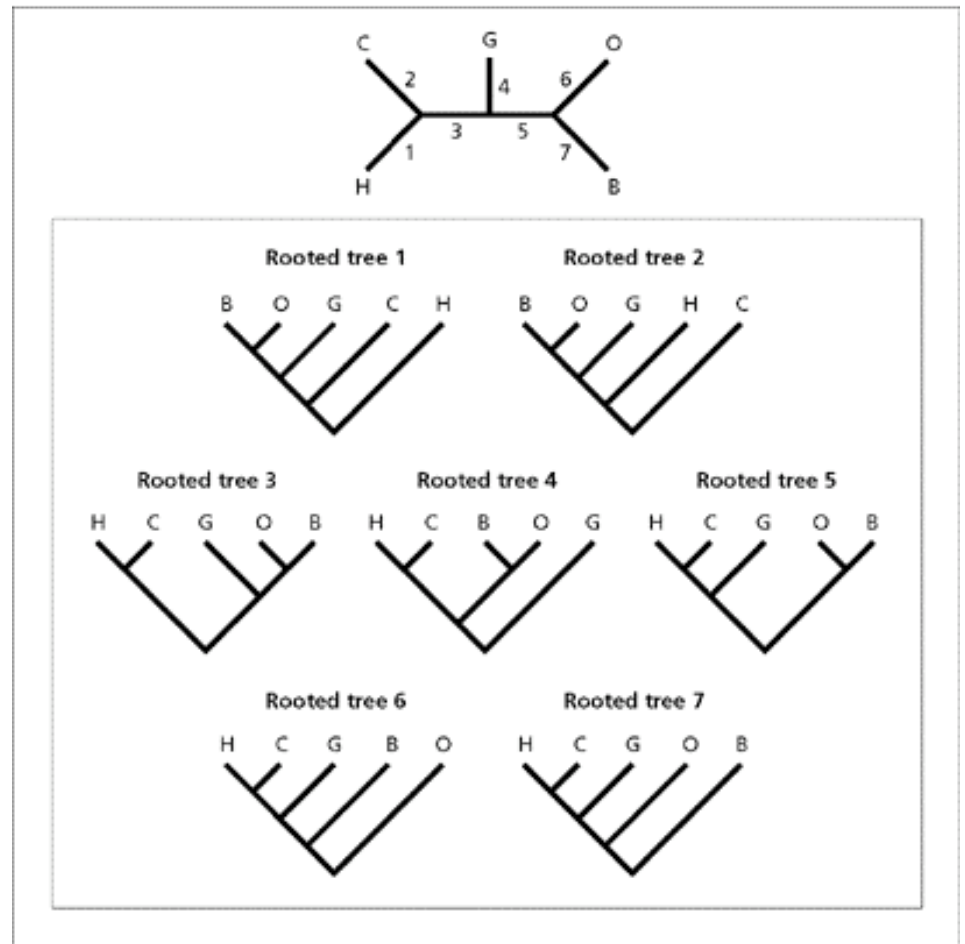
The outgroup:
Primate phylogeny

- The tree of human, chimpanzee and gorilla are rooted with baboon because we know from the fossil record that baboons split from primate lineage before the other 4 species

Figure 19.5 *Genomes 3* (© Garland Science 2007)

The possible # of trees increases exponentially as number of OTUs increase

Tree topology

Page, Holmes
Molecular Evolution
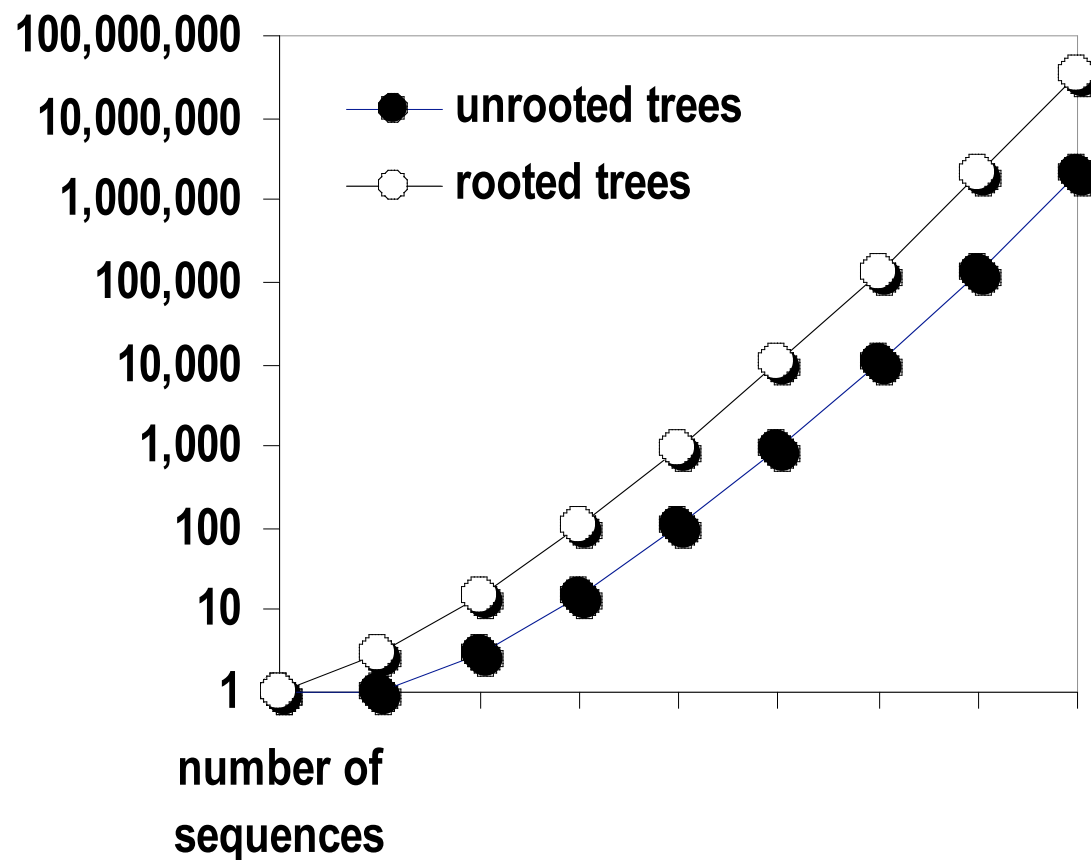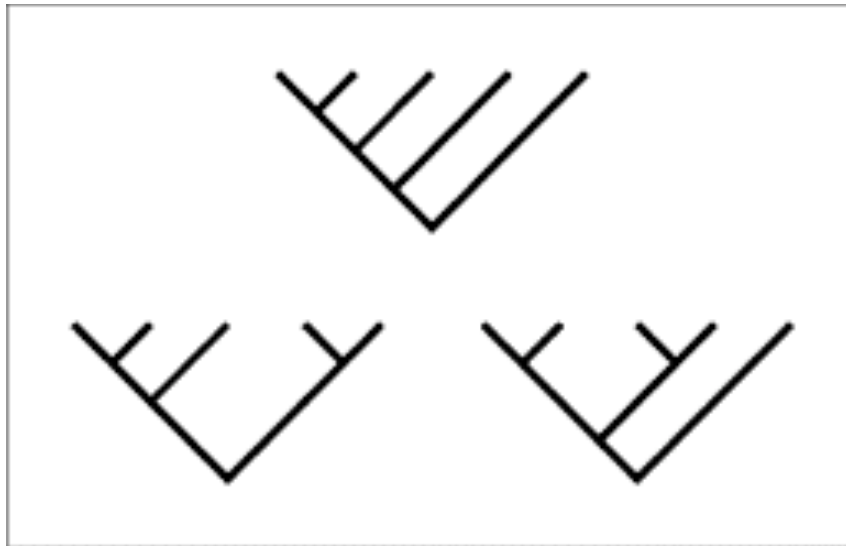
- **Tree topology** – another name for the shape of the tree
- **Example:** In a sample of five sequences, all of the resulting 105 possible rooted trees would have one of the three topologies on the left

Notating the history of character changes

a) Each node is labeled by the corresponding nucleotide

b) Each branch is colored corresponding to the nucleotide at the end of each branch

c) Indicating the branch on which the change took place

b Page, Holmes
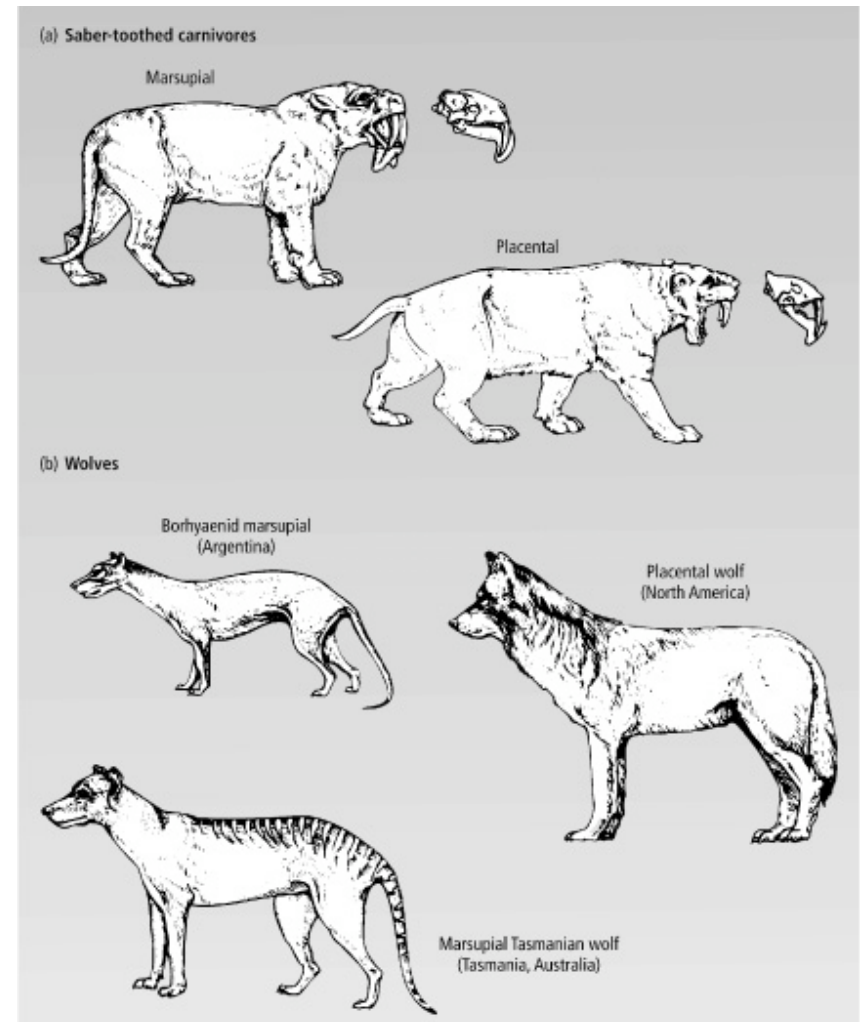**Molecular Evolution**
Blackwell
Science
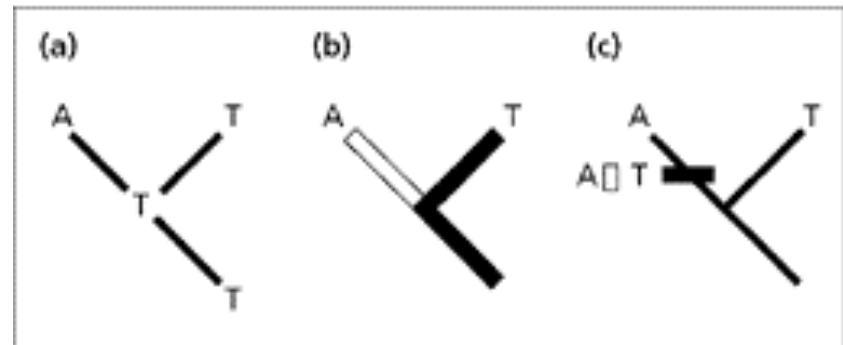
Ancestral & derived states

- Derived character state (**apomorphic**):
    - as opposed to ancestral character state (**plesiomorphic**) evolves from the ancestral state in a more recent common ancestor

- Convergent evolution (**homoplasy**):
   - same character evolved in two different lineages
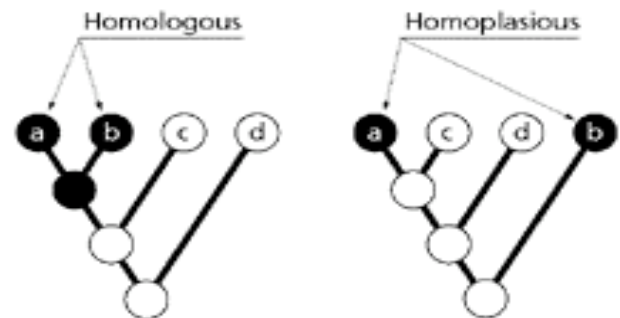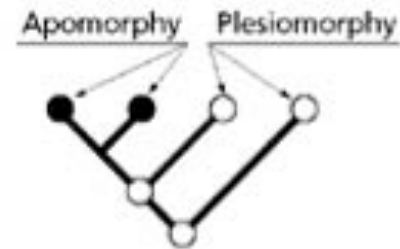
Convergence in marsupial & placental carnivores

- Reconstructed bodies and skulls of saber-toothed marsupial and saber-toothed placental
  - *Thylacosmilus* vs. *Smilodon*
- Convergence in dog-like mammals:
  - *Canis lupus*
  - *Thylacinus cynocephalus*
  - *Prothylacinus patagonicus*

(a) Saber-toothed carnivores

Marsupial

Placental

(b) Wolves

Borhyaenid marsupial (Argentina)

Placental wolf (North America)

Marsupial Tasmanian wolf (Tasmania, Australia)

Derived states:
Convergent evolution

- Both birds and bats possess wings, but bats are more related to wingless mammals than to winged birds

Wings   No wings   Wings

Figure 19.2a *Genomes 3* (© Garland Science 2007)
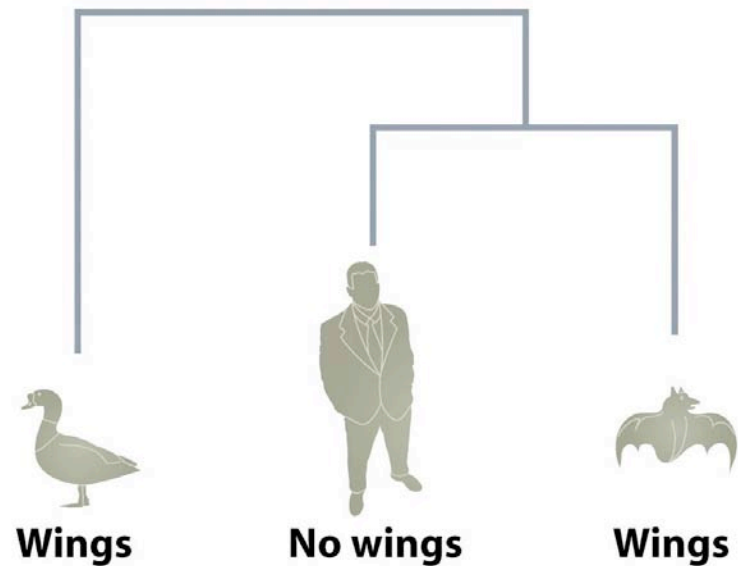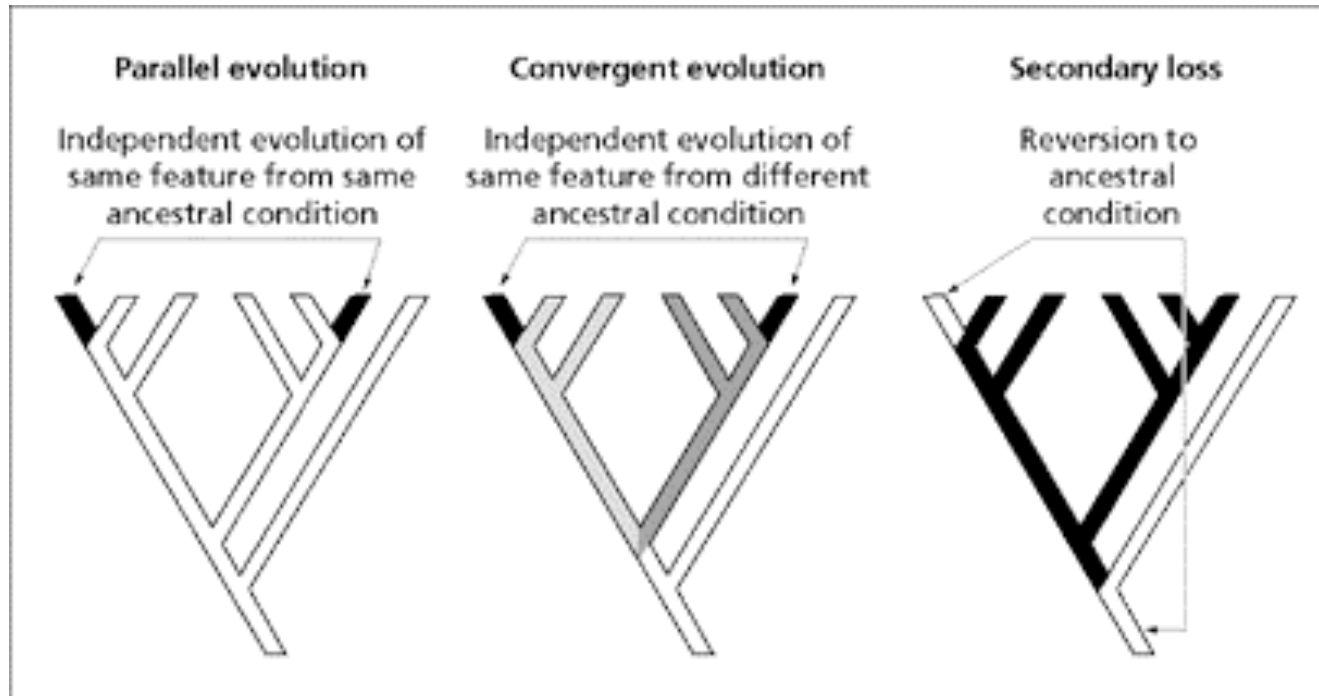
Derived states:
Convergent evolution

- Among vertebrates, the possession of a single toe as displayed by the modern horses is a derived character

- Otherwise it may appear that humans are more related to 5 toed lizards than to horses

5 toes   1 toe   5 toes

Figure 19.2b *Genomes 3* (© Garland Science 2007)
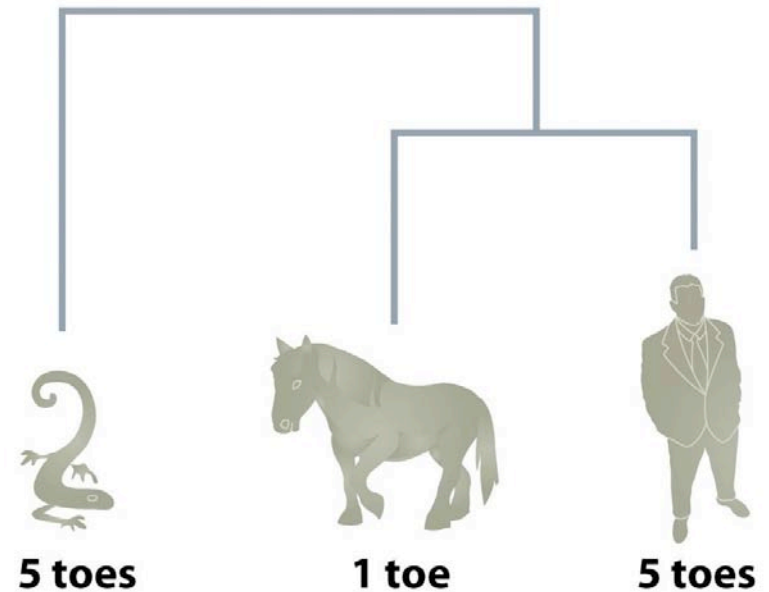
Derived states:
Homoplasy

Homoplasy can arise in three different ways:

Page, Holmes
Molecular Evolution
Blackwell Science

Ancestral & derived character states:
Definitions

- **Plesiomorphic** – character has the same state (e.g., sequence has the same base) as the common ancestor of all OTUs in the three

- **Apomorphic** – derived sequence, opposite of plesiomorphic

- **Autapomorphic** – a unique apomorphic character state

- **Homology** – characters are similar and derived from the ancestral state directly

- **Homoplasy** – characters are similar, but their states have evolved independently

Ancestral & derived states

- Trees showing the terminology used to describe different patterns of ancestral (○) and derived (●) character

- Only homology reflects the true common ancestry

Page, Holmes
Molecular Evolution