

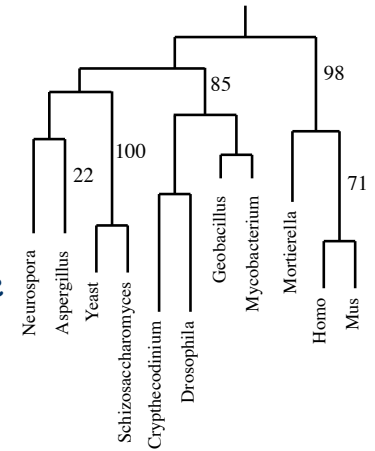


Phylogenies



How To Build a Phylogenetic Tree

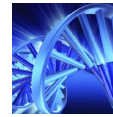
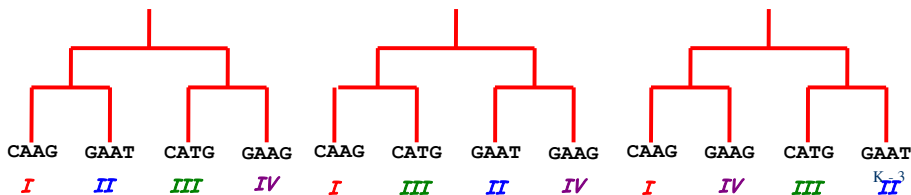
- Start with a set of sequences
- Build a guide tree
- Perform a multiple sequence alignment
- Construct a phylogenetic tree
- Assess the robustness of the phylogenetic tree



Constructing a Phylogenetic Tree

When analyzing a set of data, there are *many* possible phylogenies to consider. We would like to identify a good (the best) phylogeny.

- I** CAAG
- II** GAAT
- III** CATG
- IV** GAAG



Constructing a Phylogenetic Tree

Character based methods are one type of approach for constructing a phylogenetic tree.

Character based methods are often based on the idea of *maximum parsimony*.

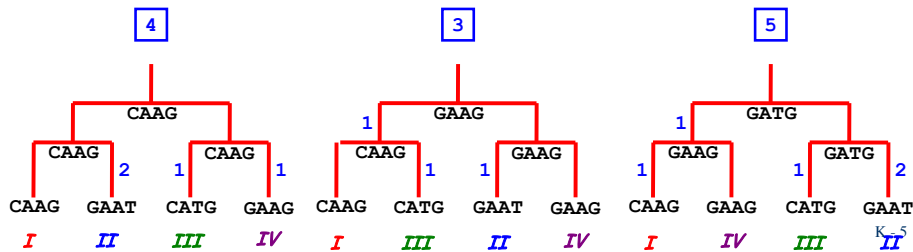
“ IT IS VAIN TO DO WITH MORE WHAT CAN BE DONE WITH FEWER”
 OR
 Principle of parsimony
 OR
 ...smallest number of evolutionary changes...

Optimality criterion: The ‘most parsimonious’ tree is the one that requires the fewest number of evolutionary events (*e.g.*, nucleotide substitutions, amino acid replacements) to explain the sequences.



Character Based Methods

- I** CAAG
- II** GAAT
- III** CATG
- IV** GAAG



Maximum Parsimony Methodology

Step 1: Identify informative sites

Sites with at least two different characters at the site, each of which is represented in at least two of the sequences

	Site								
Seq.	1	2	3	4	5	6	7	8	9
I	A	A	G	A	G	T	T	C	A
II	A	G	C	C	G	T	T	C	T
III	A	G	A	T	A	T	C	C	A
IV	A	G	A	G	A	T	C	C	T

↑ ↑ ↑

K-6

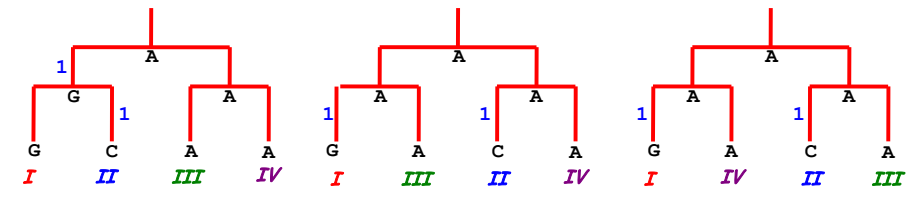


Sites Where All Trees Require the Same Number of Changes Are Not Informative

Tree I

Tree II

Tree III



	Site								
Seq.	1	2	3	4	5	6	7	8	9
I	A	A	G	A	G	T	T	C	A
II	A	G	C	C	G	T	T	C	T
III	A	G	A	T	A	T	C	C	A
IV	A	G	A	G	A	T	C	C	T

↑ ↑ ↑

K-7

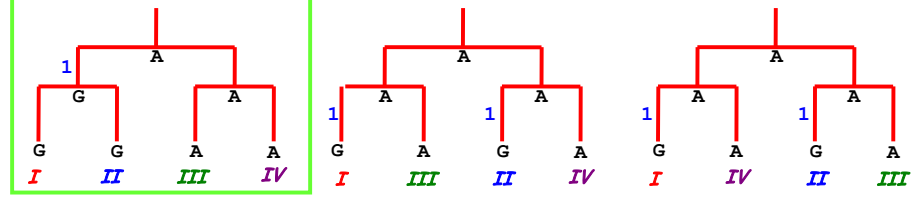


Maximum Parsimony Analyzes Sites At Which One Substitution Model Requires Fewer Changes

Tree I

Tree II

Tree III



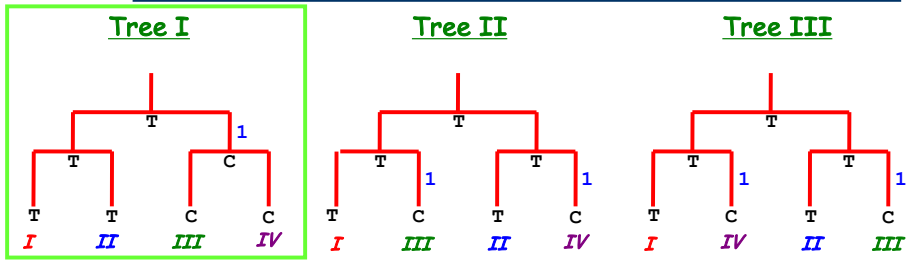
	Site								
Seq.	1	2	3	4	5	6	7	8	9
I	A	A	G	A	G	T	T	C	A
II	A	G	C	C	G	T	T	C	T
III	A	G	A	T	A	T	C	C	A
IV	A	G	A	G	A	T	C	C	T

↑ ↑ ↑

K-8



Maximum Parsimony Analyzes Sites At Which One Substitution Model Requires Fewer Changes

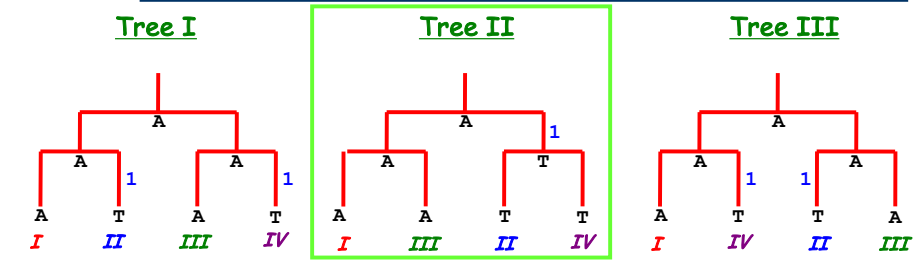


	Site								
Seq.	1	2	3	4	5	6	7	8	9
<i>I</i>	A	A	G	A	G	T	T	C	A
<i>II</i>	A	G	C	C	G	T	T	C	T
<i>III</i>	A	G	A	T	A	T	C	C	A
<i>IV</i>	A	G	A	G	A	T	C	C	T

K - 9



Maximum Parsimony Analyzes Sites At Which One Substitution Model Requires Fewer Changes



	Site								
Seq.	1	2	3	4	5	6	7	8	9
<i>I</i>	A	A	G	A	G	T	T	C	A
<i>II</i>	A	G	C	C	G	T	T	C	T
<i>III</i>	A	G	A	T	A	T	C	C	A
<i>IV</i>	A	G	A	G	A	T	C	C	T

K - 10



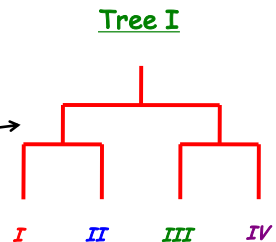
Maximum Parsimony Methodology

Step 2: Calculate minimum number of substitutions at each informative site

Step 3: Sum number of changes at each informative site for each possible tree

The tree with the least number of total changes is the most parsimonious tree

	Number of Changes at Each Informative Site			
	5	7	9	Σ
Tree I	1	1	2	4
Tree II	2	2	1	5
Tree III	2	2	2	6

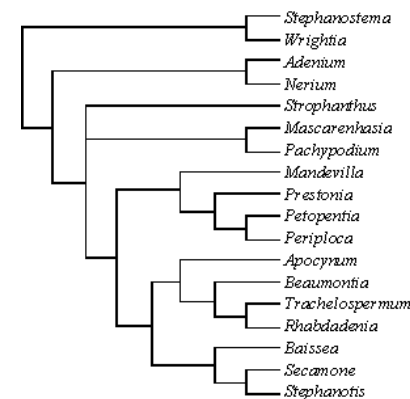


K - 11



How Confident Are We In Our Tree?

Bootstrapping. Given a particular tree, how consistently does a tree-building algorithm find that branching order using a randomly sampled version of the original dataset?



K - 12



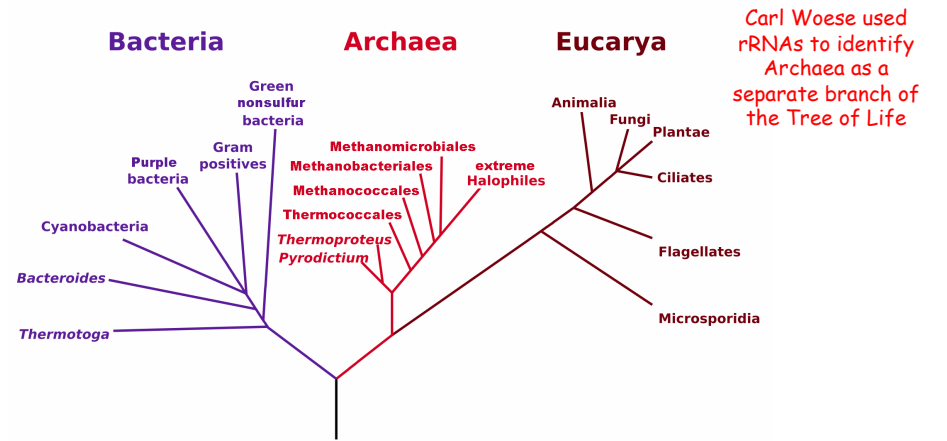
What Can Phylogeny Do For You?

Why do we care about evolution and the evolutionary history of organisms?

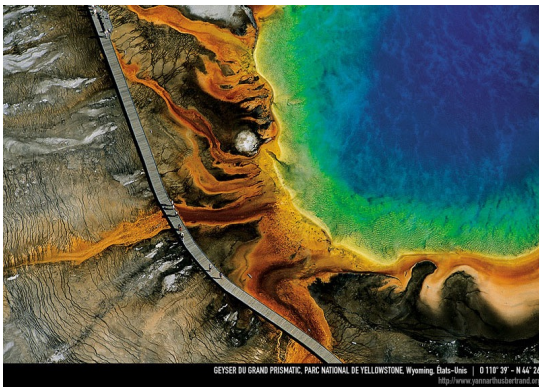


Revolutionizing the Tree of Life

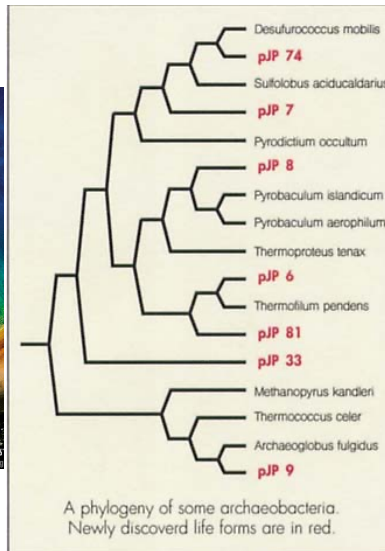
Phylogenetic Tree of Life



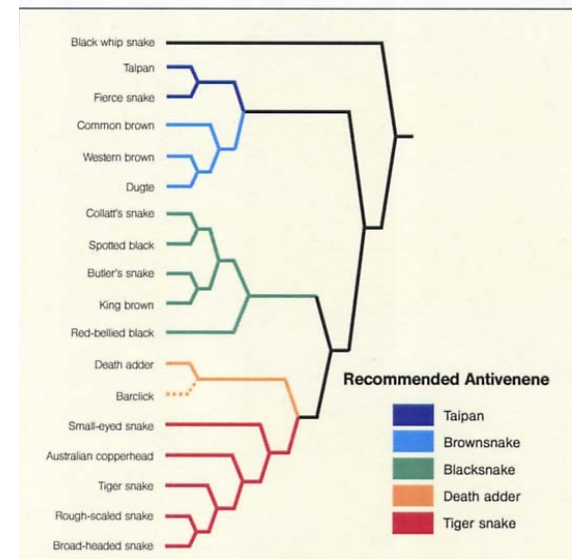
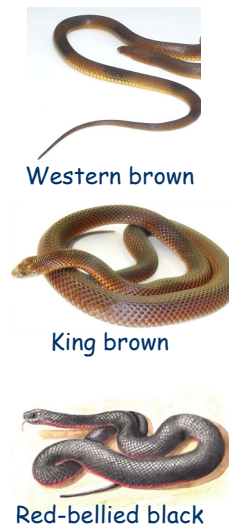
Discovering New Life Forms



GEYSER DU GRAND PRISMATIC, PARC NATIONAL DE YELLOWSTONE, Wyoming, États-Unis | 0 11° 39' - N 44° 28' | <http://www.parcnationalyellowstone.com>



Developing Effective Snakebite Antivenins

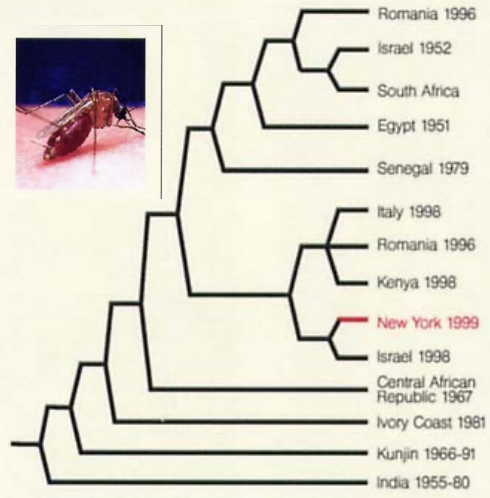




Identifying Emergent Diseases



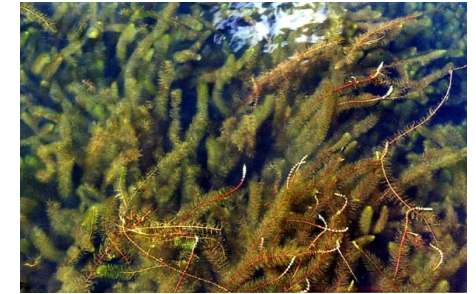
Several species of mosquitos, including the recently invasive tiger mosquito, transmit West Nile virus from bird and mammal hosts to human populations. Understanding the phylogenetic relationships of disease vectors such as mosquitos is important in tracing the origin and spread of diseases.



Protecting Ecosystems From Invasive Species



Purple loosestrife



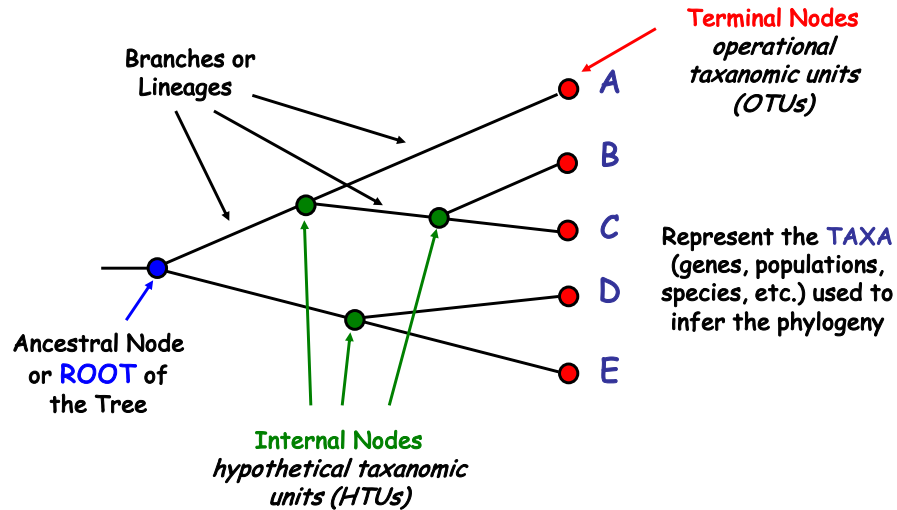
Eurasian water milfoil



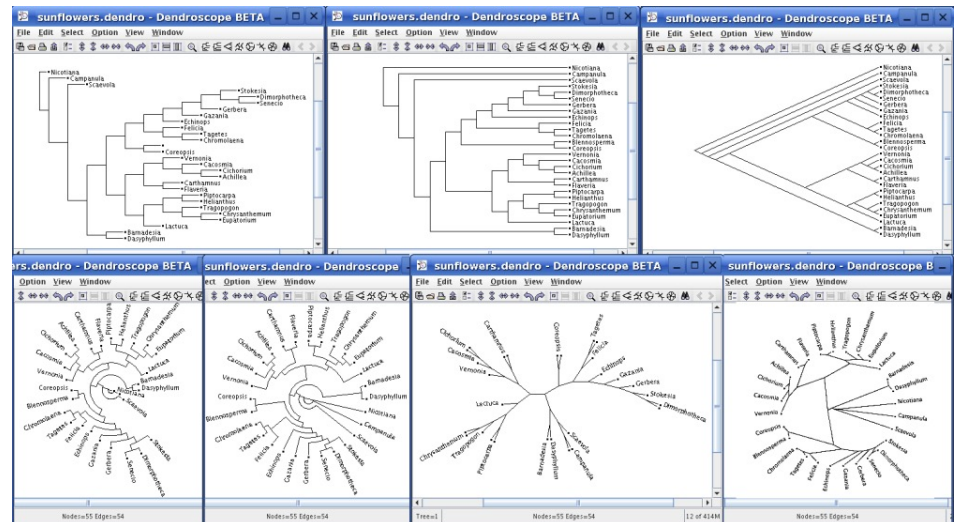
Caulerpa taxifolia



Common Phylogenetic Tree Terminology

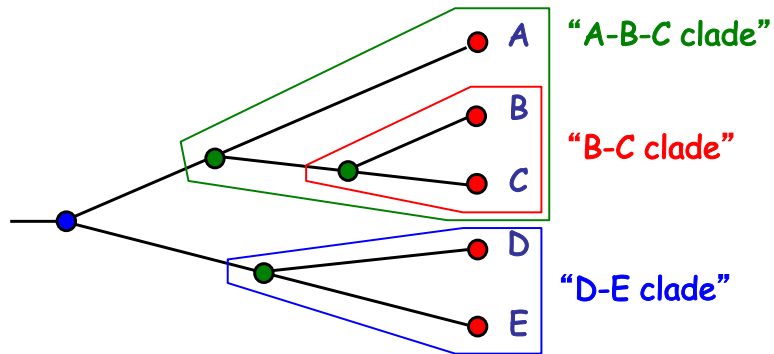


Phylogenetic Trees Can Be Drawn Many Ways





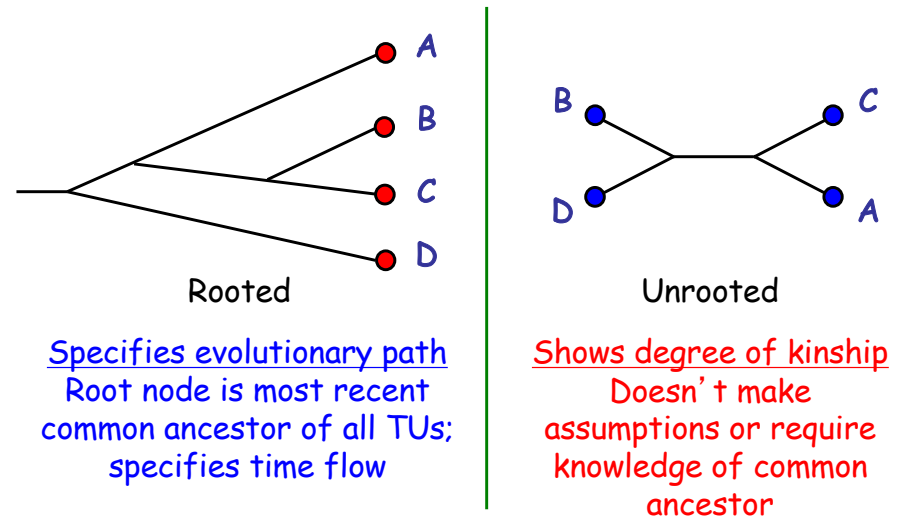
Clade: group with a single common ancestor and its descendants



K - 21



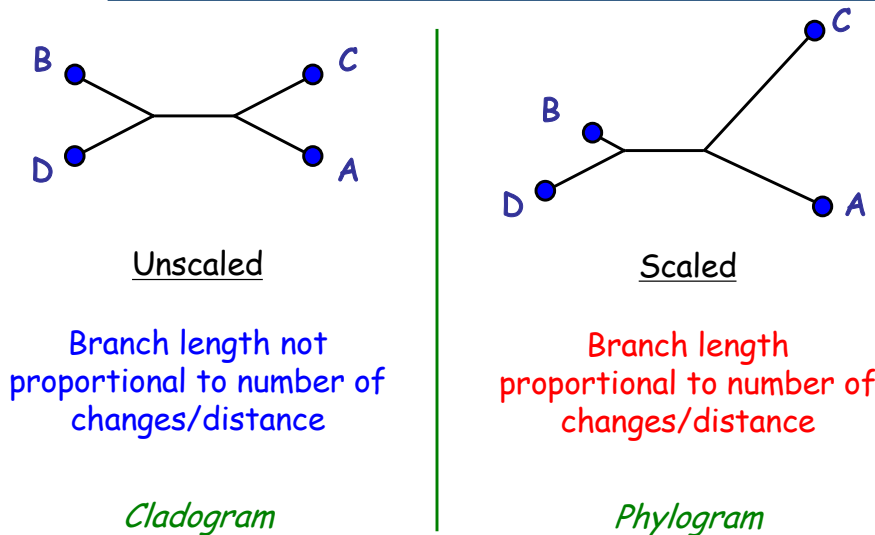
Phylogenetic Trees Can Be Rooted or Unrooted



K - 22



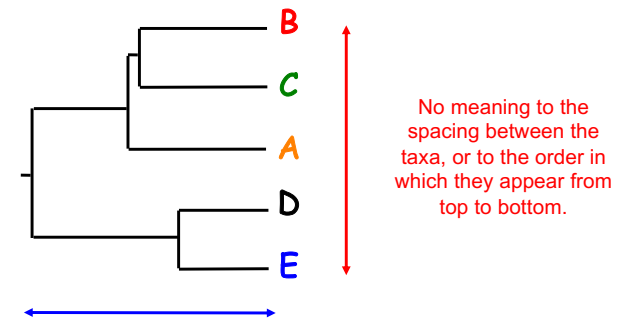
Phylogenetic Trees Can Be Scaled or Unscaled



K - 23



Phylogenetic Trees Diagram Evolutionary Relationships

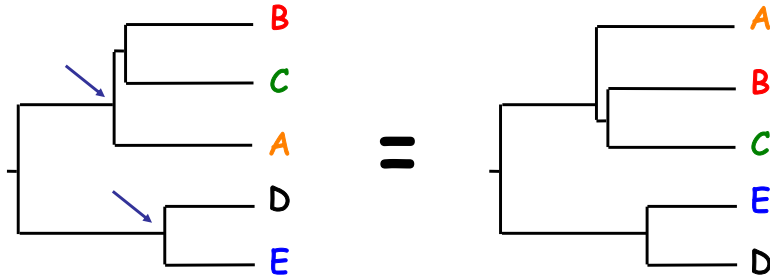


- 1) No scale (cladograms)
- 2) Proportional to genetic distance (phylograms)
- 3) Proportional to time (ultrametric trees)

K - 24



Rotating Clades: Same Meanings

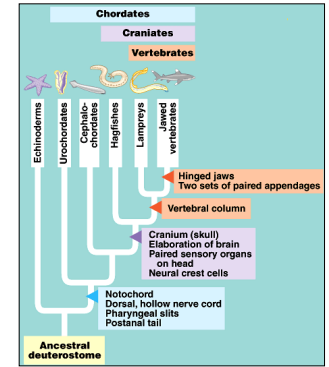
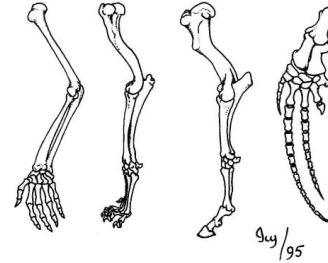


K - 25



How Are Phylogenetic Trees Built?

Traditionally: use homologous structures



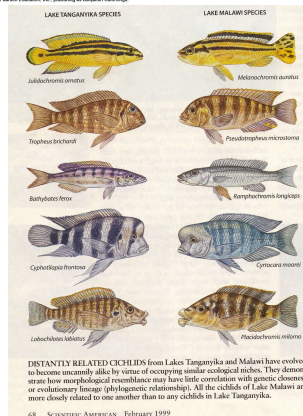
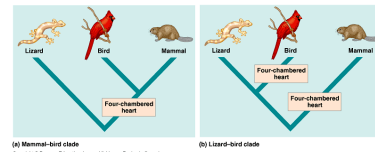
Caveats:

- Closely related organisms don't always look similar
- Similar looking organisms not always closely related
- How do you decide importance of traits?

K - 26

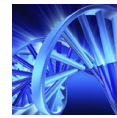


Structural Analogy Can Result From Convergent Evolution



DISTANTLY RELATED CICHLIDS from Lakes Tanganyika and Malawi have evolved to become uncannily alike by virtue of occupying similar ecological niches. They demonstrate how morphological resemblance may have little correlation with genetic distance or evolutionary lineage (phylogenetic relationship). All the cichlids of Lake Malawi are more closely related to one another than to any cichlid in Lake Tanganyika.

K - 27



Molecular Phylogenetic Trees

- Large molecular data sets: Bioinformatics!
- Molecular clock vs. punctuated equilibrium
- Eliminates analogy and trait selection issues
- Result: great improvement on classical phylogenies

Caveat:

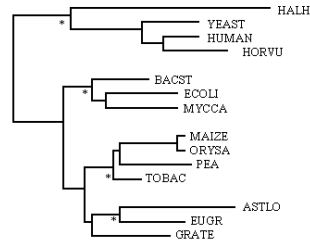
Gene divergence may not correlate with species divergence

K - 28



Molecular Phylogenies Can Be Constructed Using Different Elements

Nuclear genes
Mitochondrial DNA
Genome structure

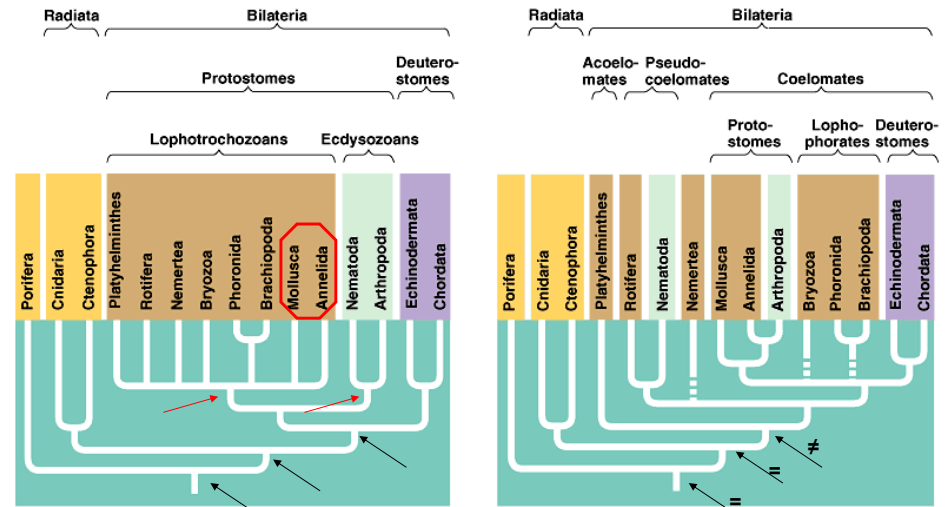


Usually integrate analyses of multiple different genes

K - 29



Molecular Comparisons vs. Body Plans



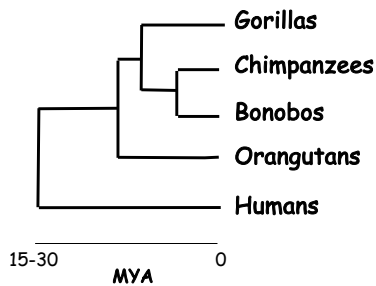
(a) Tree based on molecular comparisons

(b) Tree based on body-plan grades

Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

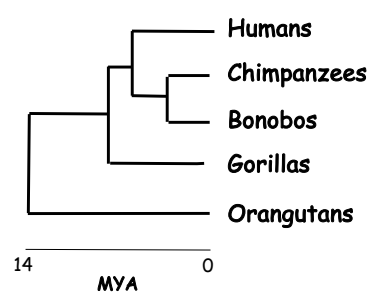


Which Species Are the Closest Living Relatives of Modern Humans?



Pre-molecular view

Great apes (chimpanzees, gorillas and orangutans) formed a clade separate from humans.



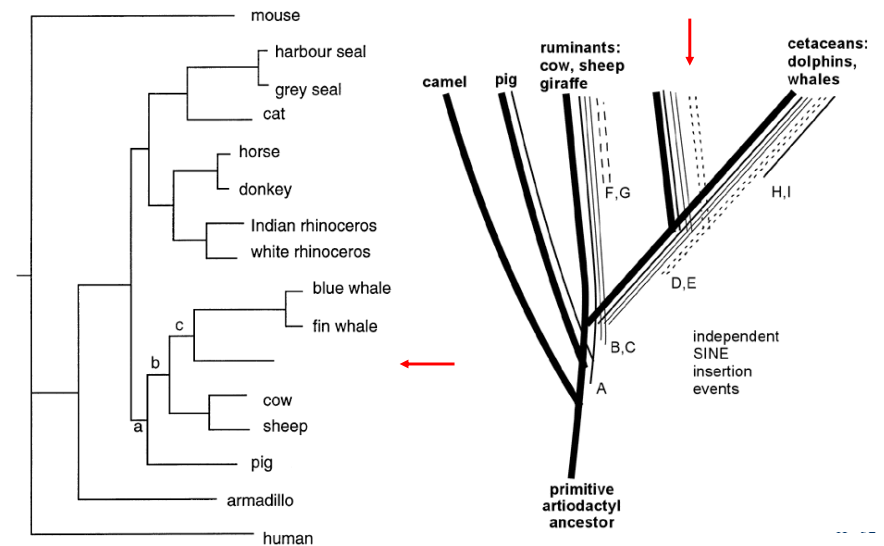
MitoDNA, most nuclear genes, and DNA hybridization

Bonobos and chimpanzees are related more closely to humans than either are to gorillas.

K - 31



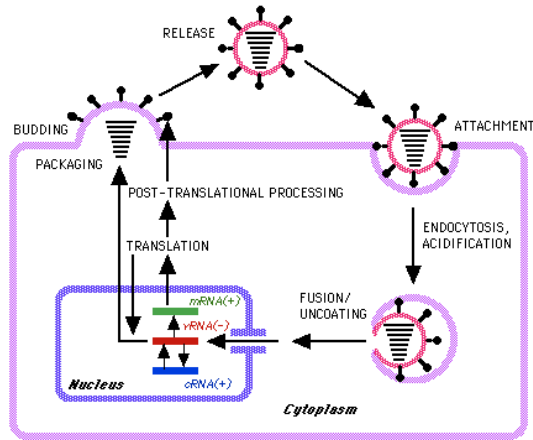
What is the Closest Living Relative of Whales?





Influenza Virus

- ssRNA genome, ~13,588 bases
- Genome in 8 segments, 10-11 genes



K - 33



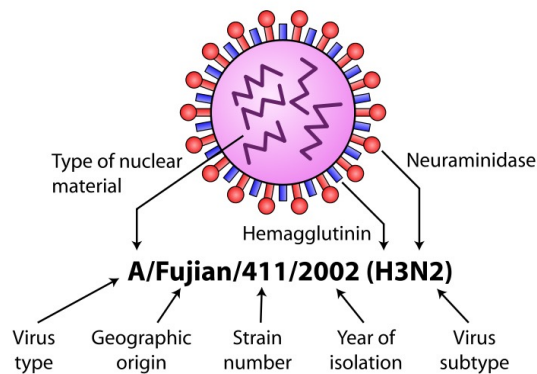
Influenza Virus Genes

Genome segment	Segment size (bases)	Gene(s)	Gene function
1	2341	PB2	Transcriptase: cap binding
2	2341	PB1	Transcriptase: elongation;
		PB1-F2	Induces apoptosis
3	2233	PA	Transcriptase: protease activity
4	1778	HA	Hemagglutinin: host cell recognition
5	1565	NP	Nucleoprotein: RNA binding; transcriptase complex; vRNA transport
6	1413	NA	Neuraminidase: release of virus
7	1027	M1	Matrix protein: major component of virion
		M2	Integral membrane protein - ion channel
8	890	NS1	Non-structural: RNA transport, splicing, translation. Anti-interferon.
		NS2	Non-structural: nucleus and cytoplasm, vRNA export (NEP)

K - 34



Influenza Nomenclature



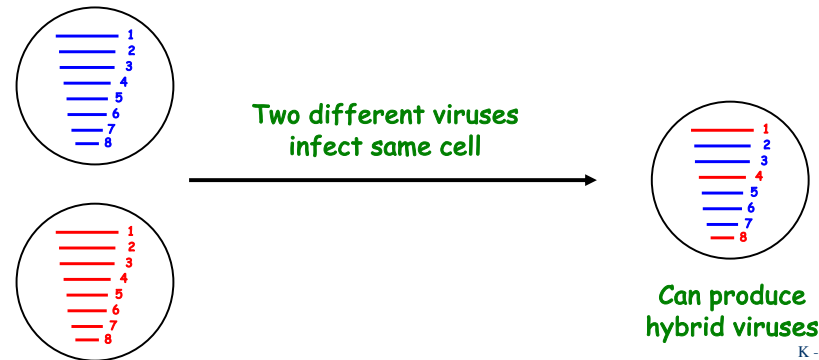
- Subtype nomenclature based on HA and NA genes
 - 16 Hemagglutinins, 9 Neuraminidases
- Human: H: 1,2,3 ; N: 1,2; Birds: all combinations

K - 35



Influenza Virus Can Change Rapidly

- High mutation rate (antigenic drift)
- Reassortment (antigenic shift)



K - 36



Reassortment Can Produce Pandemic Influenza Viruses

- **1957 Asian flu:** H2N2, 3 avian flu segments, 5 human flu segments
- **1968 Hong Kong flu:** H3N2, 2 avian flu segments, 6 human flu segments
- **Reassortment in pigs** - susceptible to avian, human, and swine flus

K - 37



1918 Influenza Pandemic

- Highly virulent flu virus (“Spanish flu”)
- Estimated deaths: 50-100 million worldwide (of 1.8 billion)
- Many people died within a few days from acute pneumonia
- Many fatalities were young and healthy people
- Lowered average U.S. life expectancy by 10 years

K - 38



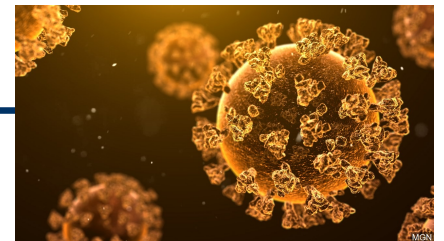
1918 Influenza Questions

- Where did the 1918 flu come from?
- Why was the 1918 flu so pathogenic?
- Is it possible for a 1918-like pandemic to happen again?

K - 39



SARS-CoV-2 Virus



- ssRNA genome: 30,000 NTs, 11 genes
- Recombination events between bat and pangolin coronaviruses
- Half a billion cases, 6 million deaths
- Substitution rate: 1 in 1,000
- Variants: Delta, Omicron, etc.
- Spike proteins

3