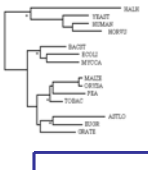


Multiple Sequence Alignment



Sequences

> Yeast YOR020c
mstllksaaksiivplmdrvlvqrikaqaktasglylpe
knveklngaevvavvggftdangnkvvqvkvgdqvl
ipqfggstiklgnddvilfrdaeilakiakd

> Neurospora crassa
mattvsvksliplldrvlvqrvkaeaktasgfilpe
ssvkdlineakvlavpgaldkdgkrlpimgvماغdrvl
ipqyggspvkvgeeeylfrdseilakiae

> Aspergillus nidulans
msllrnvknlaplldrvlvqrvkpeaktasgfilpes
svkeqmeakvlavpgavdrngqripmgvaagdrvlv
pqfggsplkigeeyhlfrdseilakine

> Schizosaccharomyces pombe (fission yeast)
matllksaaksiivplmdrvlvqrikaadtktasgfilpe
ksveklasegrvisvgkgyknegkklagsvavagdrvl
lpaygggnikvgeeeyslyrdhellaiike

> Mortierella alpina
masritkfsktivpnmmdrvlvqrikpqktaggiyip
ekaqaenalnegyvavvgkllttqeqkvpselaegdkv
llppygsvkvvdneelilfrdseilakiq

> Cryptocodium cohnii
matgiakrftplldrvlvqrlkpeaktasgflpesa
akapnyatvlavpggrtrdgdilpmnvkvgdkvvvp
eyggmtlkfedeeefqvfrdaaimgilne

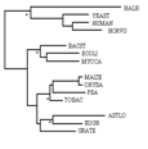
> Drosophila melanogaster
maaaikkiipmldriliqraealtktkkggivilpekav
gkvlqatvavvggskgkgsgeiqpvsvkvgdkvllpe
efggtkvnllegdqkelflfrdsilakle

> Homo sapiens
aggafkrflplfdrvlversaaetvtkkggimlpeksq
gkvlqatvavvggskgkgsgeiqpvsvkvgdkvllpe
yggtkvvllddkdyflfrdsgxilgky

> Geobacillus stearothermophilus
vlkplgdrvievietektaggivilpdtakekpgqg
rvvavgkgrvldsgervapevevgdriifskvagtcev
kydgkaylilresdilavig

> Mycobacterium tuberculosis
makvnikpledki1lvqaneatttaggivilpdtakek
ppegtrvavvggrwdedgekripldvaegdtviysky
ggteikynggeeylilsardvlavsvk

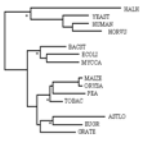
> Mus musculus (house mouse)
magqafkrfillfdrvlversaaetvtkkggimlpeks
gkvlqatvavvggskgkgsgeiqpvsvkvgdkvllp
eyggtkvvllddkdyflfrdsilgkyv



Pairwise Alignment

		A	G	G	T	T	C
		0	0	0	0	0	0
A		0	5	0	0	0	0
G		0	0	10	5	0	0
G		0	0	4	15	9	3
C		0	0	0	9	11	5
A		0	5	0	3	5	7
T		0	0	1	0	8	10
T		0	0	0	0	5	13
C		0	0	0	0	0	1
							18

AGG--TTC
 ||| |||
 AGGCATTC



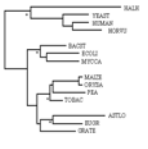
3-sequence Alignment

AGA

AGT

TCC

		A	G	T
	T	0	0	0
	C	0	0	0
A		0	5	0
G		0	0	10
A		0	5	4



Sequences

> Yeast YOR020c
mstllksaksivplmdrvlvqrikaqaktasgylpe
knveklngaevvavpggftdangnkvvqkvgdqvl
ipqfggstiklgnddevilfrdaeilakiakd

> Neurospora crassa
mattvsvksaliplldrvlvqrvkaeaktasgflpe
ssvkdlineakvlavpggaldkdgkrlpmgvnagdrvl
ipqyggspvkveeeytlfdrseilakiae

> Aspergillus nidulans
msllrnvknlaplldrvlvqrvkpeaktasgflpes
svkeqneakvlavpggavdrngqripmvaagdrvlv
pqfggsplkigeeeyhlfdrseilakine

> Schizosaccharomyces pombe (fission yeast)
matkksaksivplldrilvqrikadtktasgflpe
ksveklsagrvisvqkgyngkglagpsvavgdrvl
lpayggsnikvgeeeyslyrdhellaiike

> Mortierella alpina
masritkfsktivpmmdrvlvqrikpqqktasgiyip
ekaqealnegyvavvgklttqegkvpselaegdkv
llpyggsvvkvdneelilfrseilakiq

> Crypthecodinium cohnii
matgiakrftplldrvlvqrlkpeaktasgflpesa
akapnyatvlavpggtrrdgdilpmnvkvgdkvvvp
eyggmtlkfedeeefqvrddadimgilne

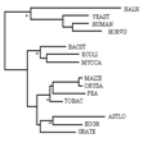
> Drosophila melanogaster
maaaikkiipmldrilligraealtktkkgivlpekav
gkvlegtlavpggtrnastgnhipigvkegdrvlp
efggtkvnlegdqkelflfresdilakle

> Homo sapiens
aggqfrkflplfdrvlversaaetvtkggimlpeksq
gkvlqatvvavsggskgkgeiepvsvkvgdkvllp
yggtkvlddkdyflfrdsilgky

> Geobacillus stearothermophilus
vlkplgdrvvievieteeektasgvlpdtakekpgqg
rvvaavggrvldsgervapevegdriifskyaagte
kydgkeylilresdilavig

> Mycobacterium tuberculosis
makvnikpledakilvqanaettasgvlvipdtakek
pqegtvvavpggrwdegedekripldvaegdtviysky
ggteikynggeeylilsardvlavvsk

> Mus musculus (house mouse)
magqafrkfllldrvlversaaetvtkggimlpeks
gkvlqatvvavsggskgkgeiepvsvkvgdkvllp
eyggtkvlddkdyflfrdsilgkyvn



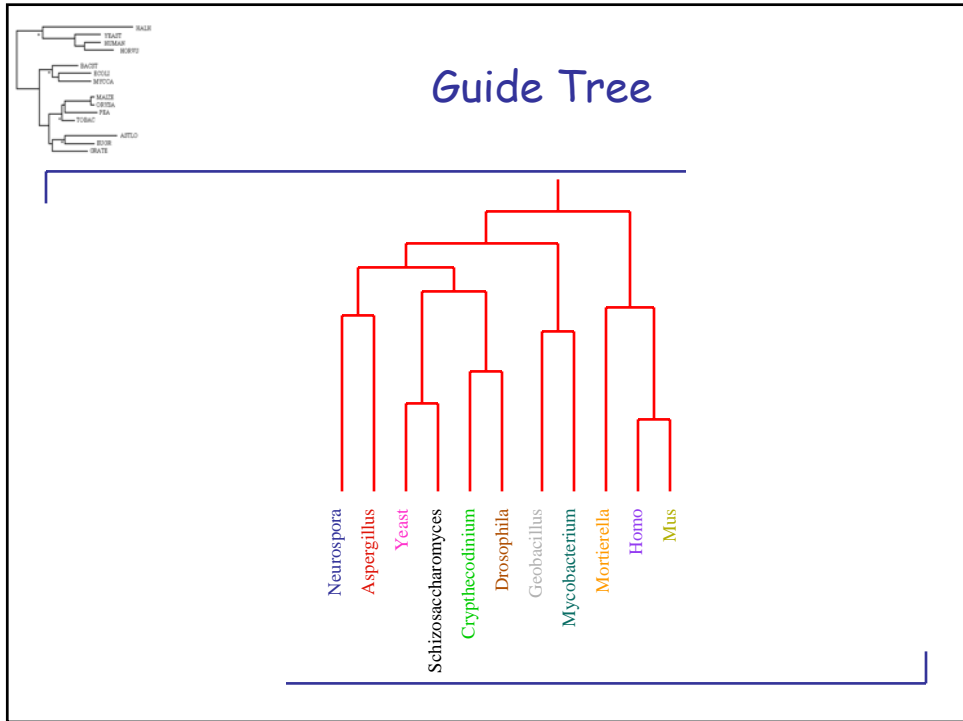
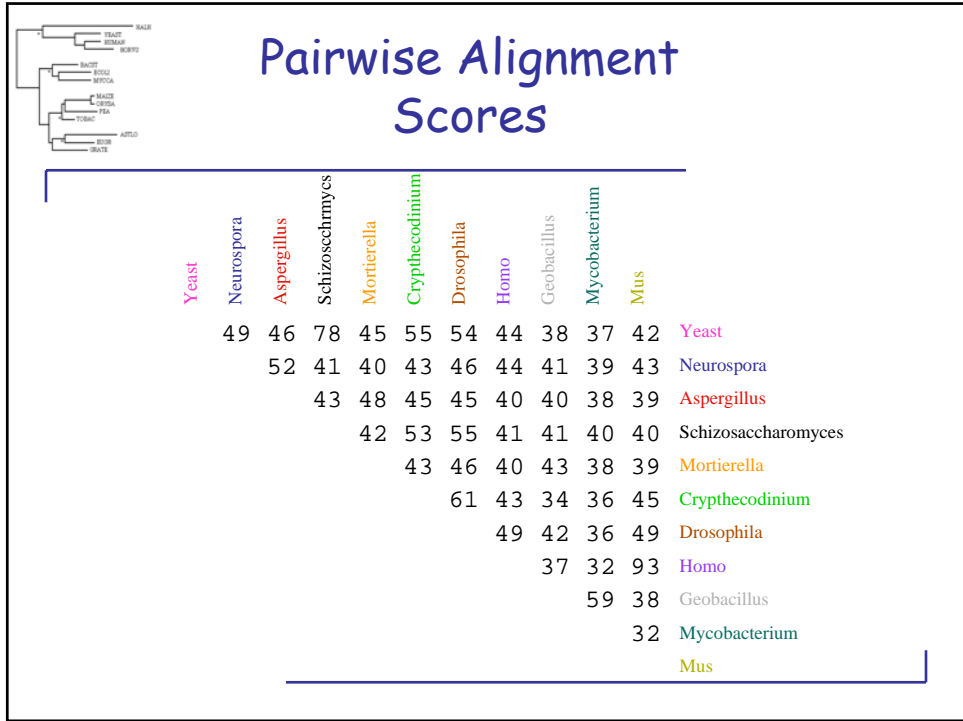
Multiple Sequence Alignment

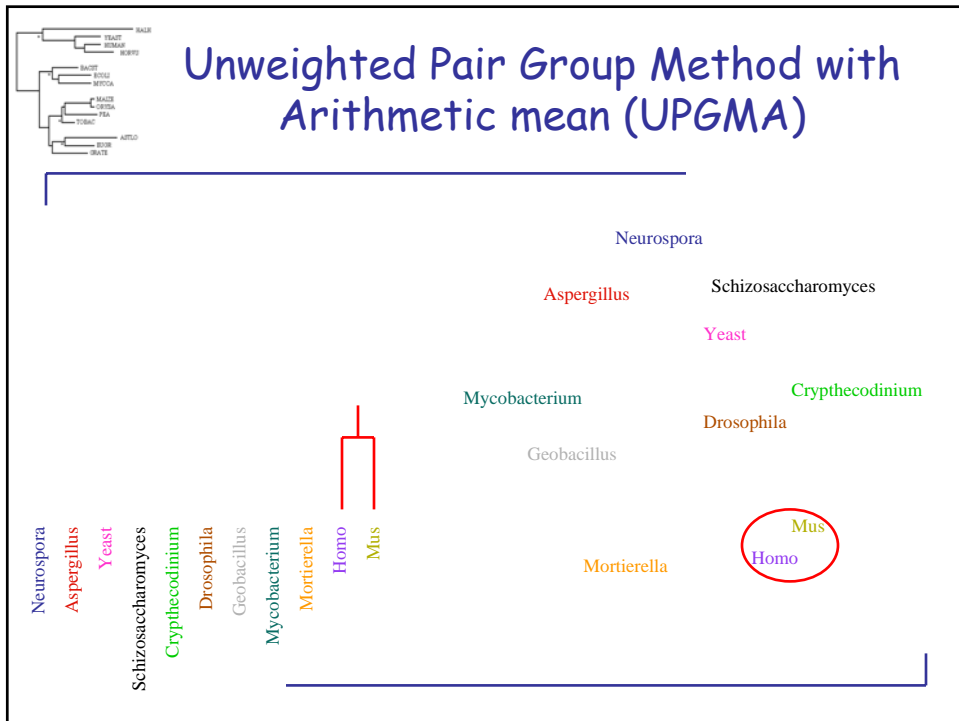
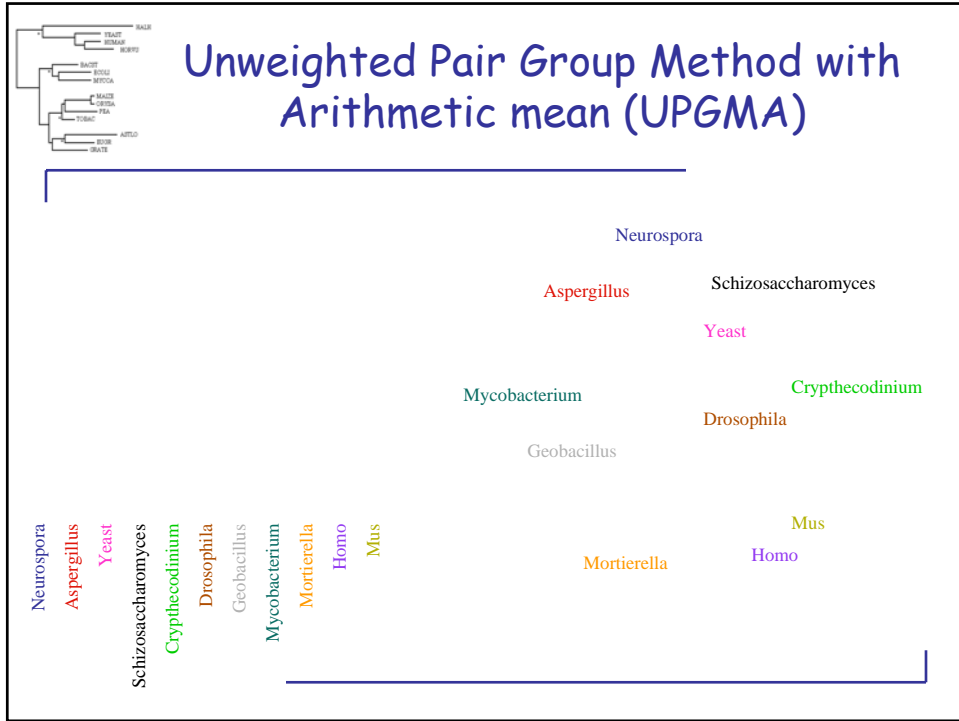
Homo/1-109
Mus/1-109

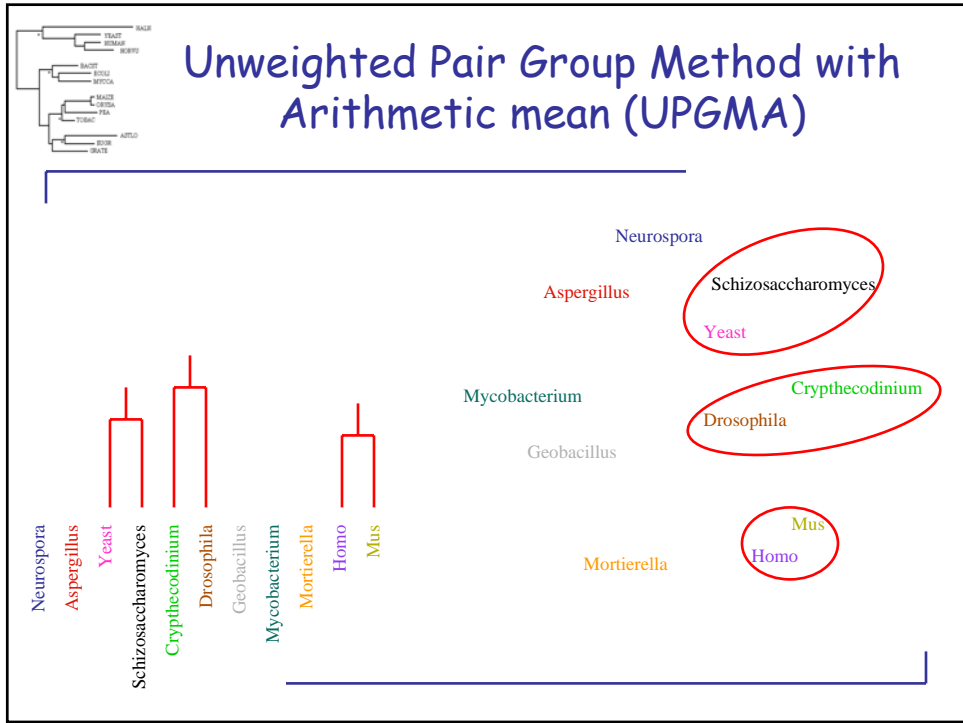
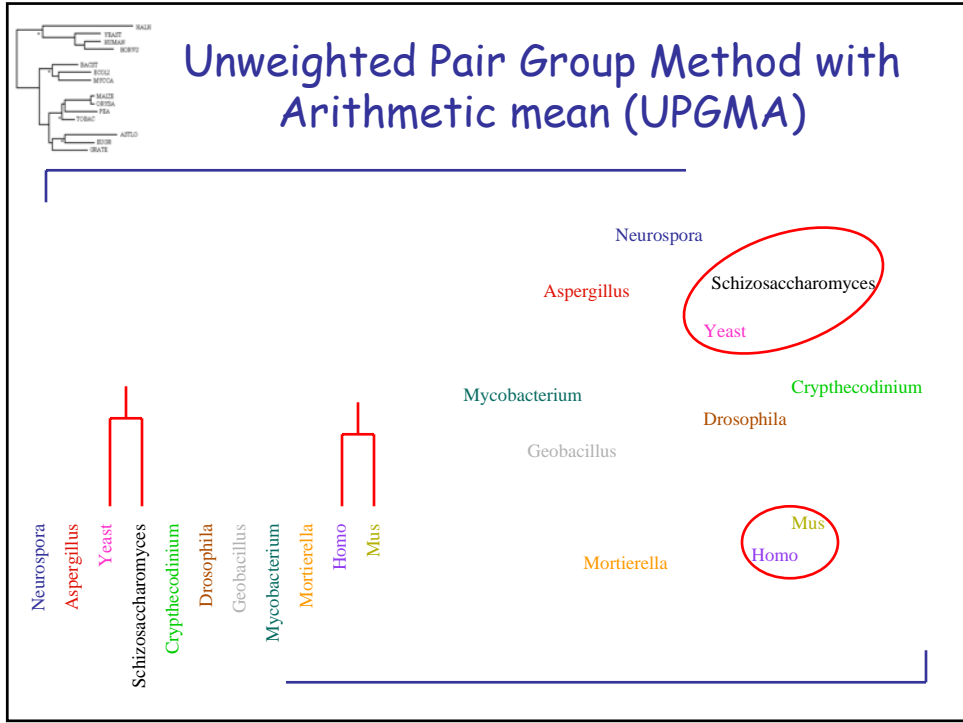
```

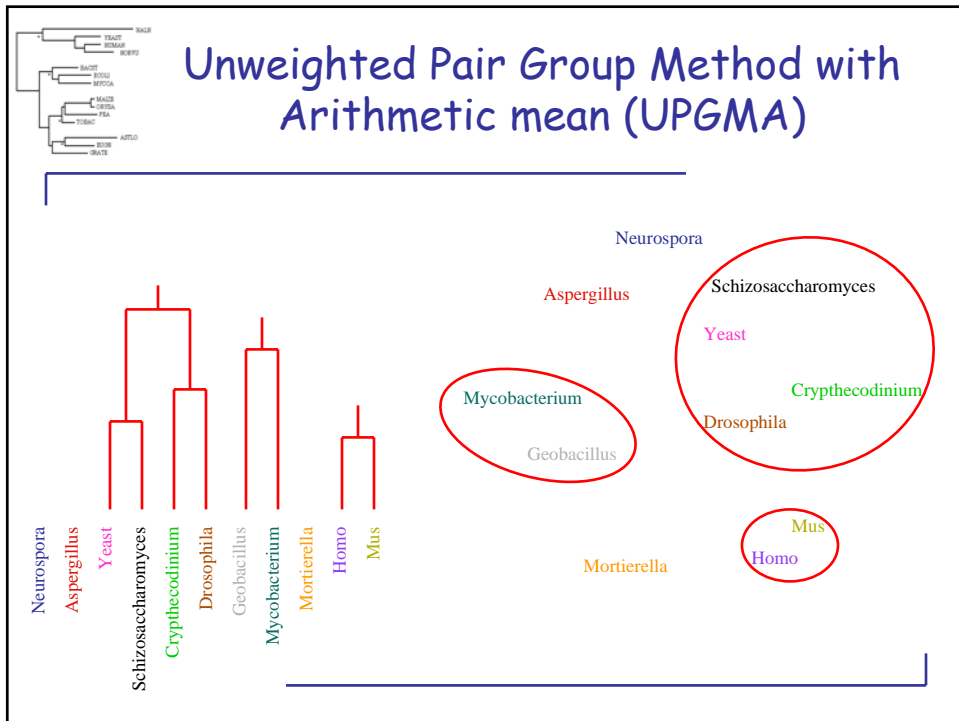
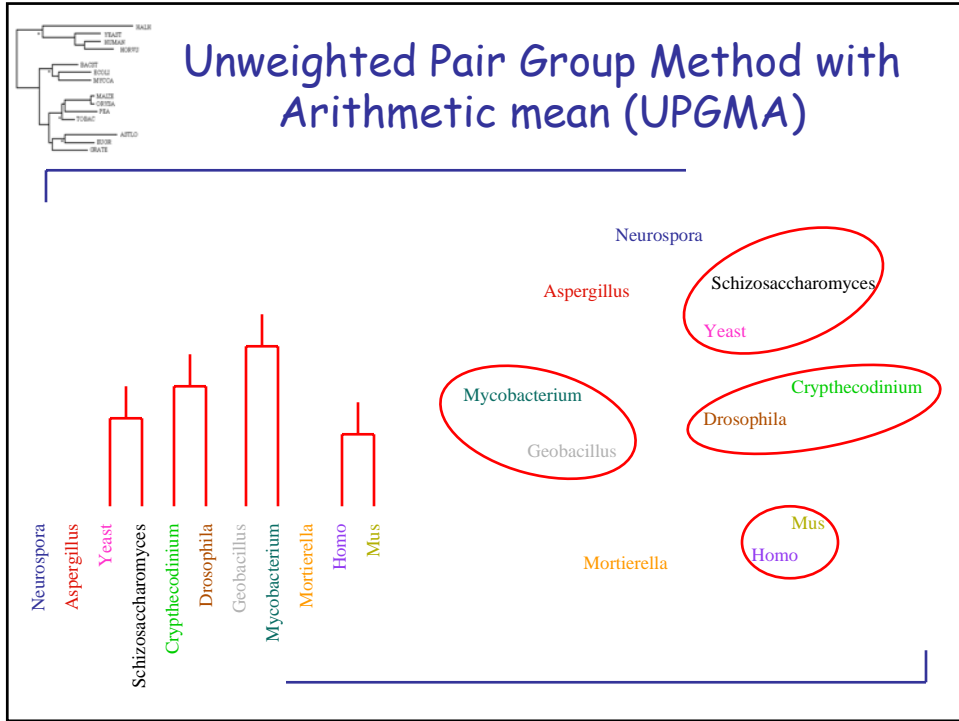
          10      20      30      40      50      60      70      80      90     100
----AGQAFRKFLLPLFDRVLVERSAAEVTKGGIMLPEKSGKRVLQATVVAVSGSK-GRGGEIQPVSVKVGDEVLLPEYGGTKVLD--DKDYFLFRDGIKGY--
---MAGQAFRKFLLPLFDRVLVERSAAEVTKGGIMLPEKSGKRVLQATVVAVSGSK-GRSGEIEFPVSVKVGDEVLLPEYGGTKVLD--DKDYFLFRDSDILGKYVN-

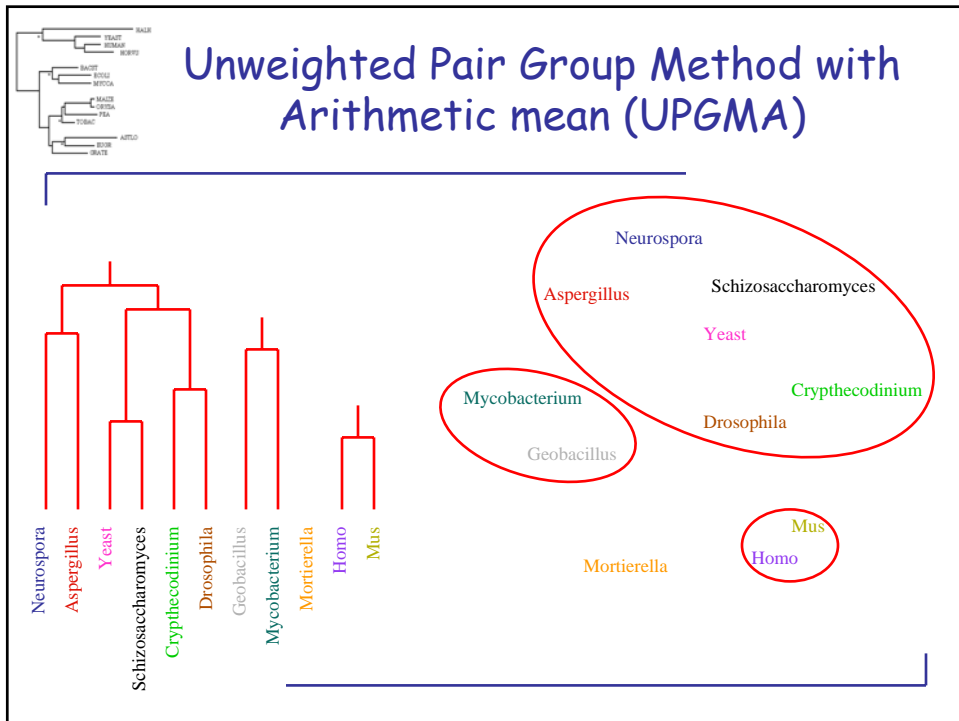
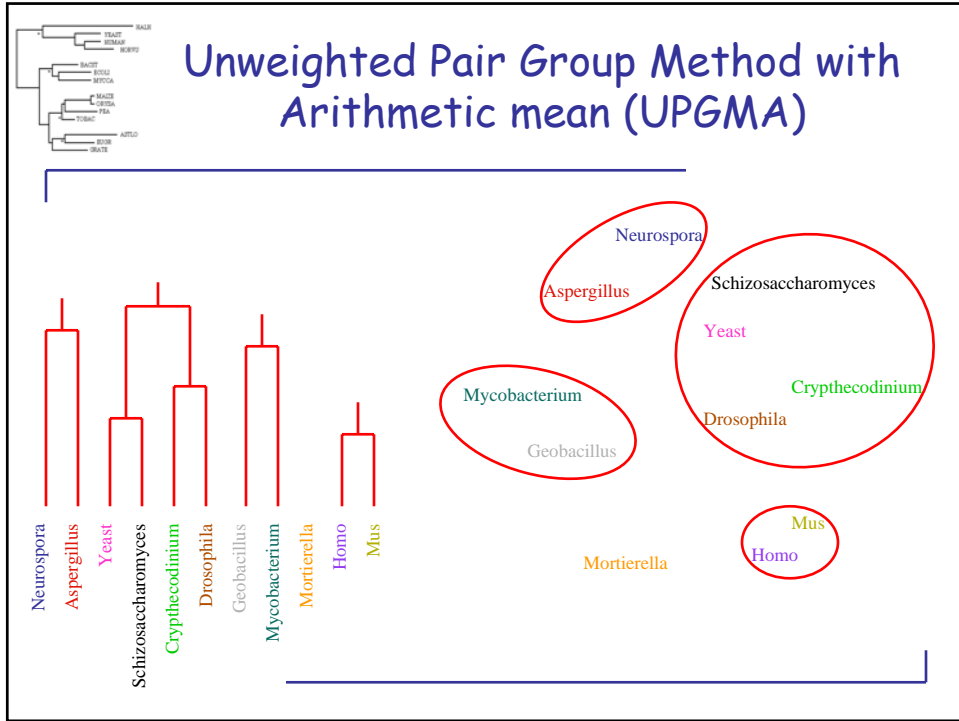
```

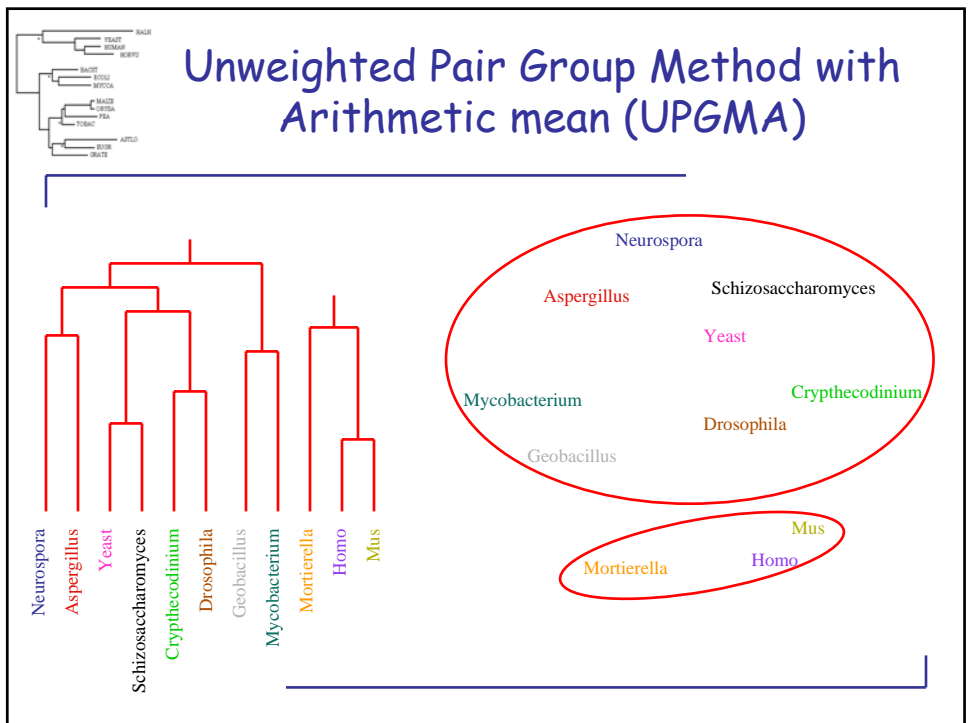
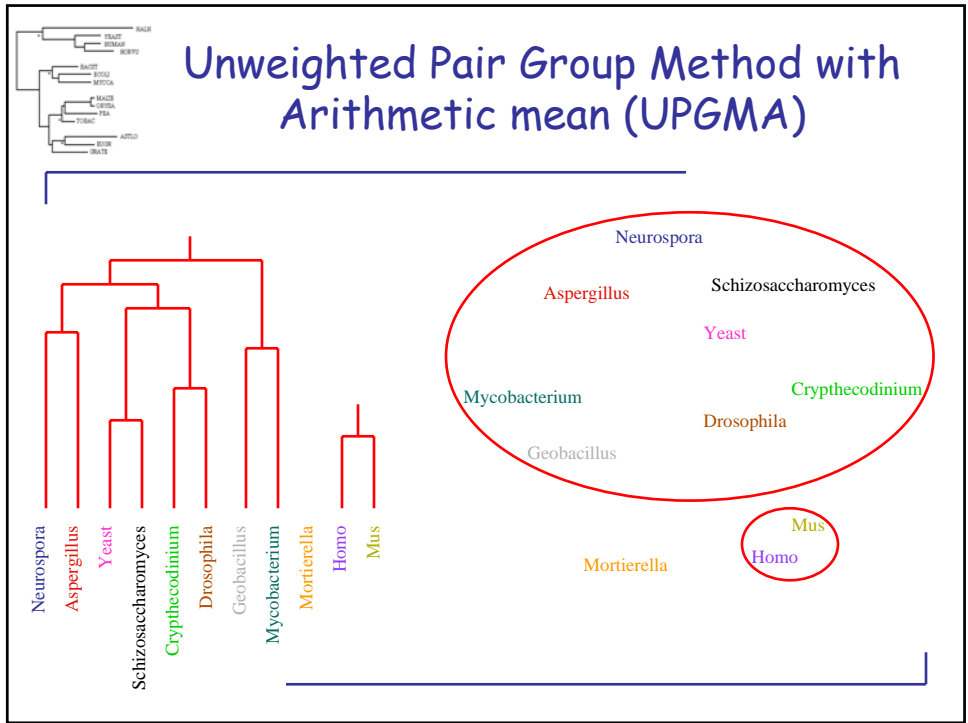


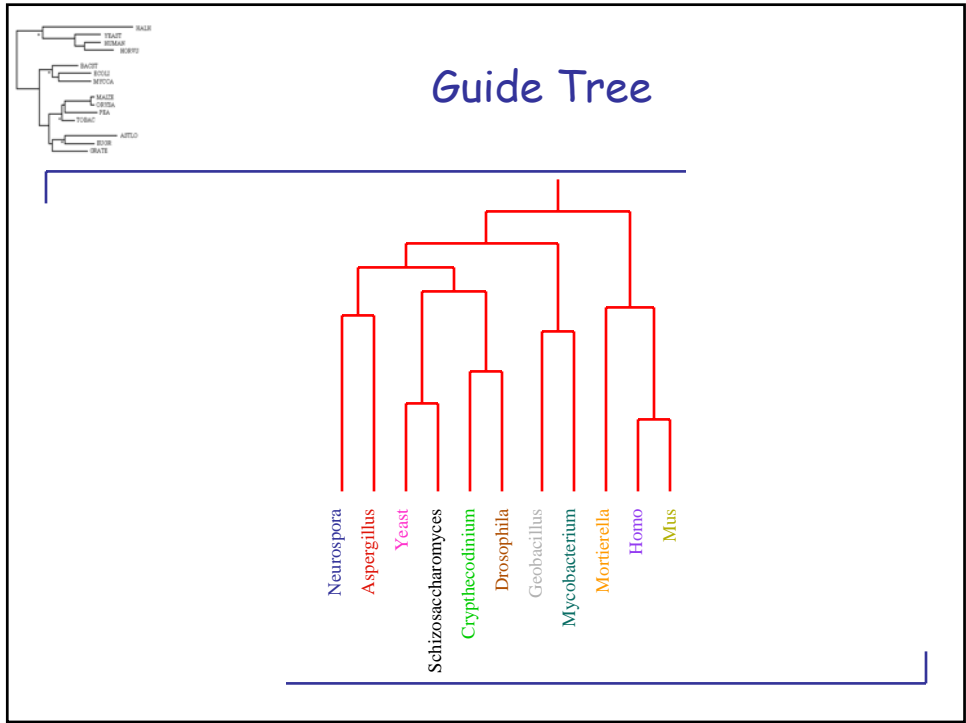
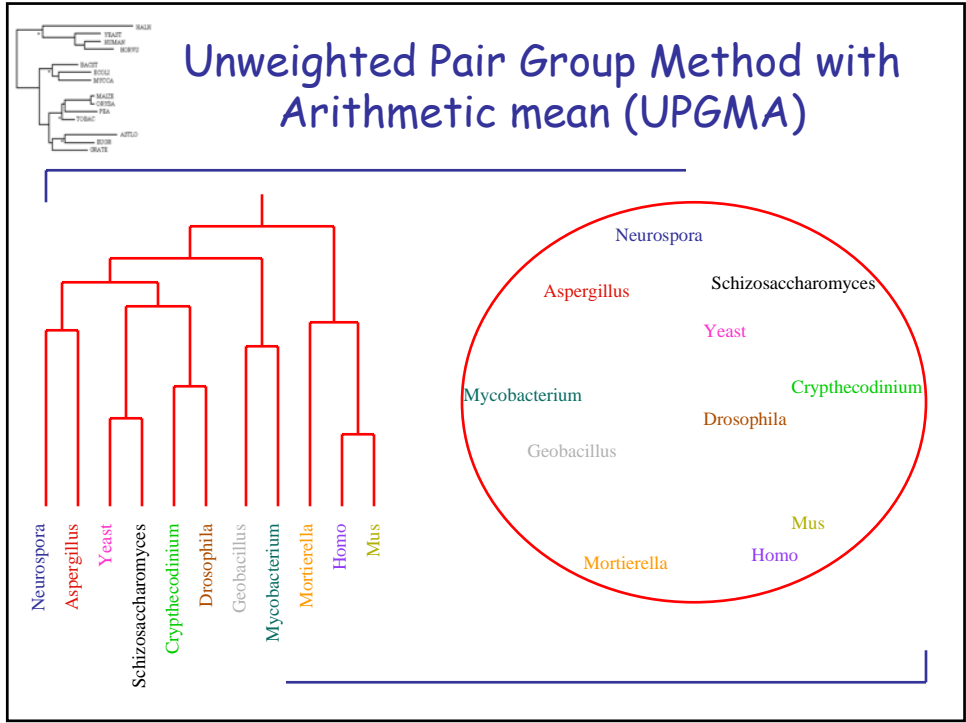


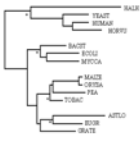






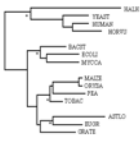




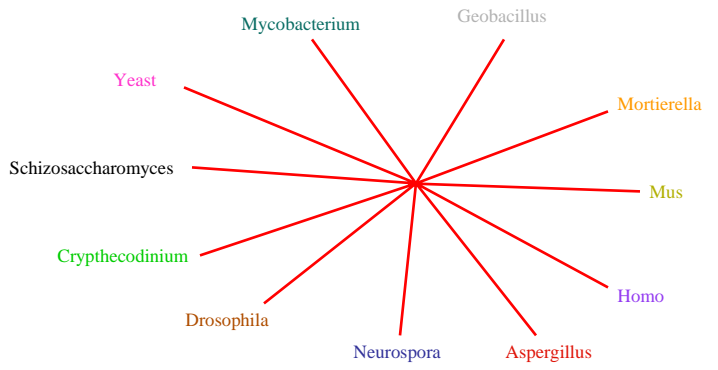


Neighbor Joining (NJ)

- Generate full tree with starlike structure
- Repeat the following step
 - Connect two closest groups (i.e., neighbors) through a single node

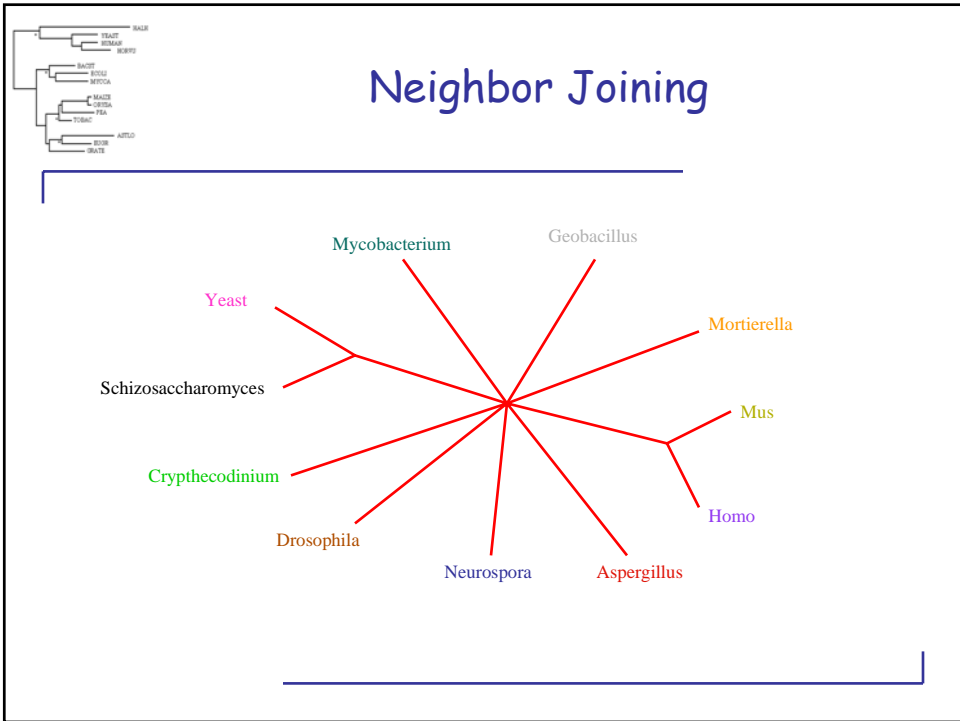
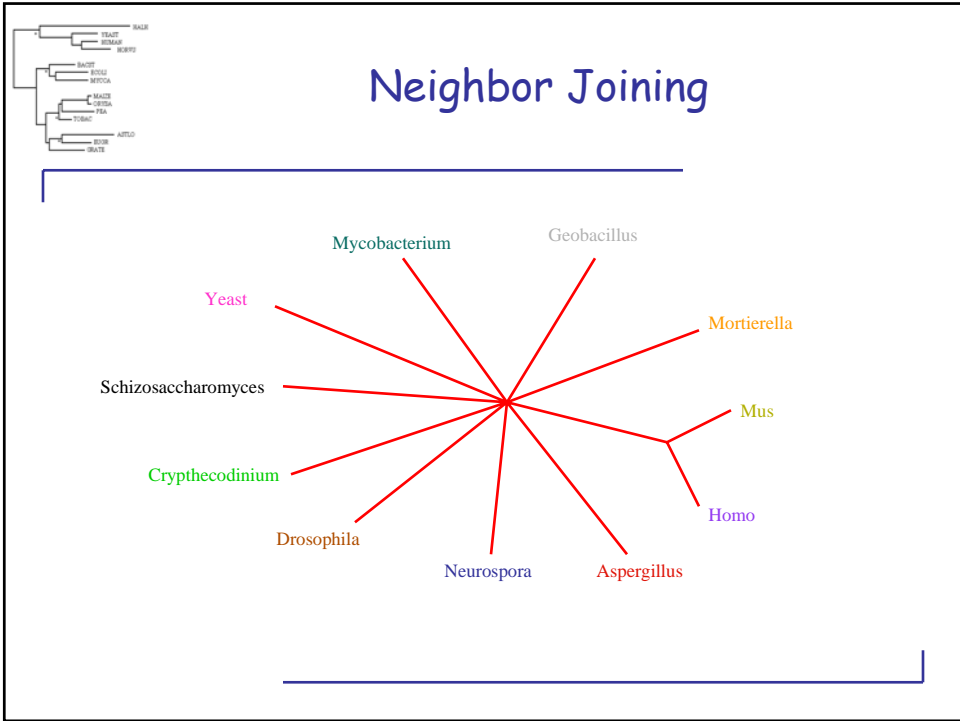


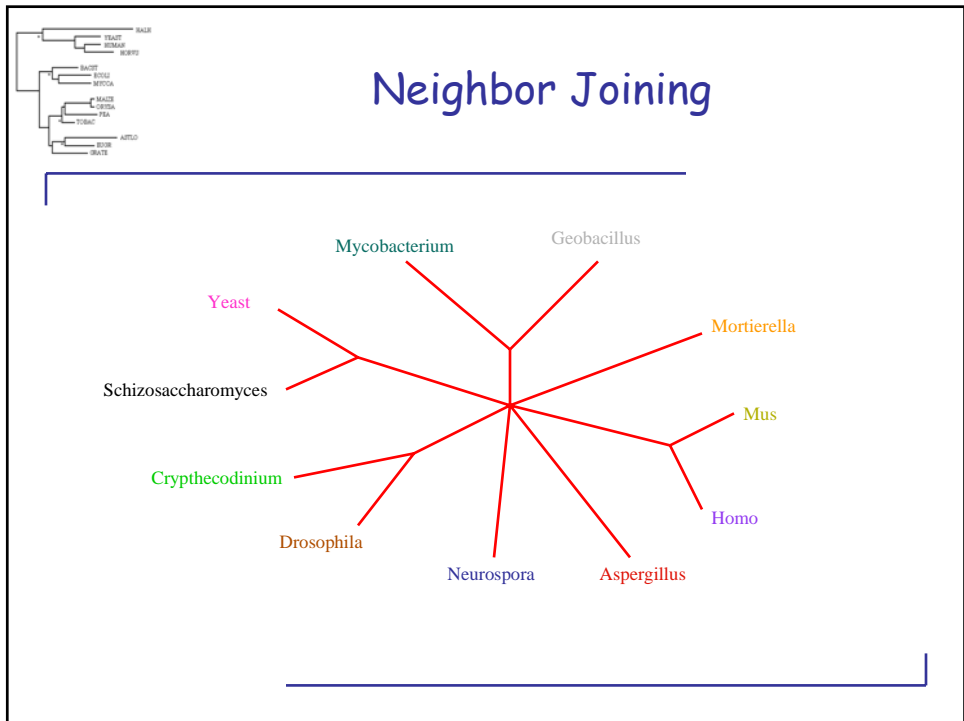
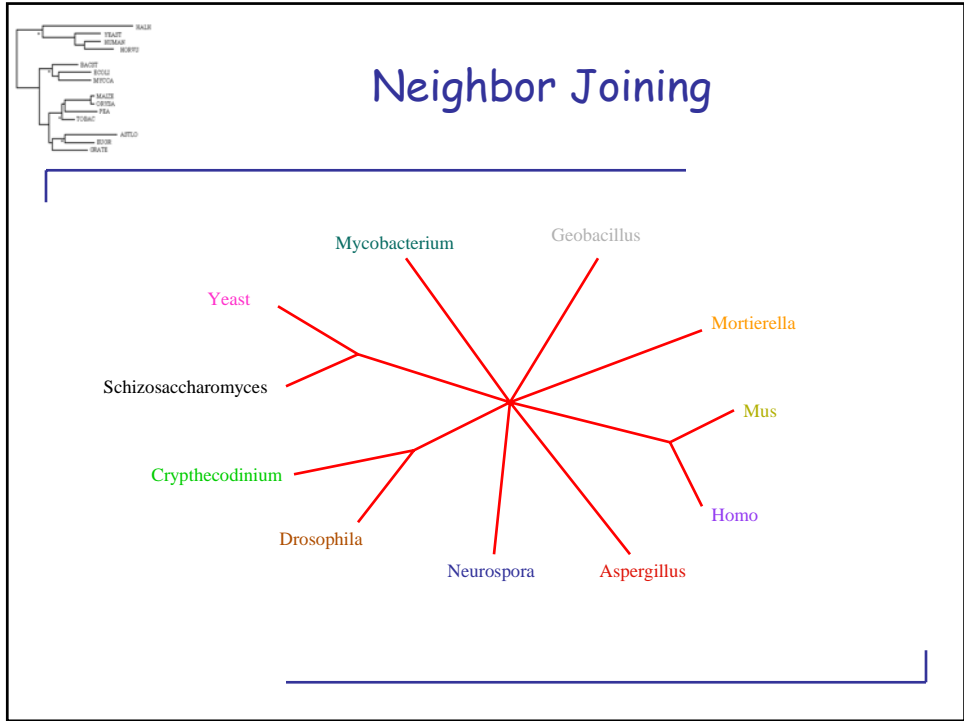
Neighbor Joining

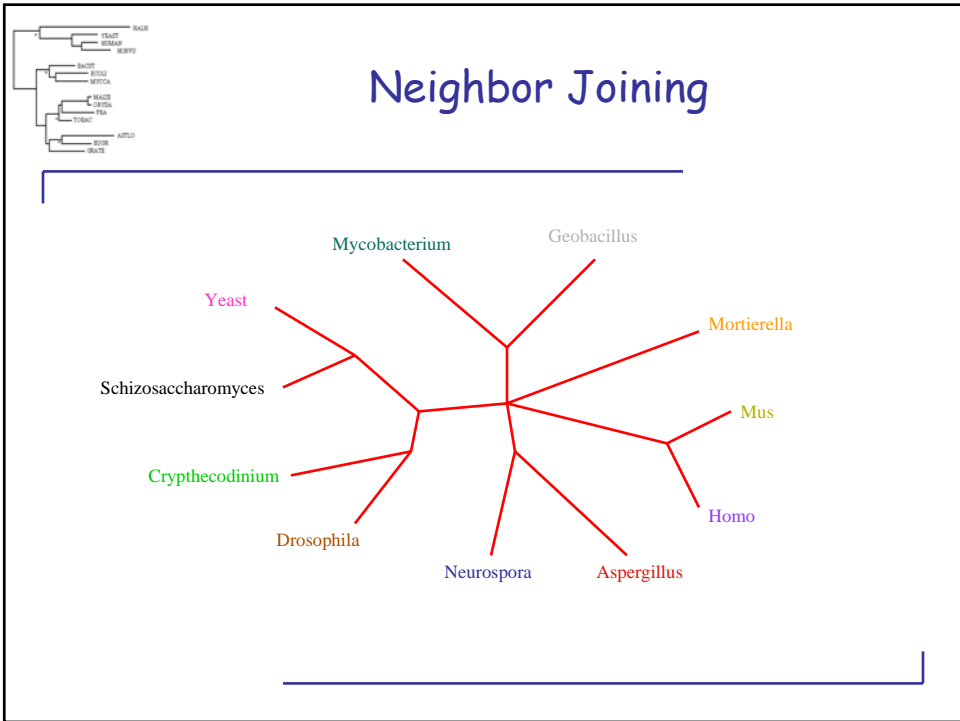
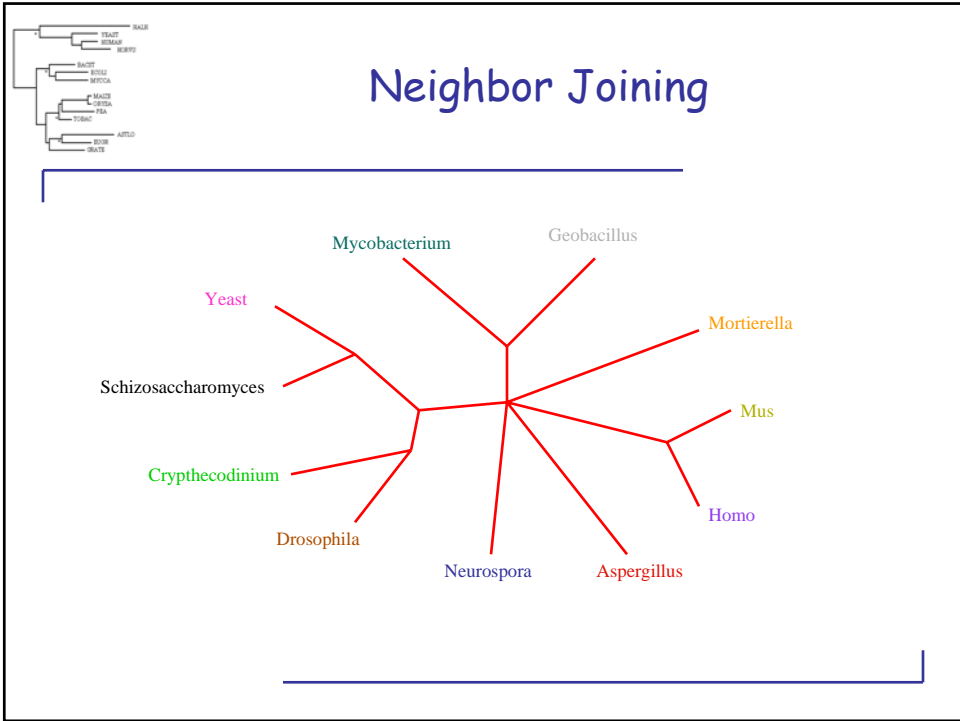


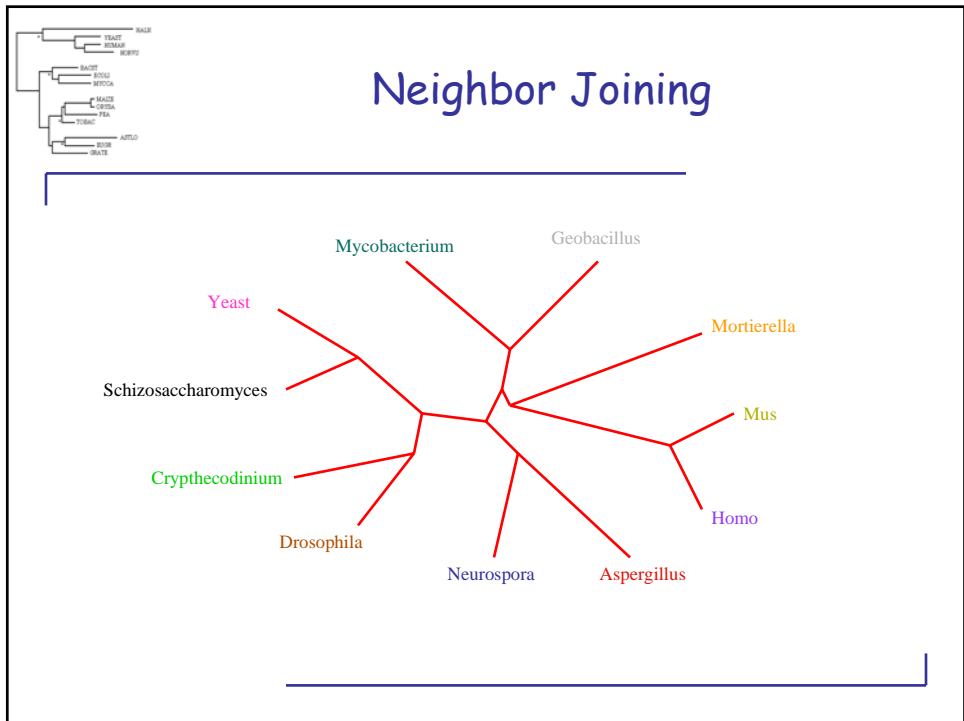
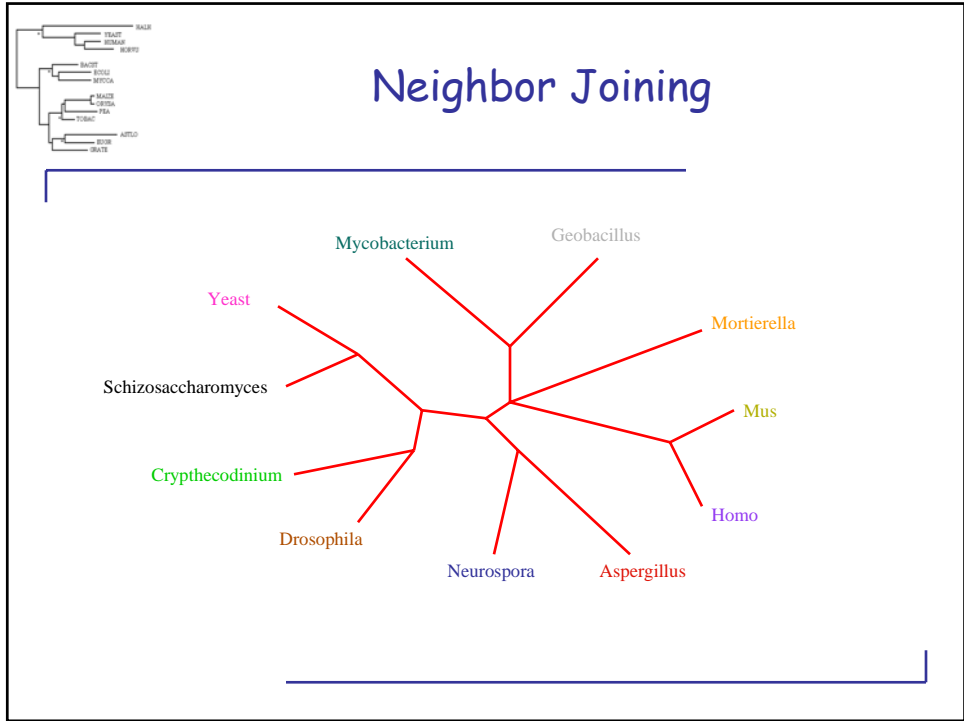
The diagram shows a starlike phylogenetic tree where all species are connected to a single central node. The species and their approximate positions are:

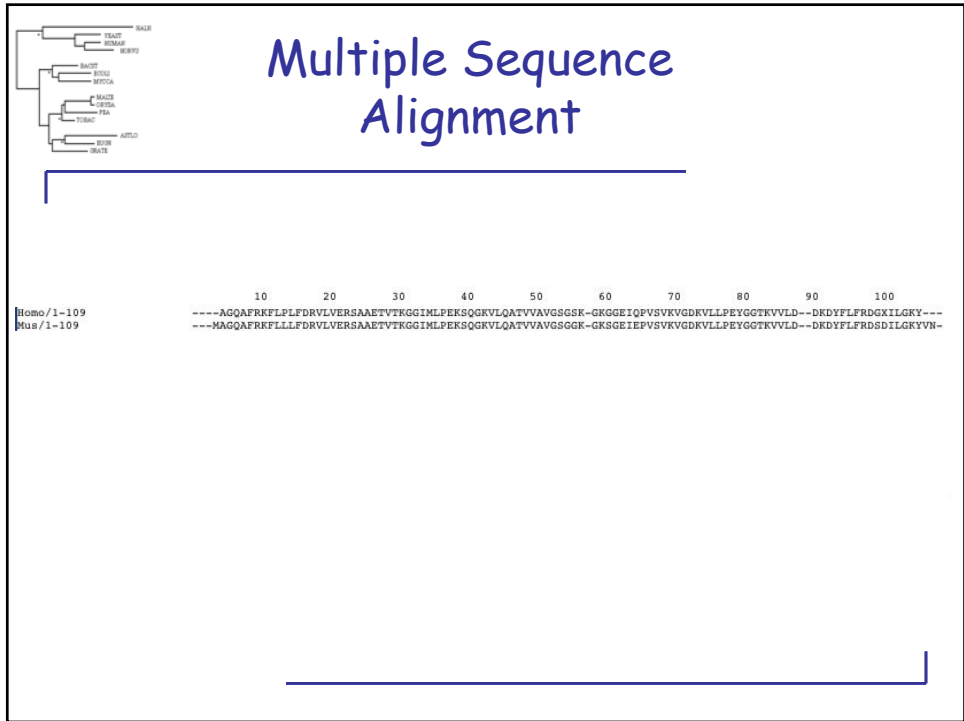
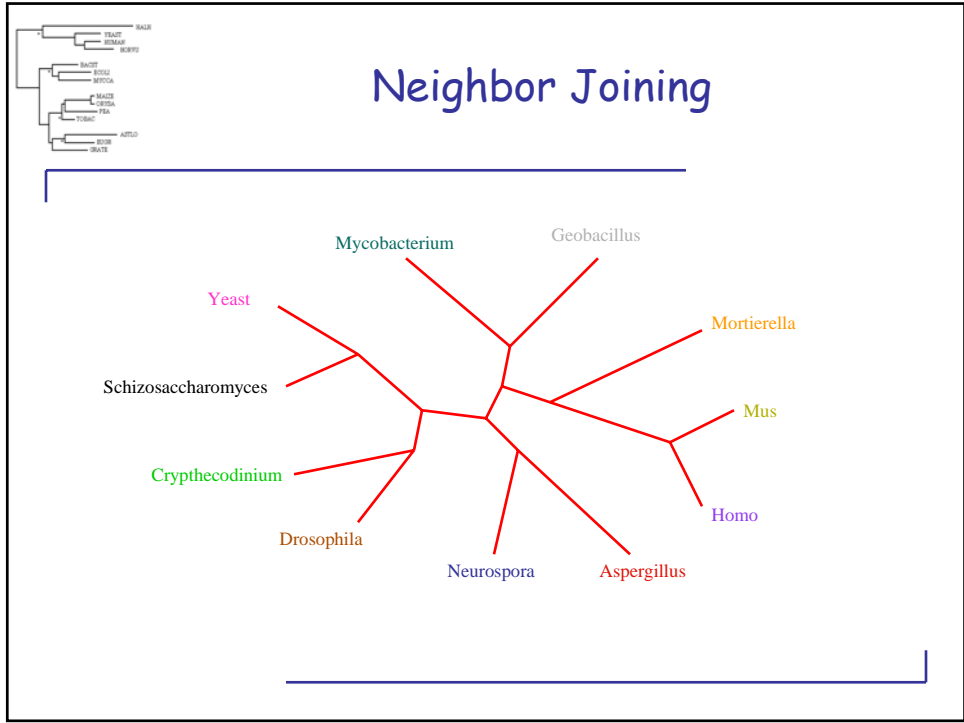
- Mycobacterium (top left)
- Geobacillus (top right)
- Mortierella (right)
- Mus (right)
- Homo (right)
- Aspergillus (bottom right)
- Neurospora (bottom center)
- Drosophila (bottom left)
- Cryptocodium (left)
- Schizosaccharomyces (left)
- Yeast (left)

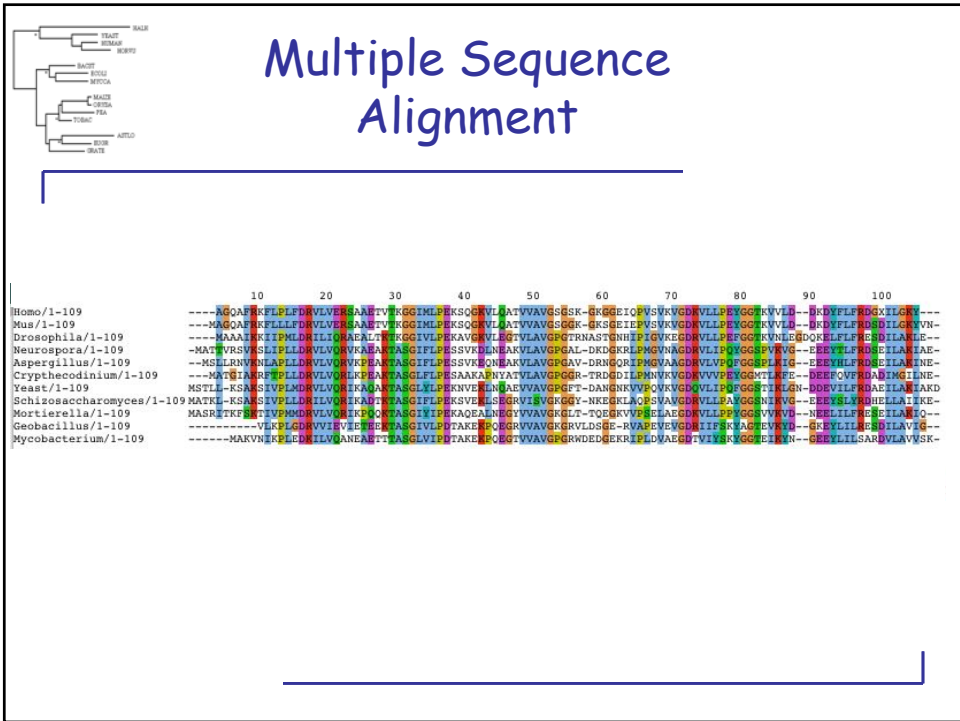
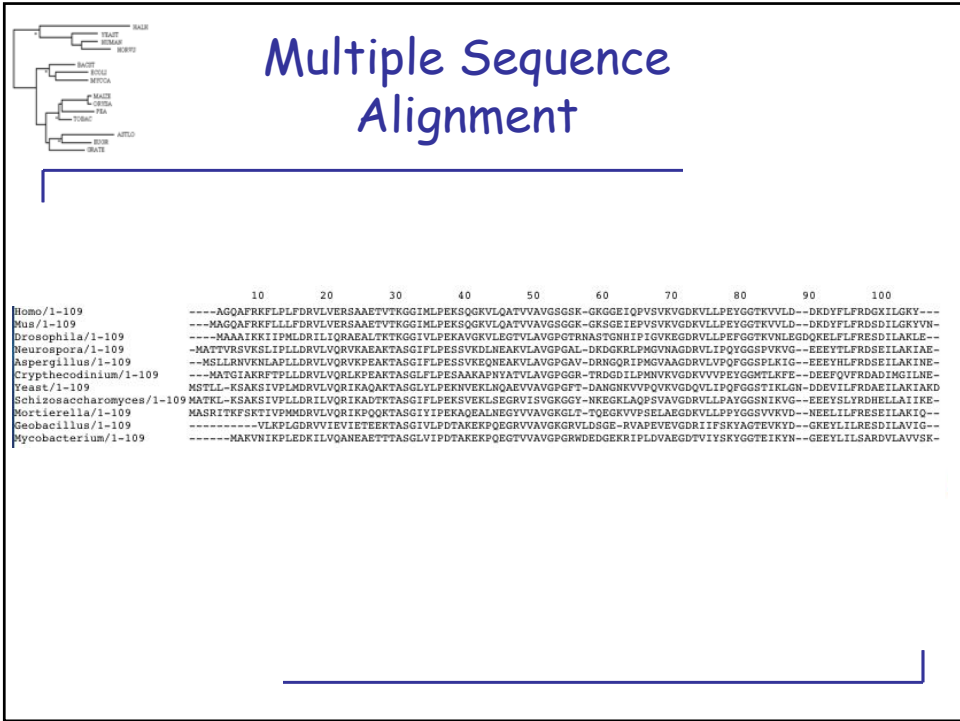












What can phylogeny do for you?

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

OR

How do we benefit from phylogeny?

AND

How is bioinformatics related to any of this?

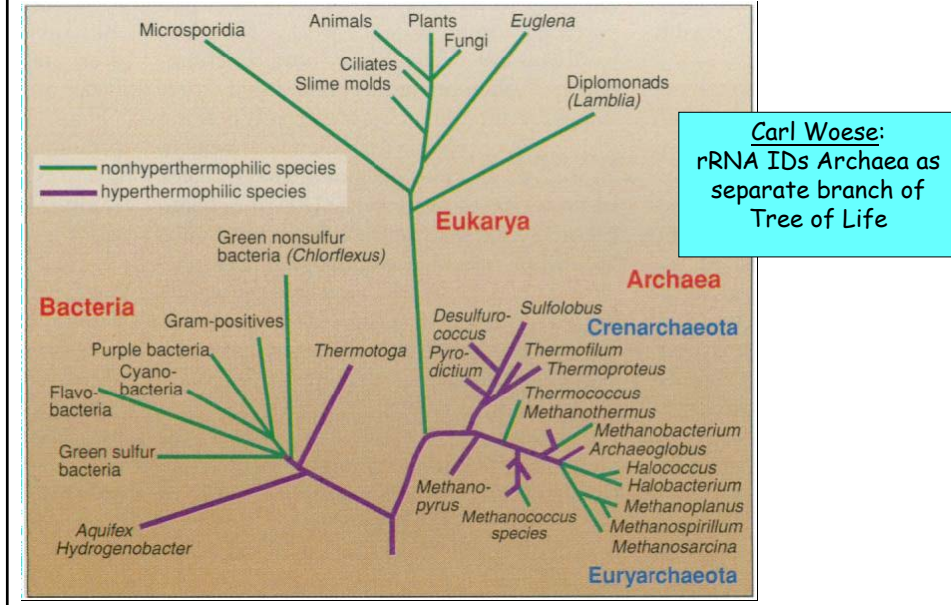
What are the goals of phylogeny?

All life forms share a common origin and are part of the Tree of Life

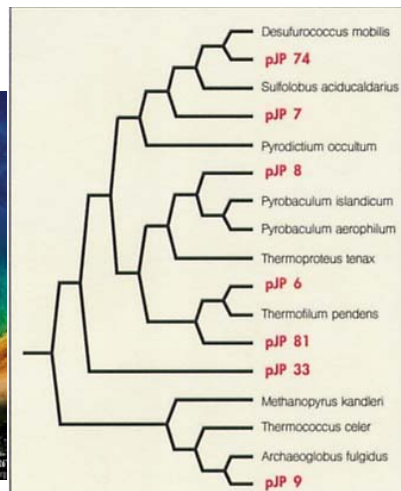
- 1) Deduce correct trees of life for all species
- 2) Infer or estimate divergence times

How can we use phylogenetic analyses?

Revolutionalizing the Tree of Life

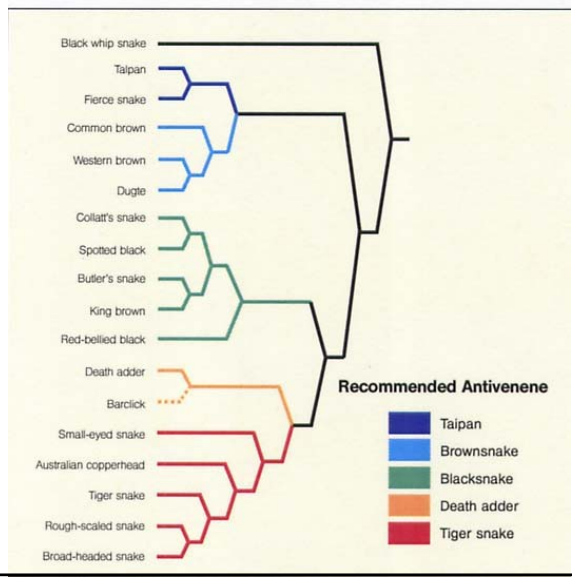


Discovering new life forms



A phylogeny of some archaeobacteria. Newly discovered life forms are in red.

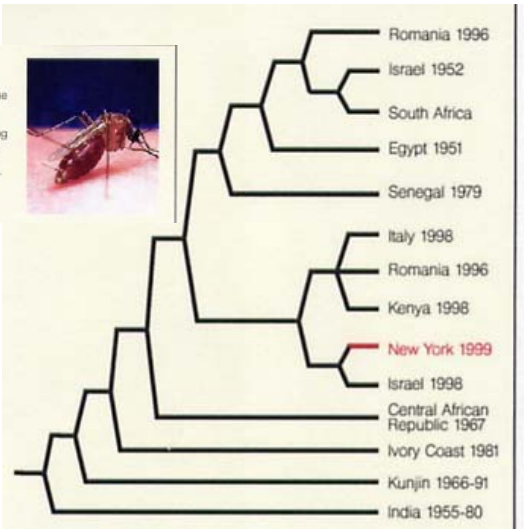
Developing effective snakebite antivenins



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

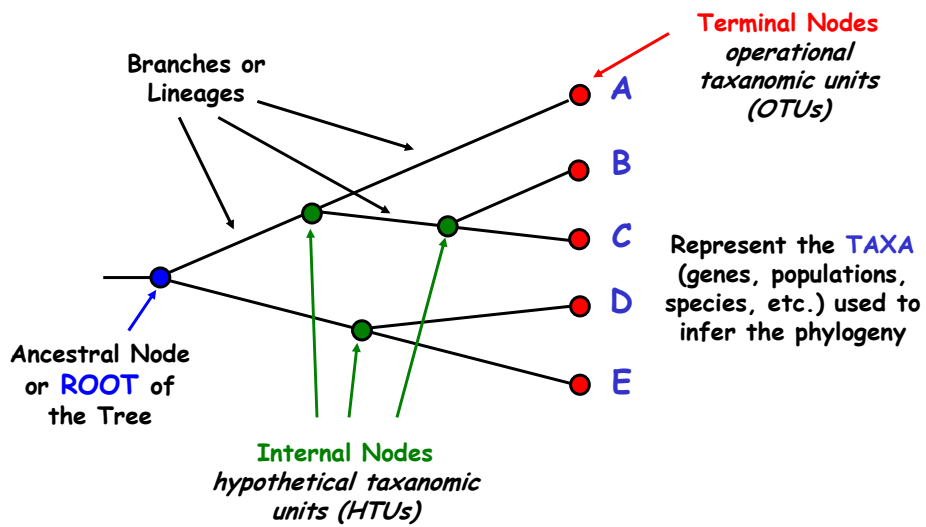


Eurasian water milfoil

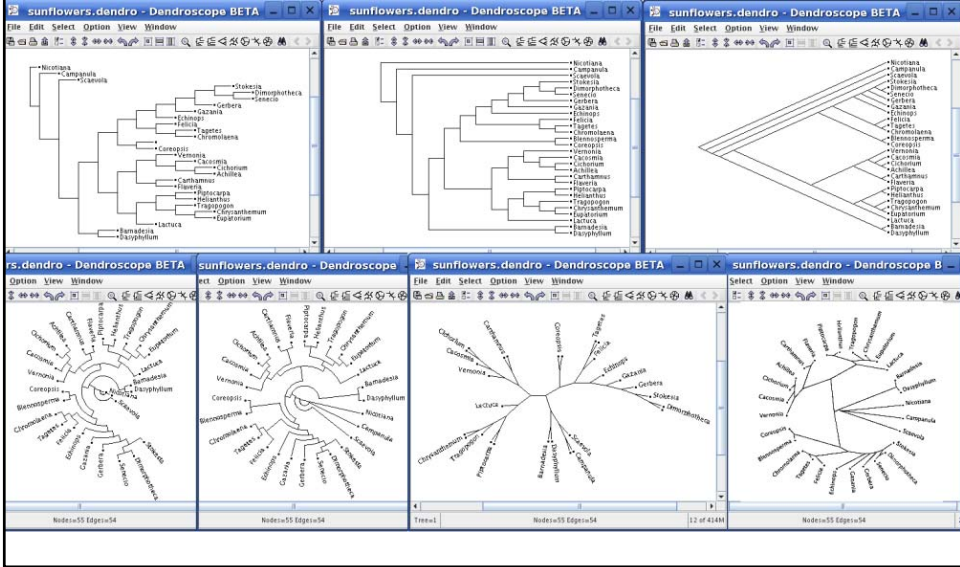


Caulerpa taxifolia

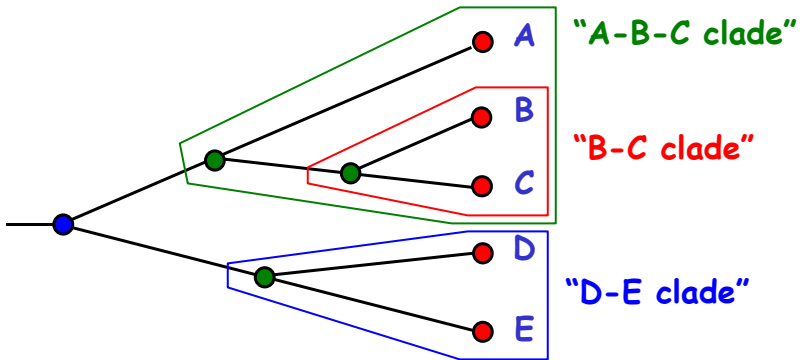
Common phylogenetic tree terminology



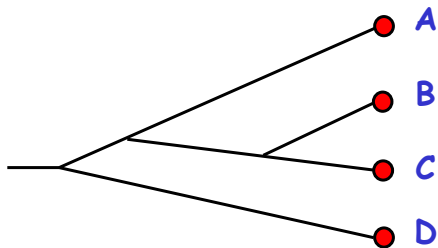
Phylogenetic trees can be drawn many ways



Clade: group with a single common ancestor and its descendants

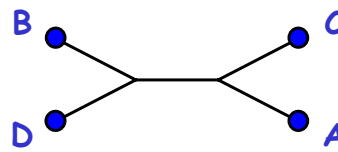


Phylogenetic trees can be rooted or unrooted



Rooted

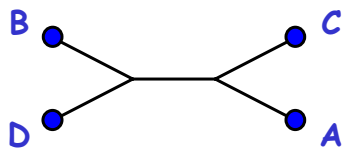
Specifies evolutionary path
Root node is most recent
common ancestor of all TUs;
specifies time flow



Unrooted

Shows degree of kinship
Doesn't make assumptions
or require knowledge of
common ancestor

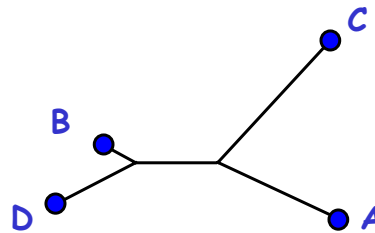
Phylogenetic trees can be scaled or unscaled



Unscaled

Branch length not
proportional to number of
changes/distance

Cladogram

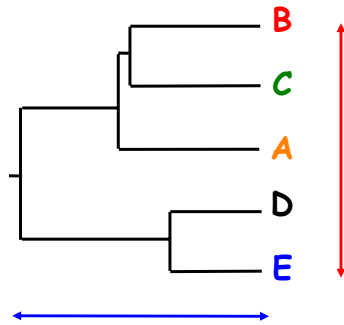


Scaled

Branch length
proportional to number of
changes/distance

Phylogram

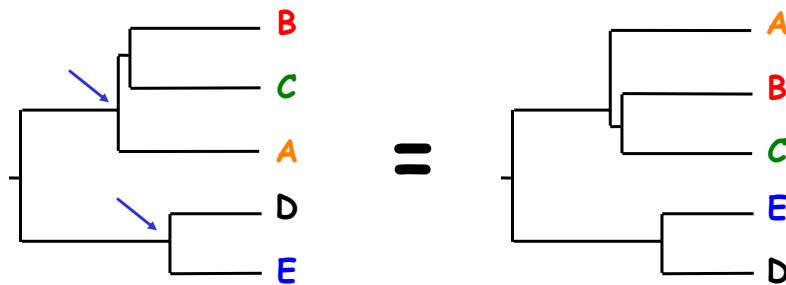
Phylogenetic trees diagram evolutionary relationships



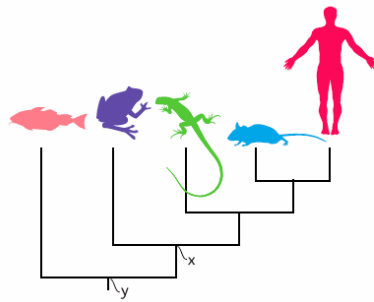
No meaning to the spacing between the taxa, or to the order in which they appear from top to bottom.

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

Rotating clades: same meanings



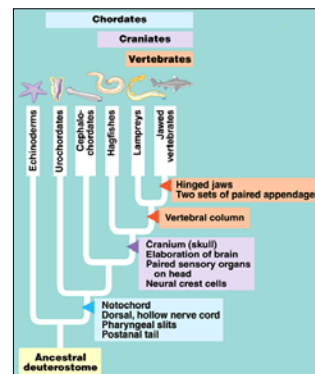
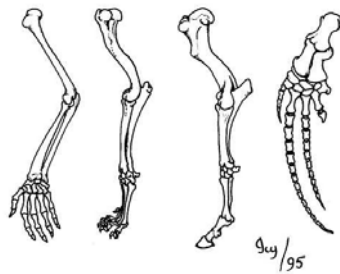
Interpreting phylogenetic trees



Is the frog more closely related to the fish or the human ?

How are phylogenetic trees built?

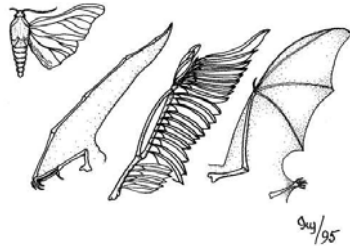
Traditionally: use homologous structures



Caveats:

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

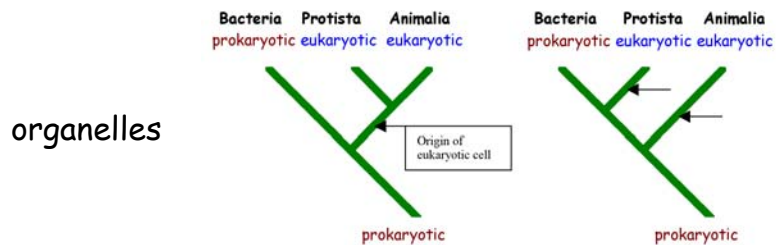
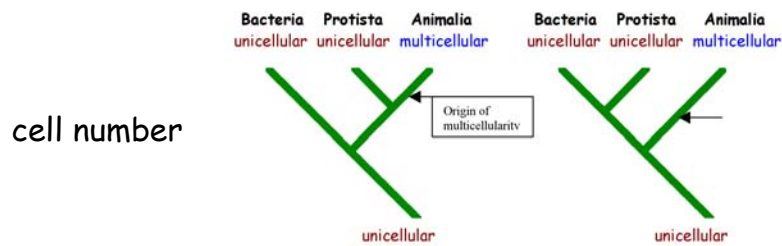
Structural analogy can result from convergent evolution



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 closeness or evolutionary lineage (phylogenetic relationship). All the cichlids of Lake Malawi are more closely related to one another than to any cichlids in Lake Tanganyika.

68 SCIENTIFIC AMERICAN February 1999

Classification based on traits can be tricky



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:

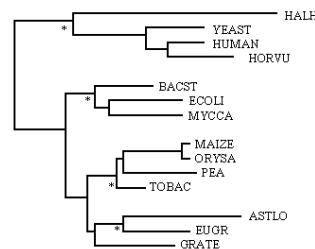
Gene divergence may not correlate with species divergence

Molecular phylogenies can be constructed using different elements

Nuclear genes

Mitochondrial DNA

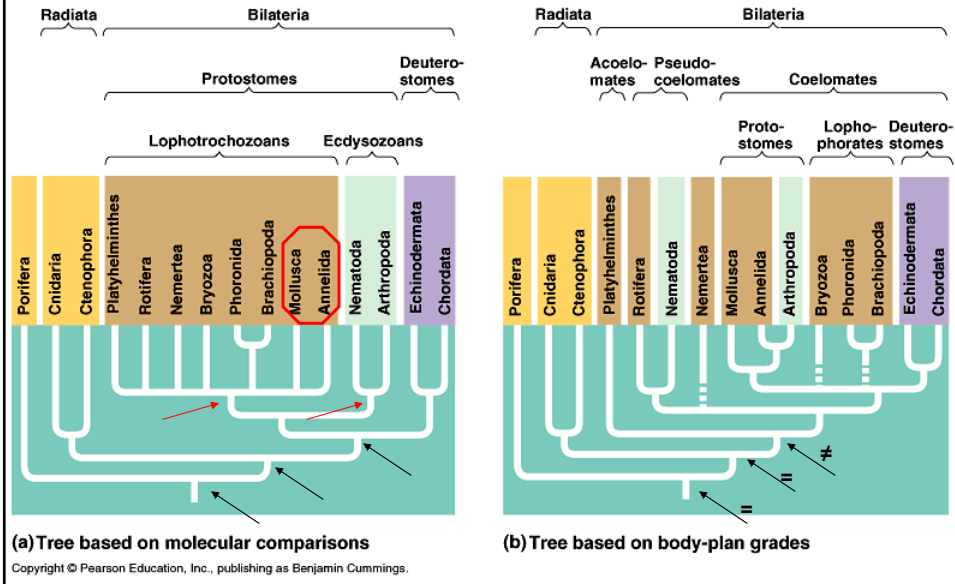
Genome structure



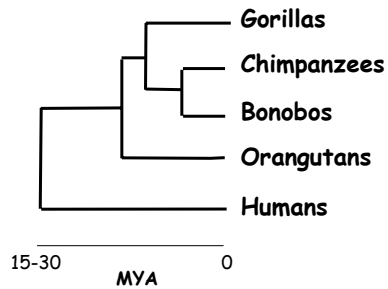
Reasonably well conserved, present in common ancestors

Usually integrate analyses of multiple different genes

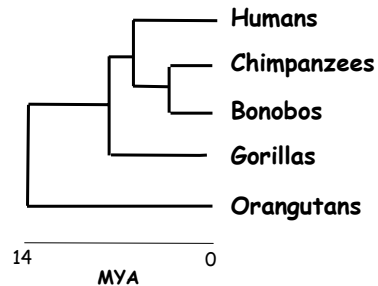
Molecular comparisons vs. body plans



Which species are the closest living relatives of modern humans?

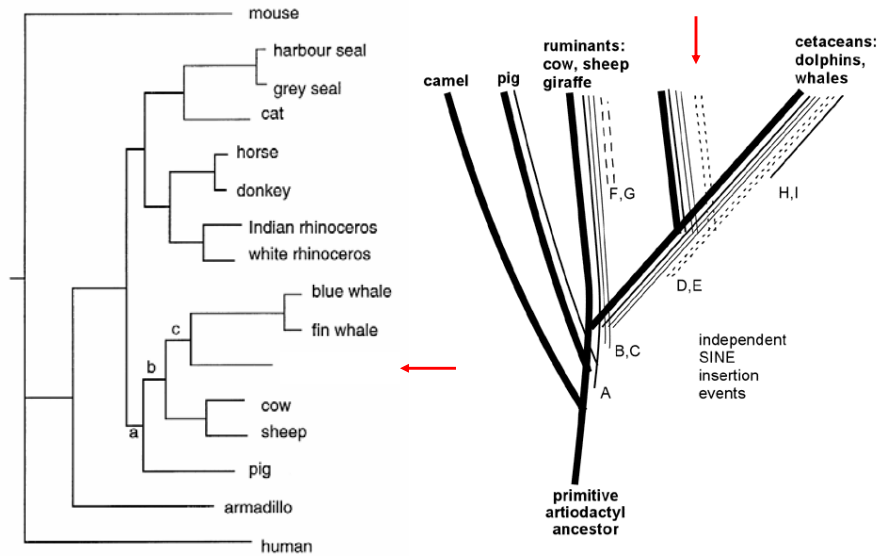


Great apes (chimpanzees, gorillas and orangutans) formed a clade separate from humans.



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

What is the closest living relative of whales?



Phylogenetic trees are hypotheses

How do you construct phylogenetic trees?

What computational strategies are used?

How do you test the robustness of hypotheses?

