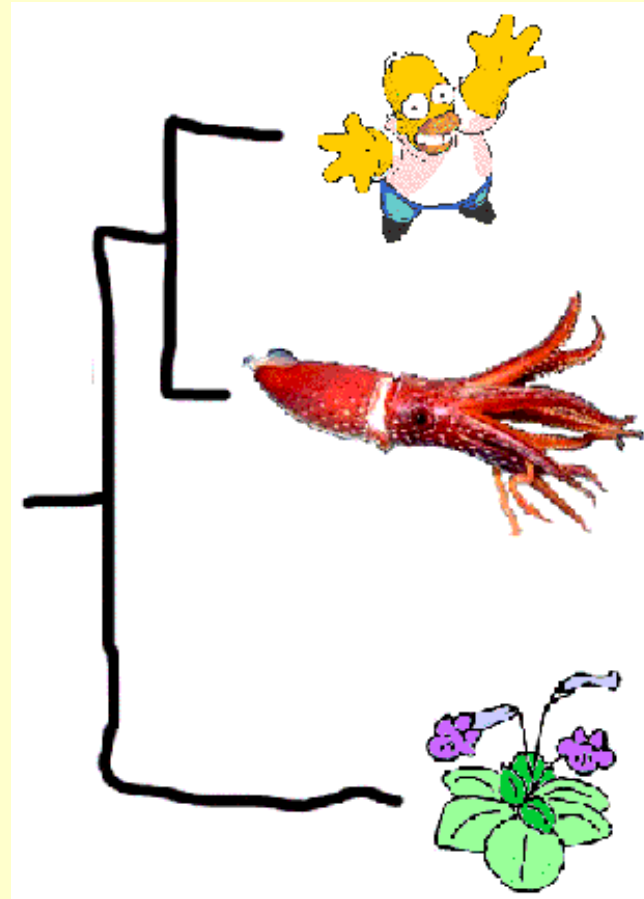


# Computational Molecular Biology

## Biochem 218 – BioMedical Informatics 231

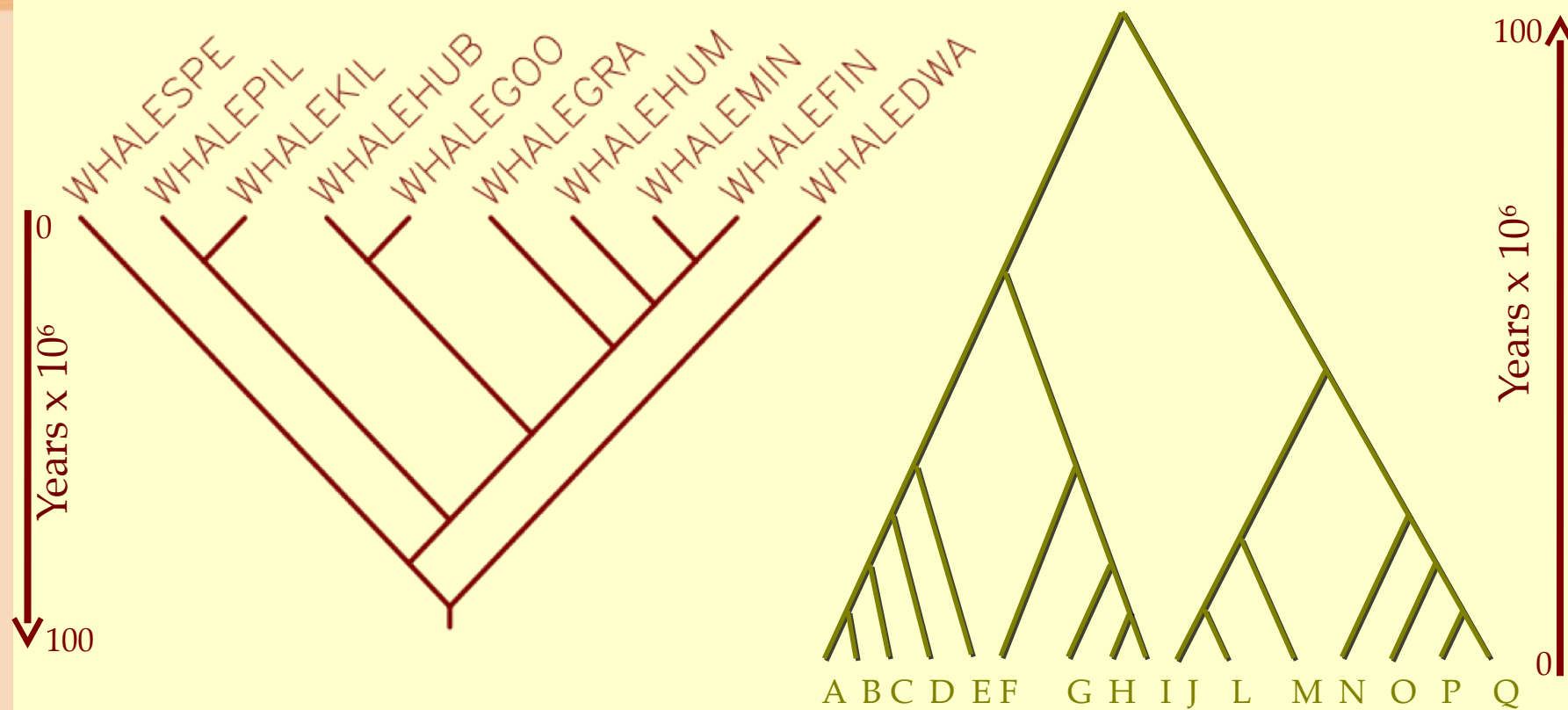
<http://biochem218.stanford.edu/>

### Phylogenies

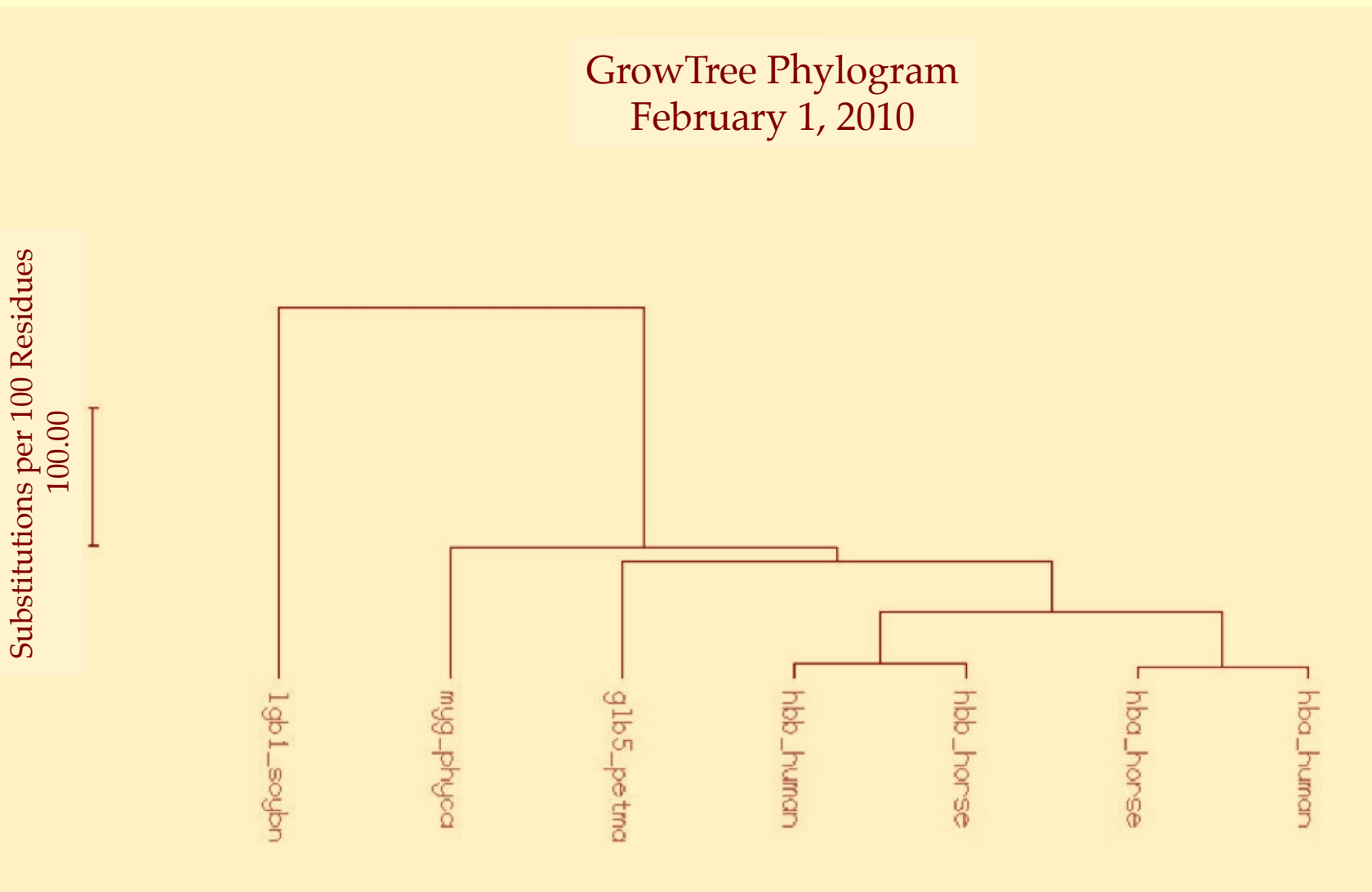


Doug Brutlag  
Professor Emeritus  
Biochemistry & Medicine (by courtesy)

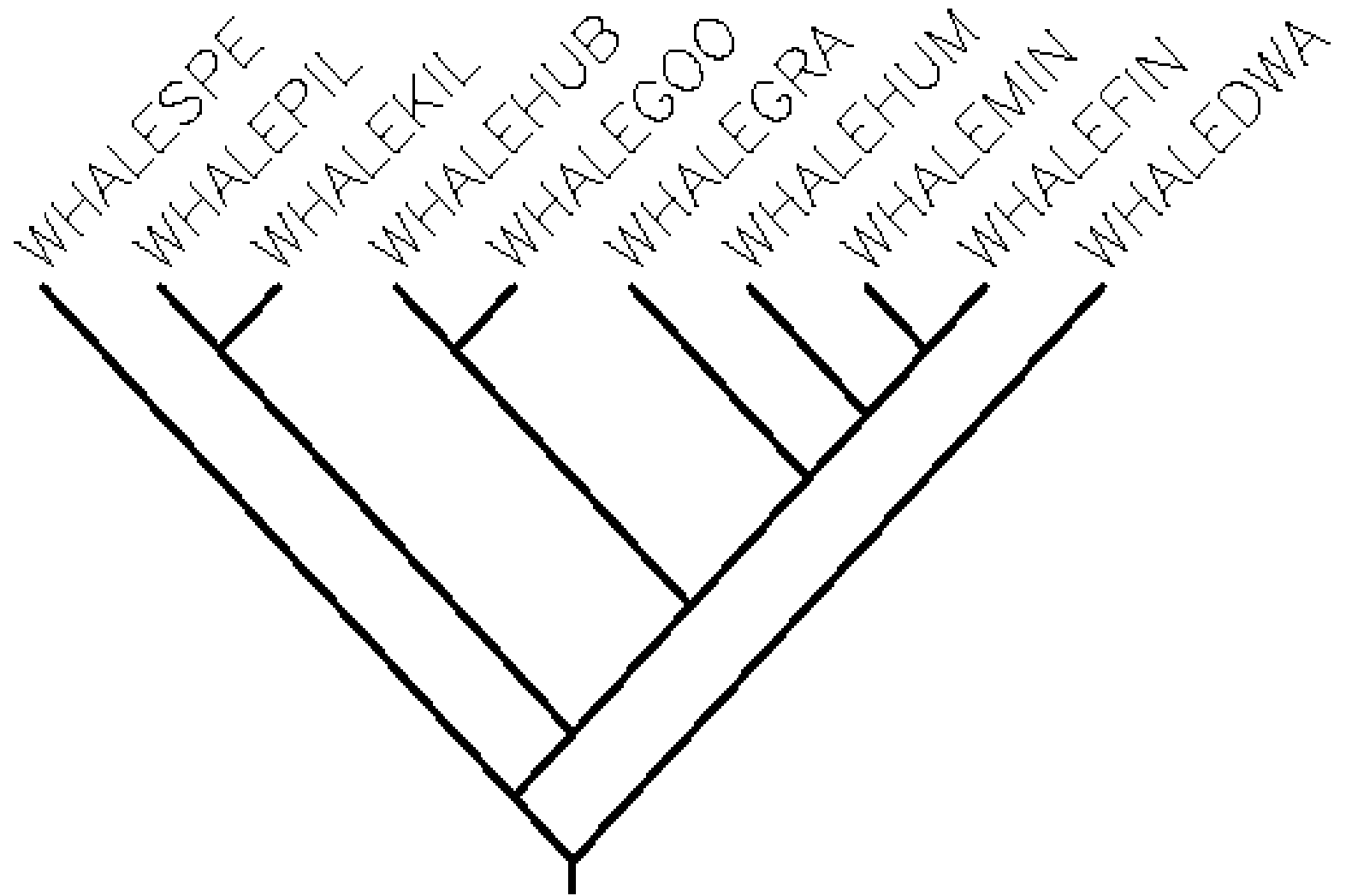
# Cladogram Representation of Phylogenies



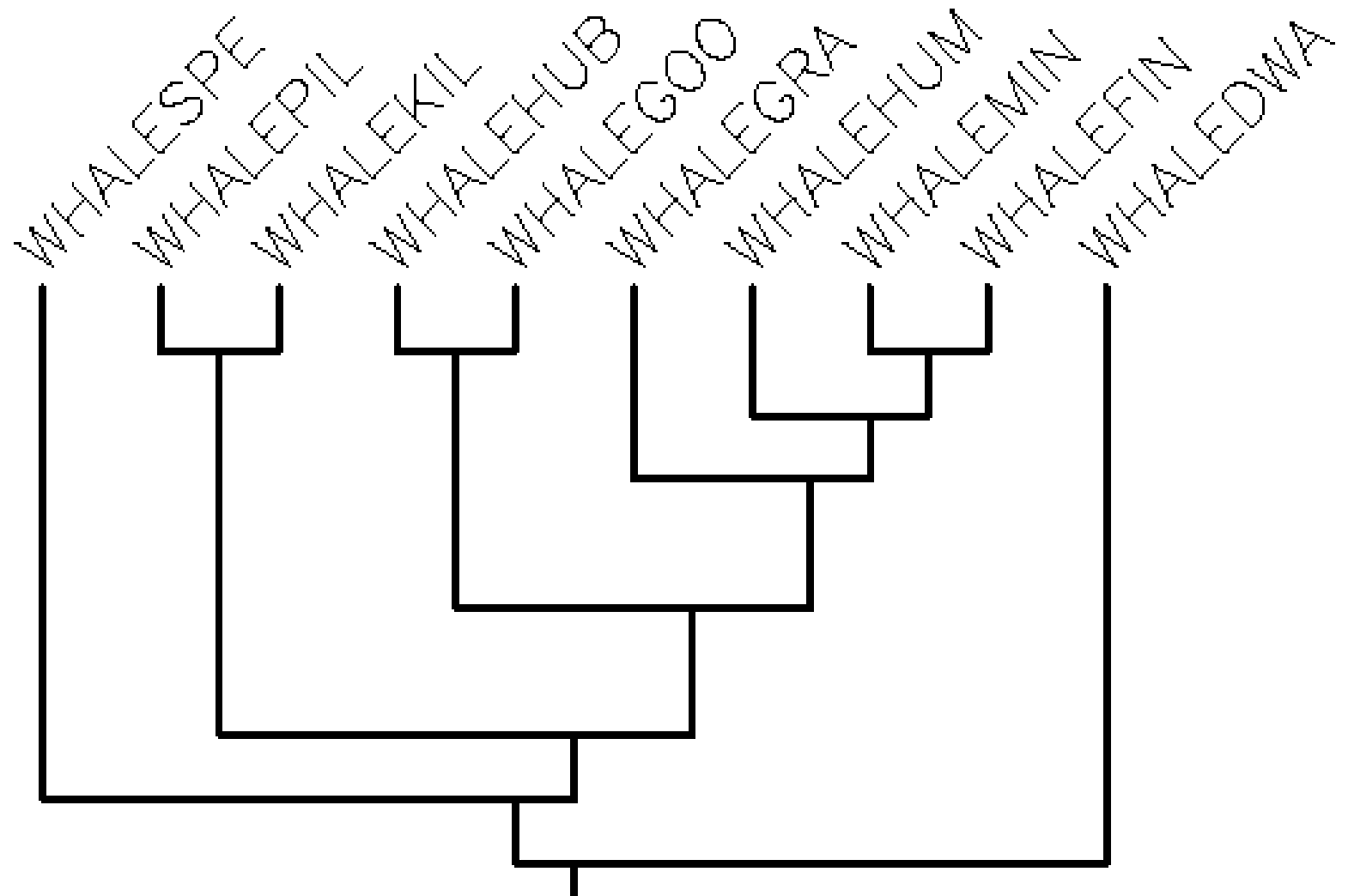
# Dendrogram Representation of Phylogenies



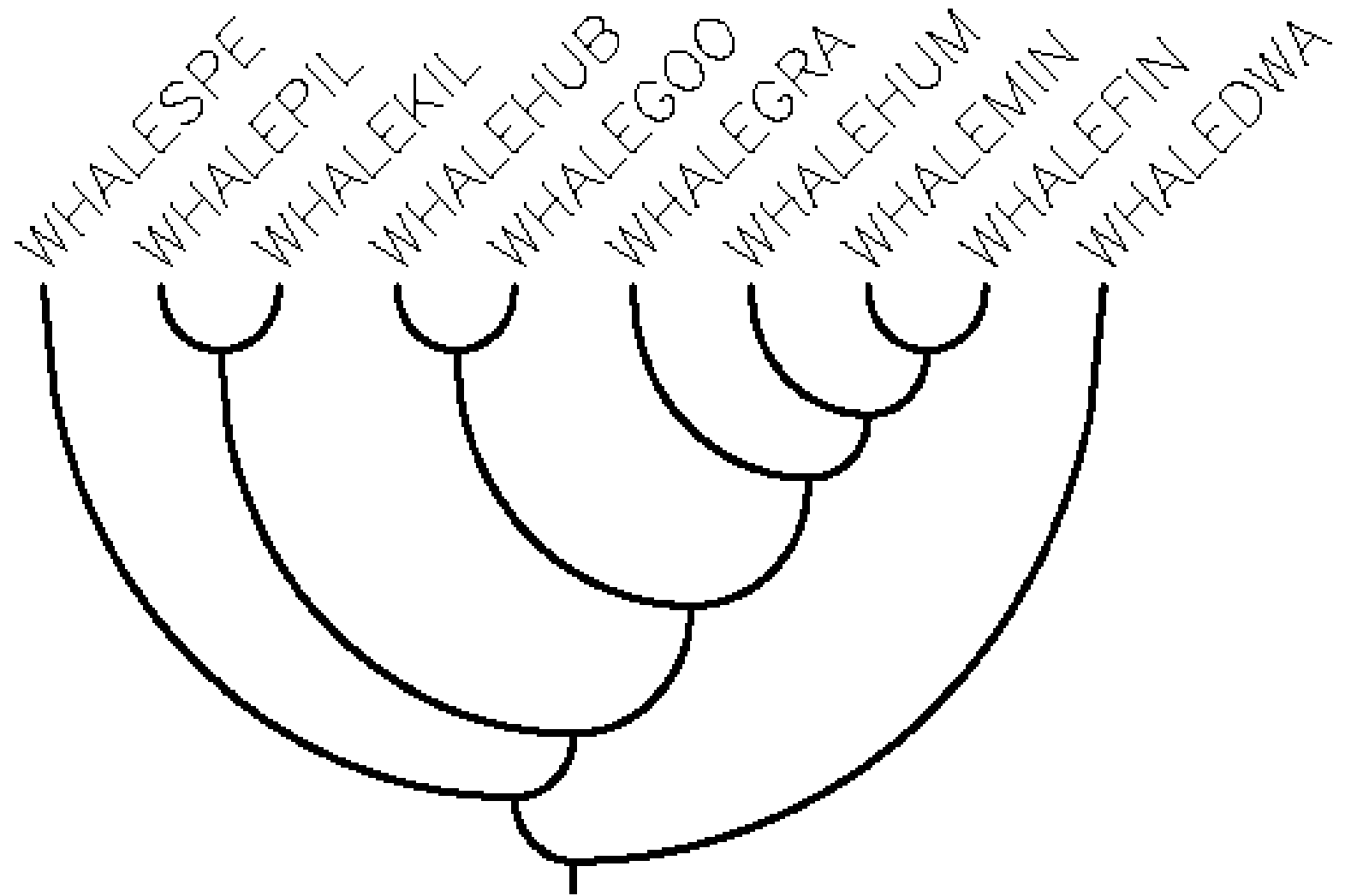
# Cladogram



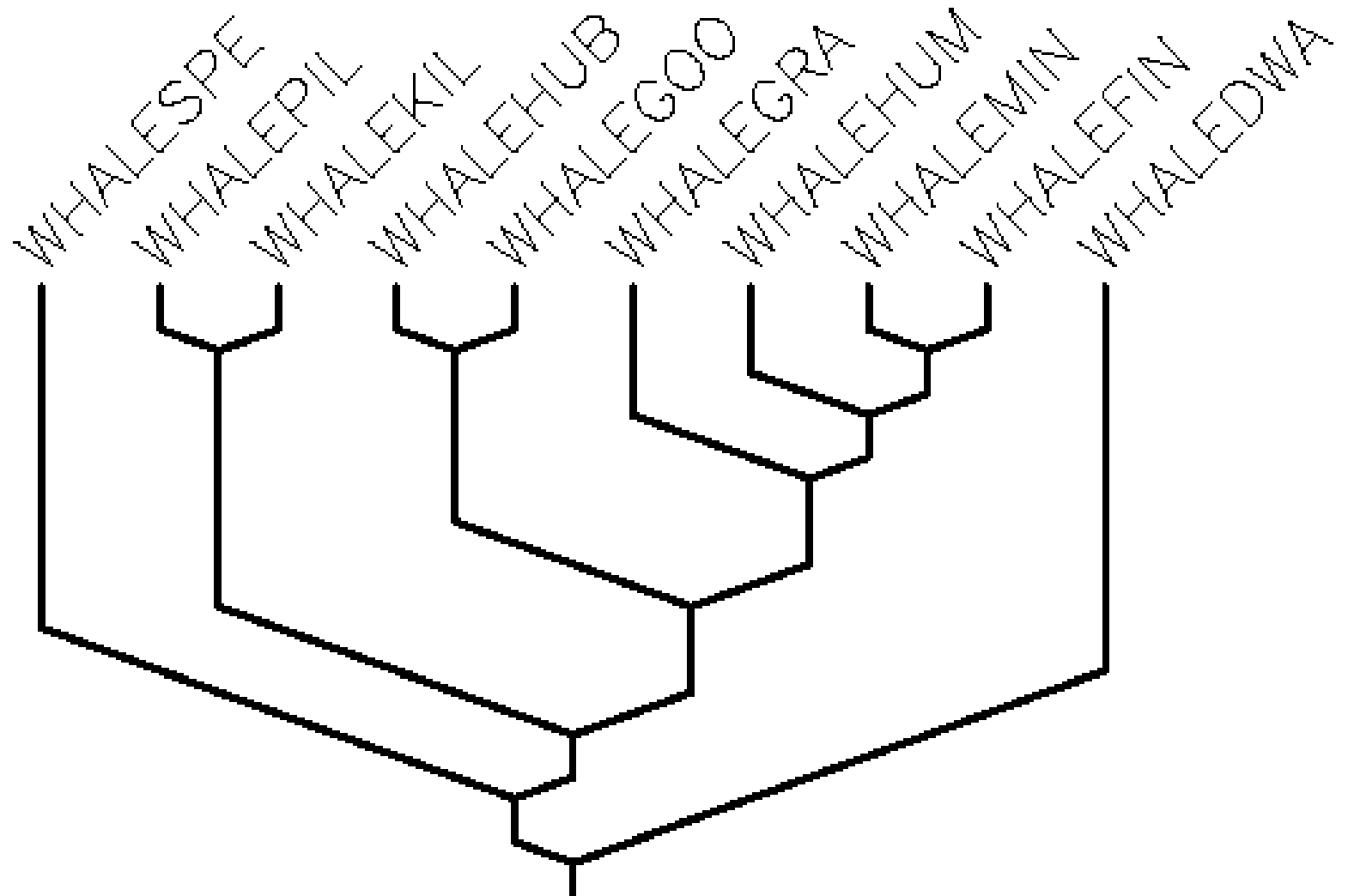
# Phenogram



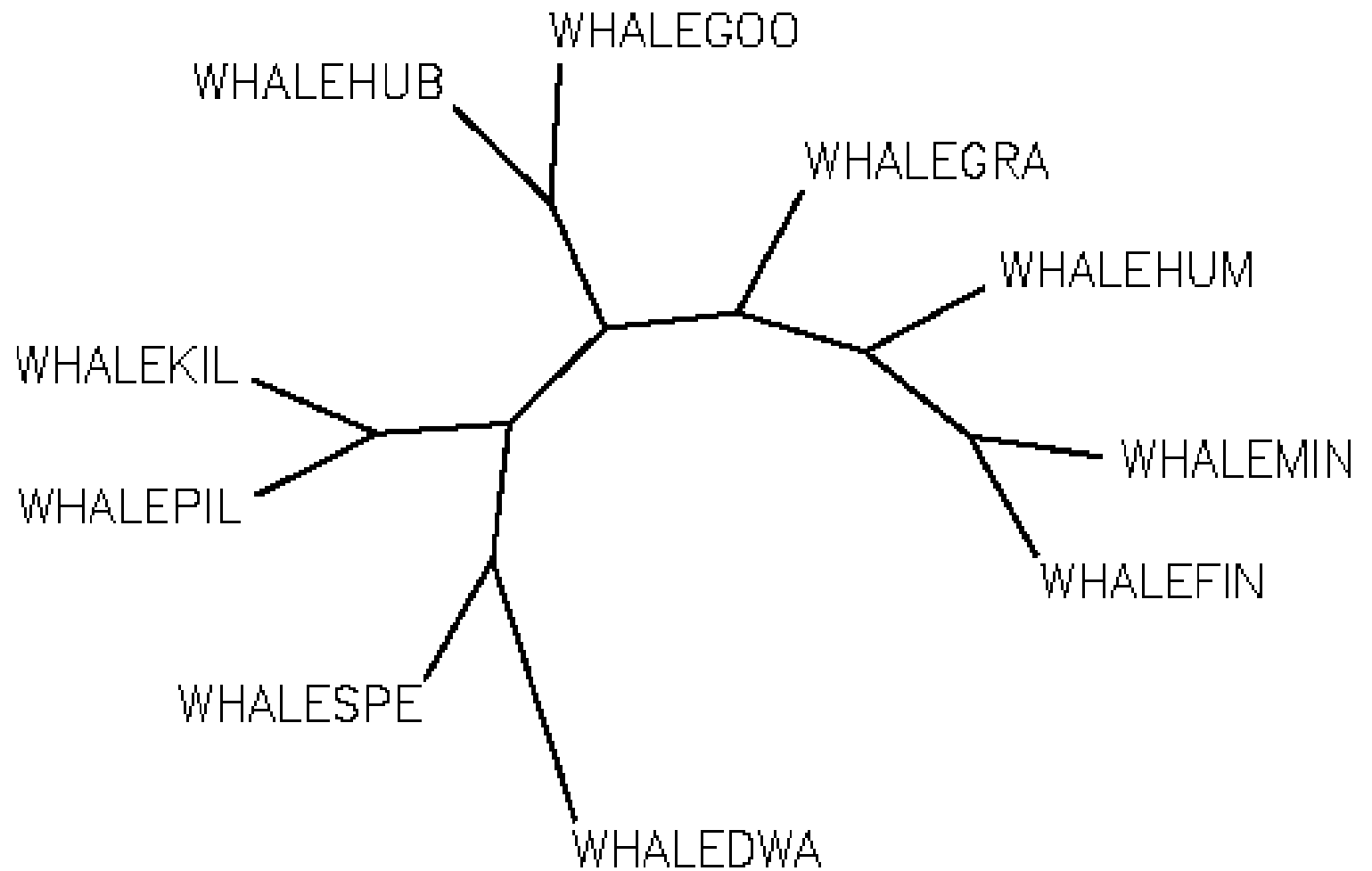
# Curve-O-Gram



# Eurogram



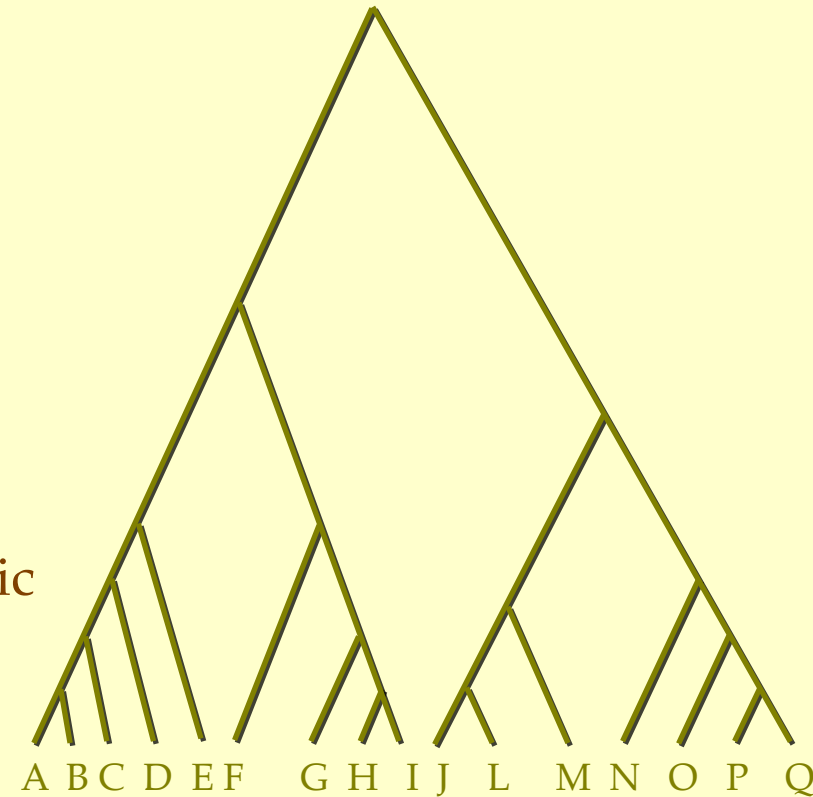
# RadialGram





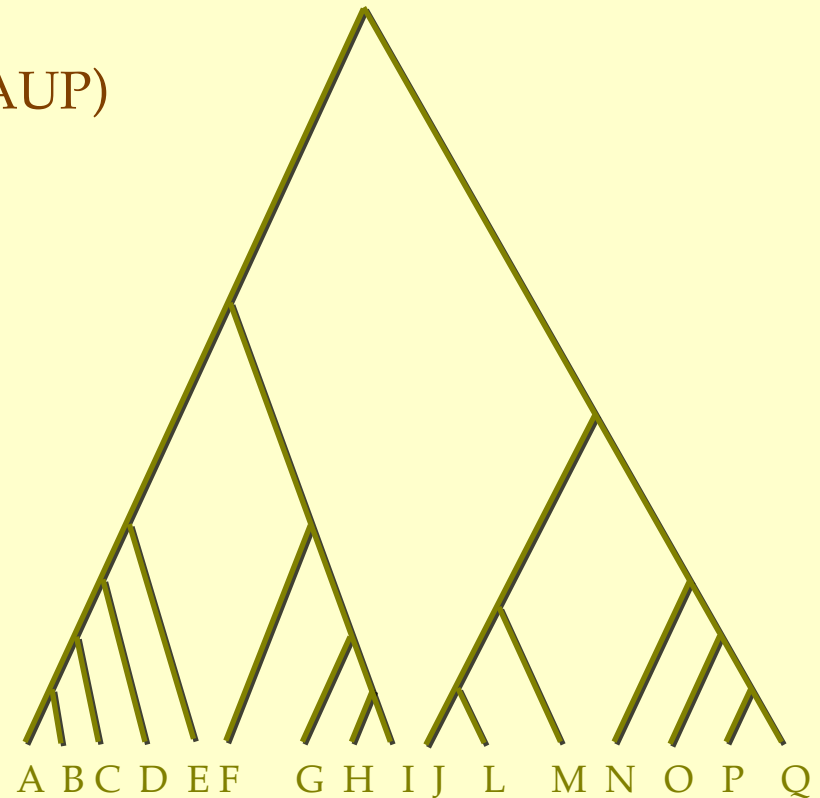
# Methods for Determining Phylogenies

- Parsimony (character based)
  - Assigns mutations to branches
  - Minimize number of edits
  - Topology maximizes similarity of neighboring leaves
- Distance methods
  - Branch lengths =  $D(i,j) / 2$  for sequences  $i, j$
  - Distances must be at least metric
  - Distances can reflect time or edits
  - Distance must be relatively constant per unit branch length



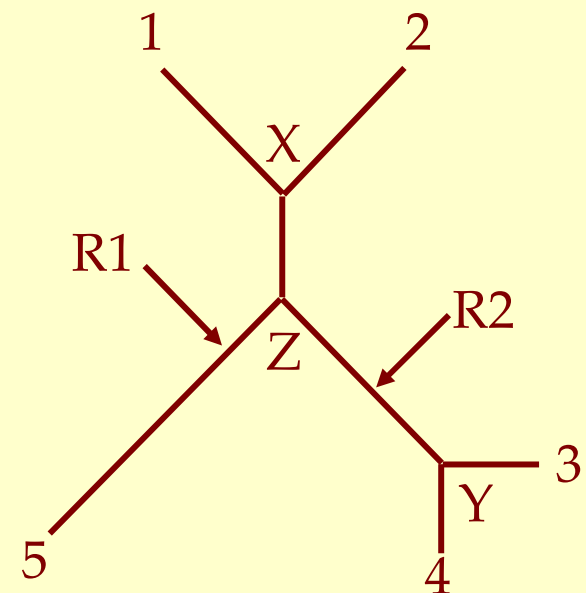
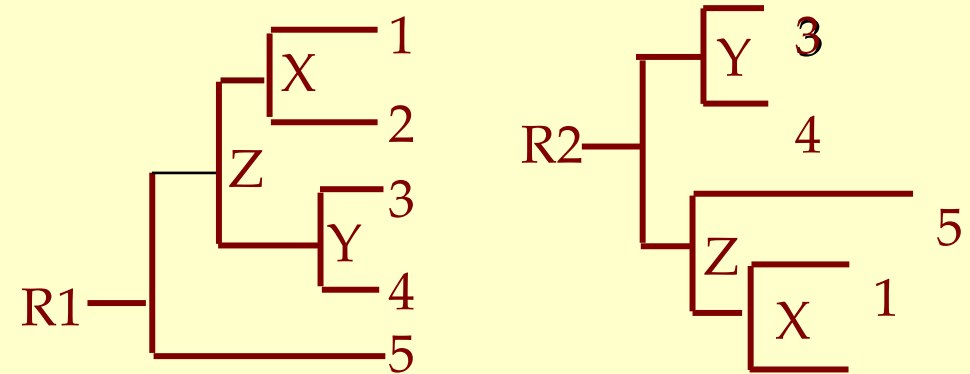
# Methods for Determining Phylogenies

- Parsimony
  - Minimum mutation (Fitch, PAUP)
  - Minimal length encoding
- Probabilistic
  - Branch and Bound
  - Maximum likelihood
- Distance methods
  - Ultrametric Trees
  - Additive Trees
  - UPGMA
  - Neighbor Joining

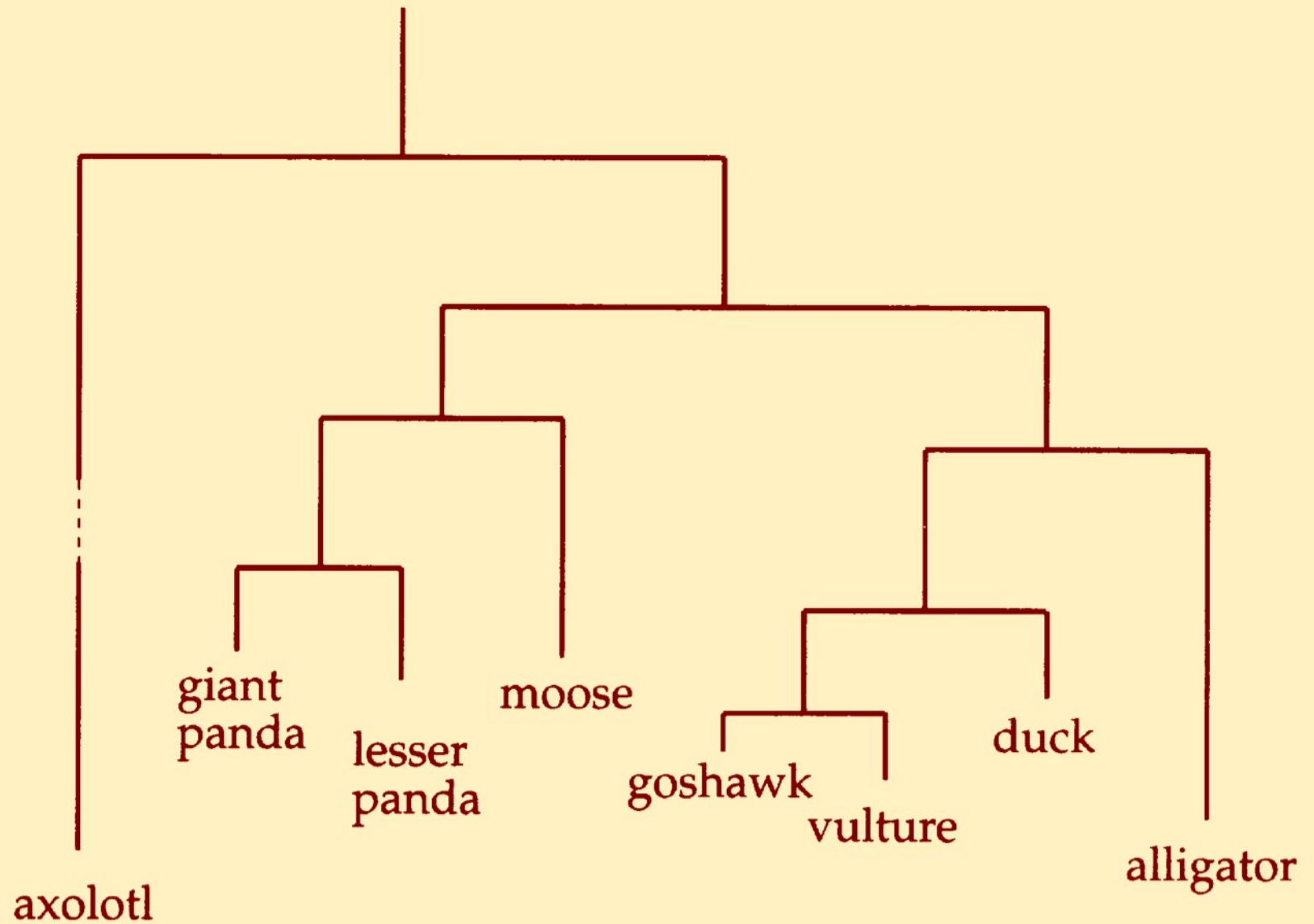


# Properties of Trees

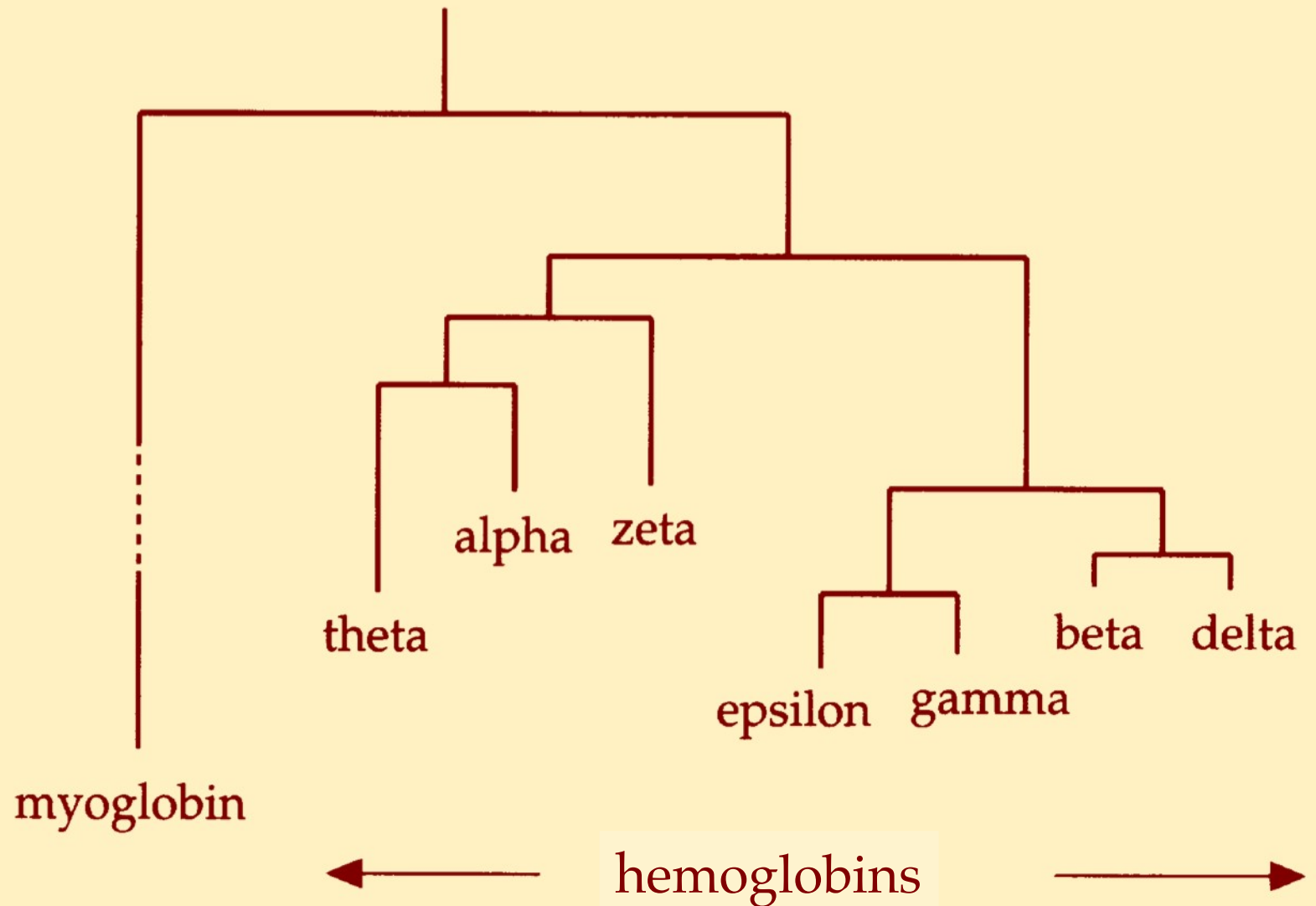
- Rooted or Unrooted
- Nodes and Branches
  - Internal Nodes
  - External Nodes - leaves
- Operational Taxonomic Units
- Outgroups
- Topology
- One path/pair
- Distances



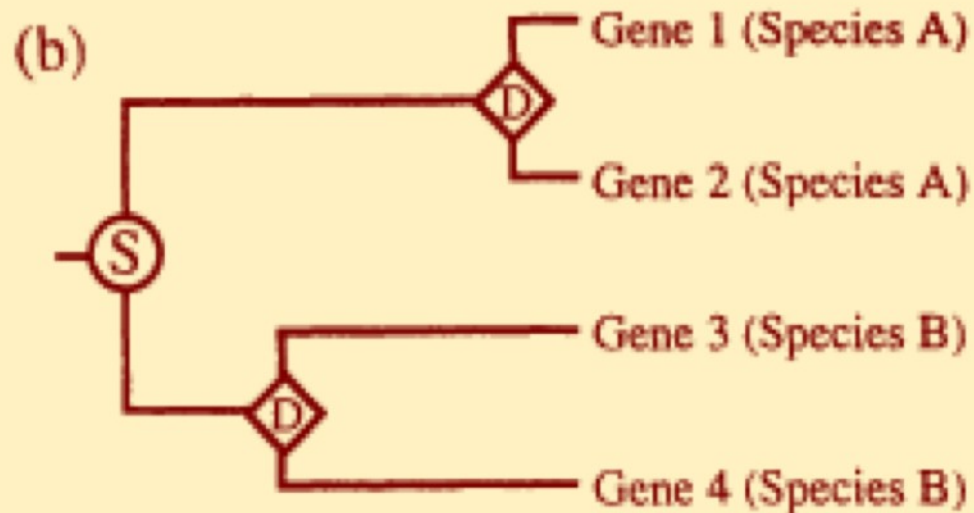
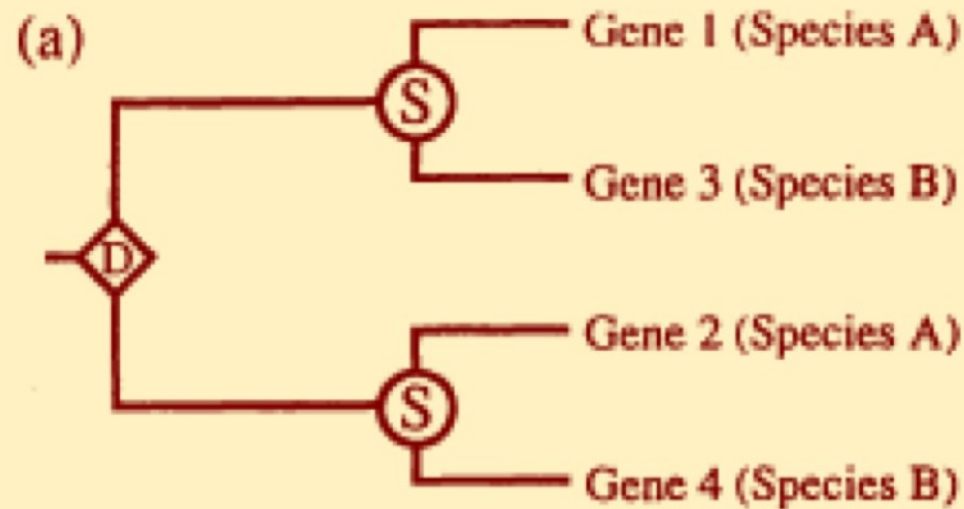
# Orthologous Evolution



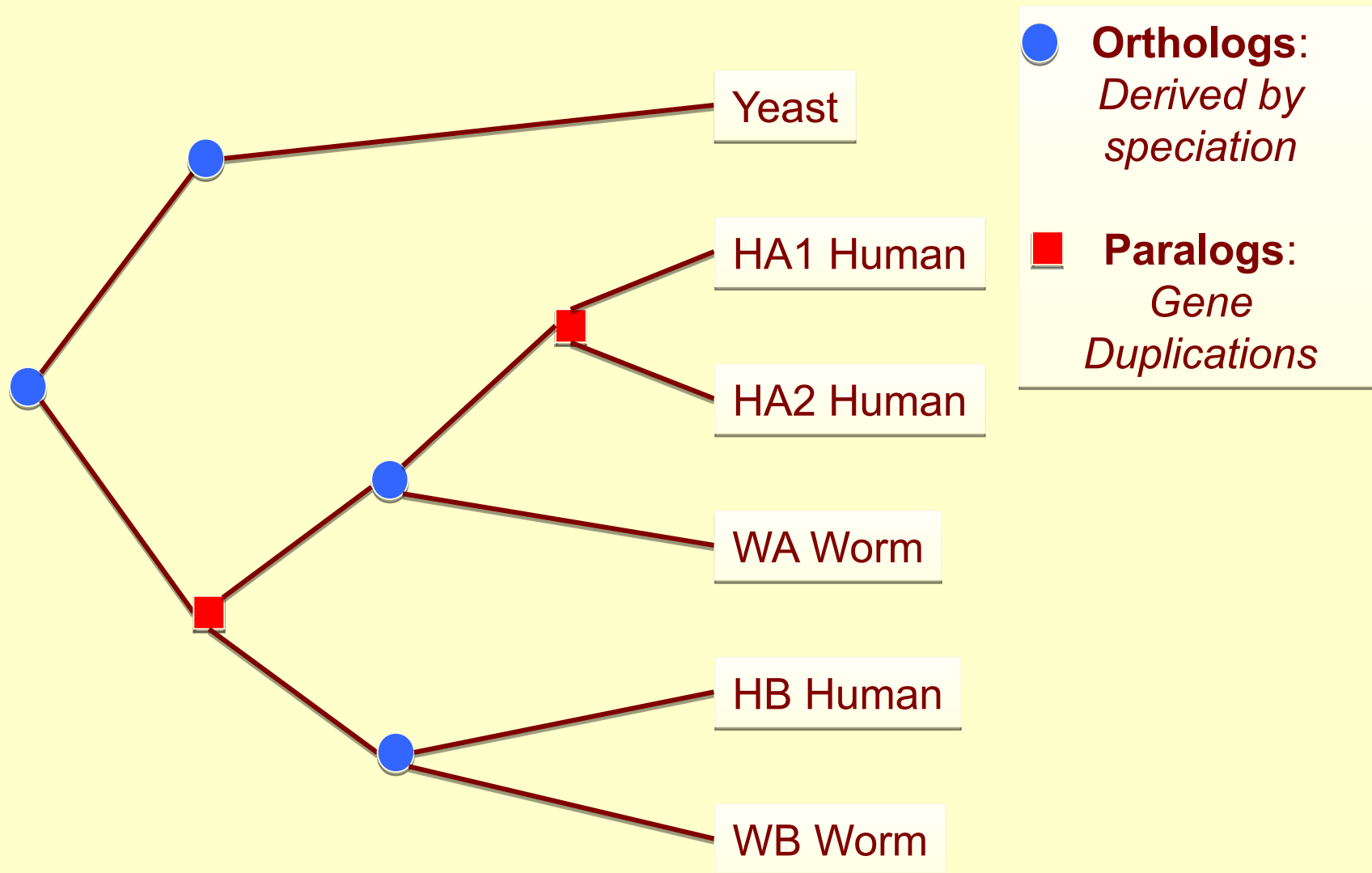
# Paralogous Evolution



# Challenges Making Trees: Gene Duplication versus Speciation



# Orthology and Paralogy



# Gene Conversion

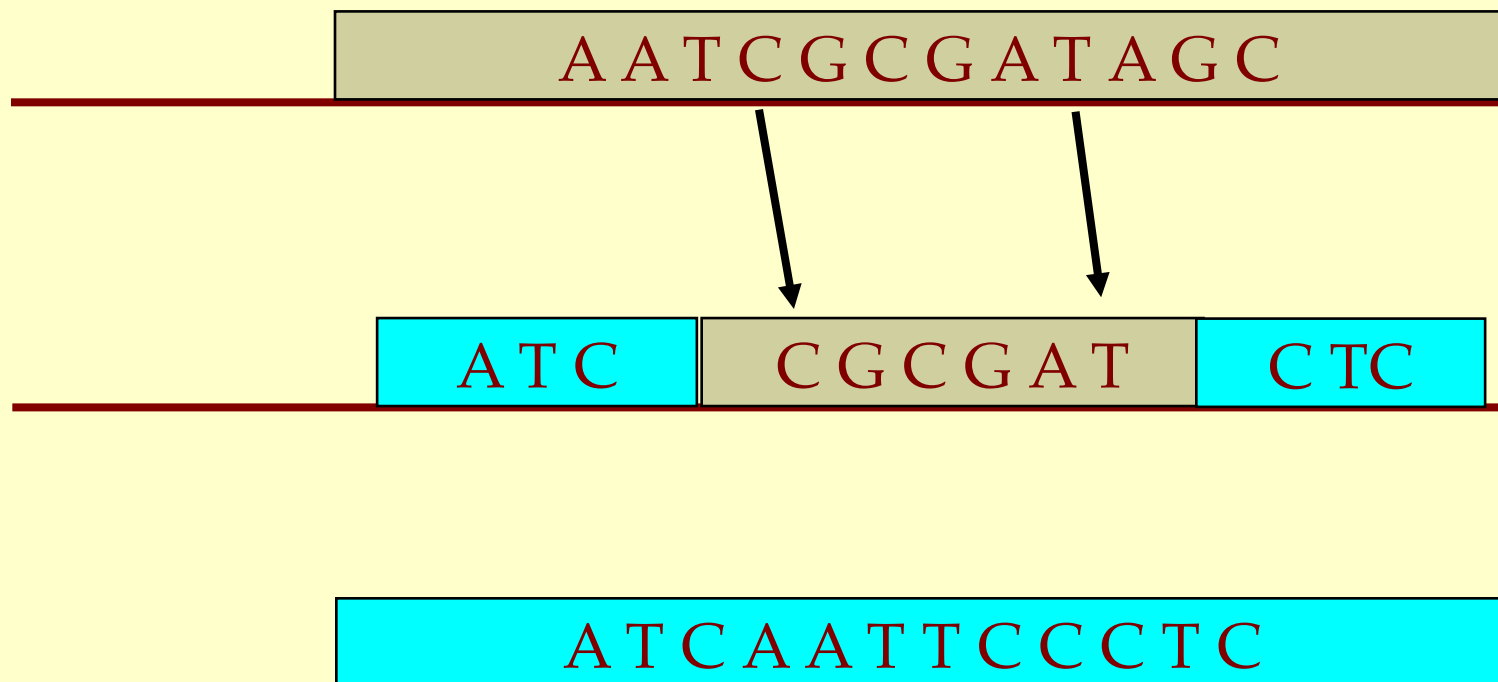
---

AATCGCGATAGC

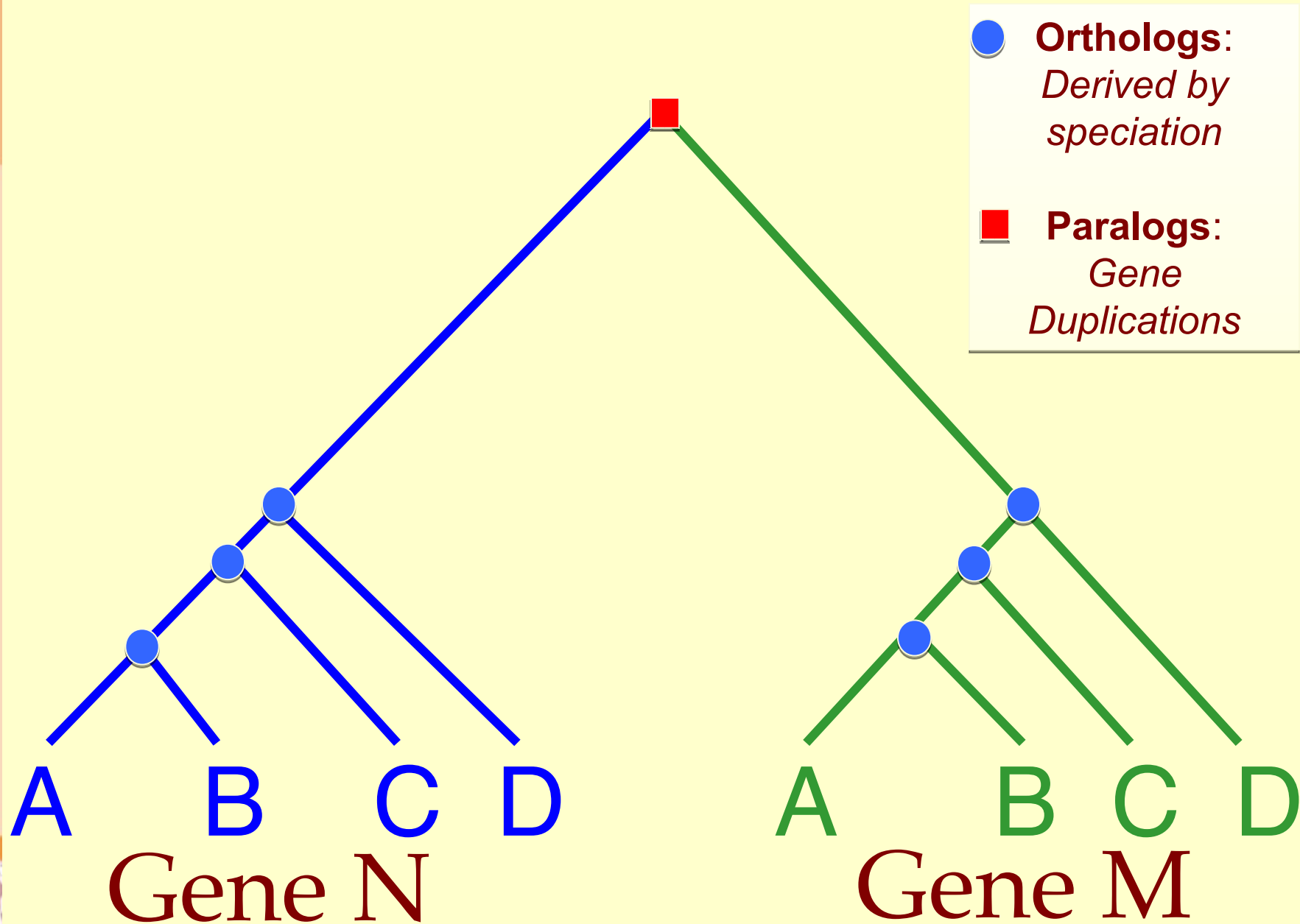
ATCAATTCCCTC



# Gene Conversion



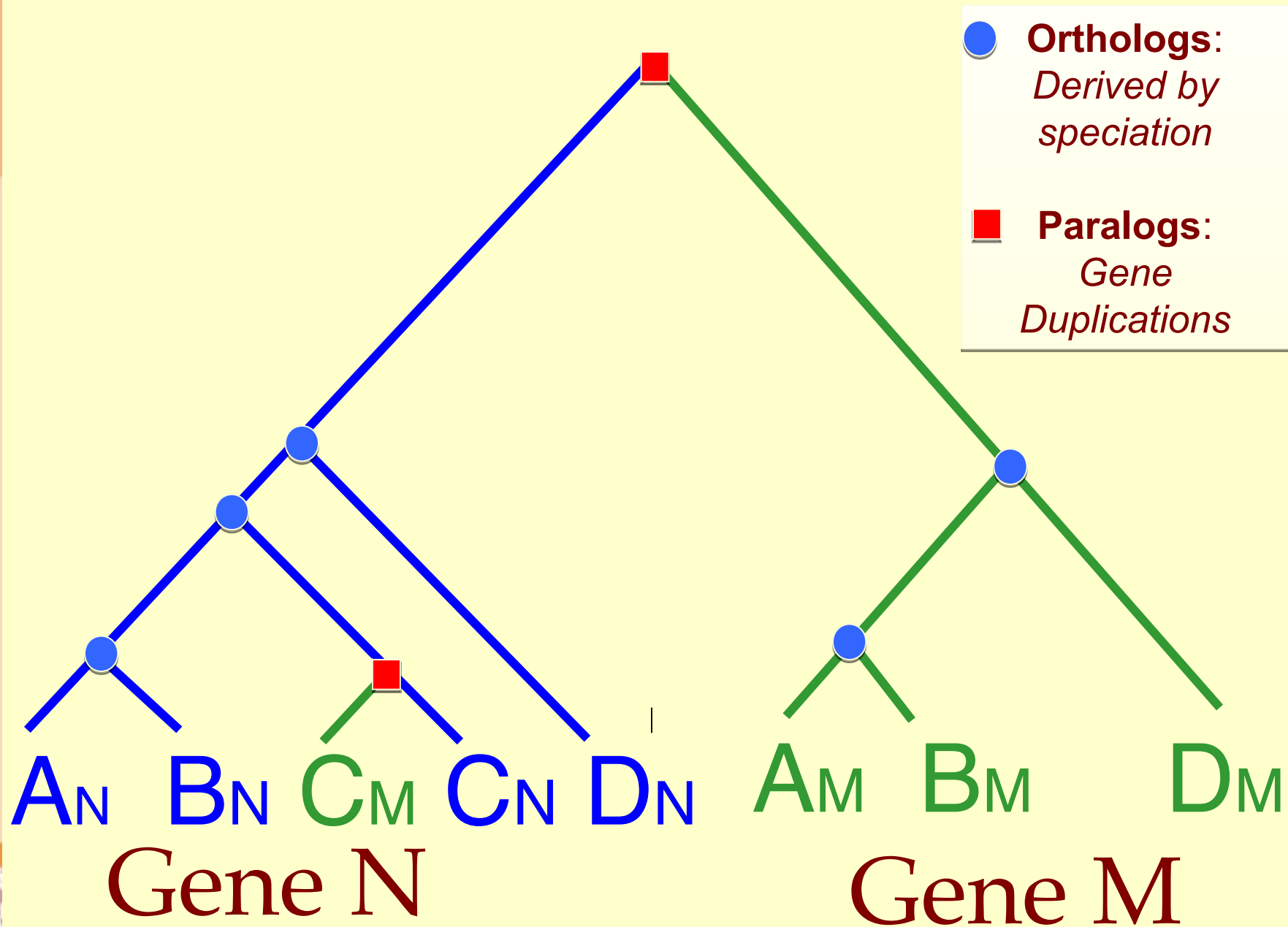
# Challenges Making Trees: Gene Conversion



Thanks to Maryellen Ruvolo

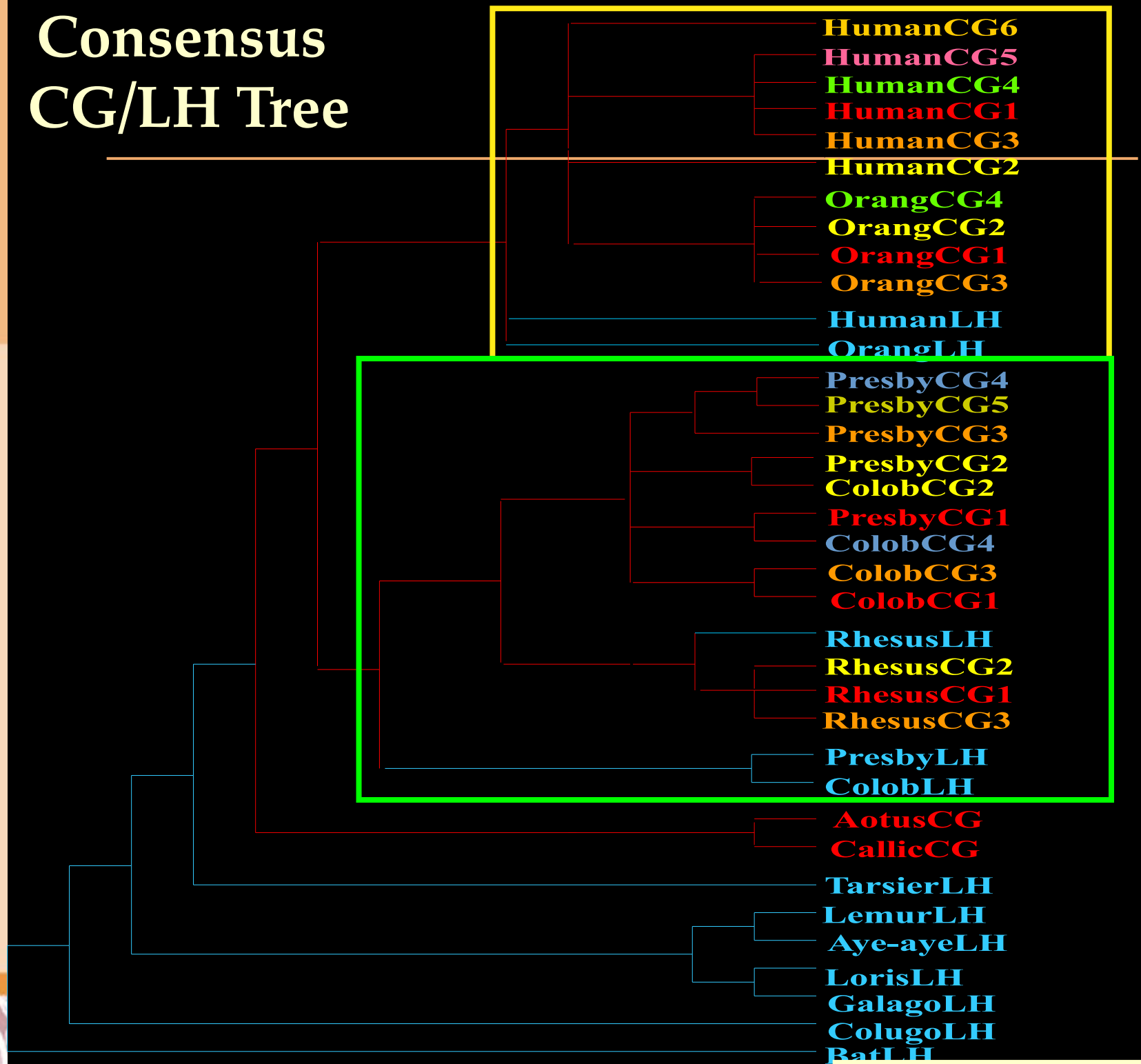
# Challenges Making Trees:

$C_M$  Has Been Converted from  $C_N$



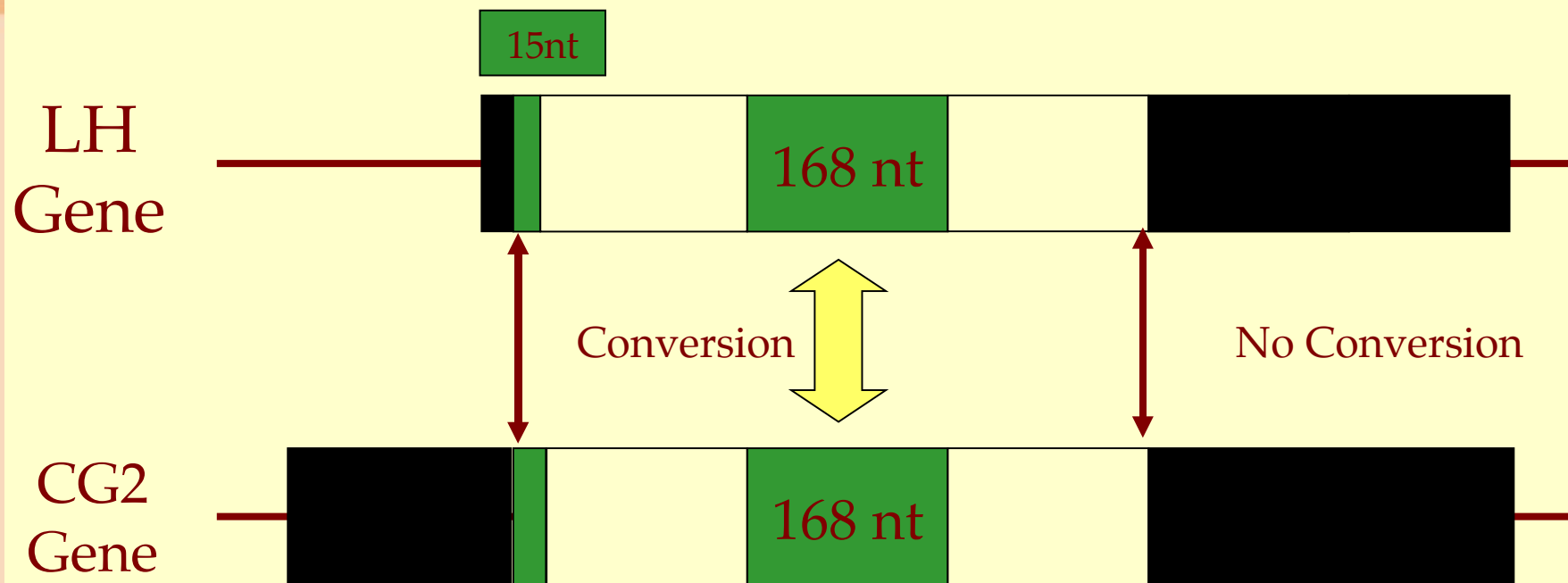
Thanks to Maryellen Ruvolo

# Consensus CG/LH Tree



Thanks to Maryellen Ruvolo

# Gene conversion between 1<sup>st</sup> & 2nd exons of LH, CG2 Genes



# Challenges Making Trees: Varying Rates of Mutation

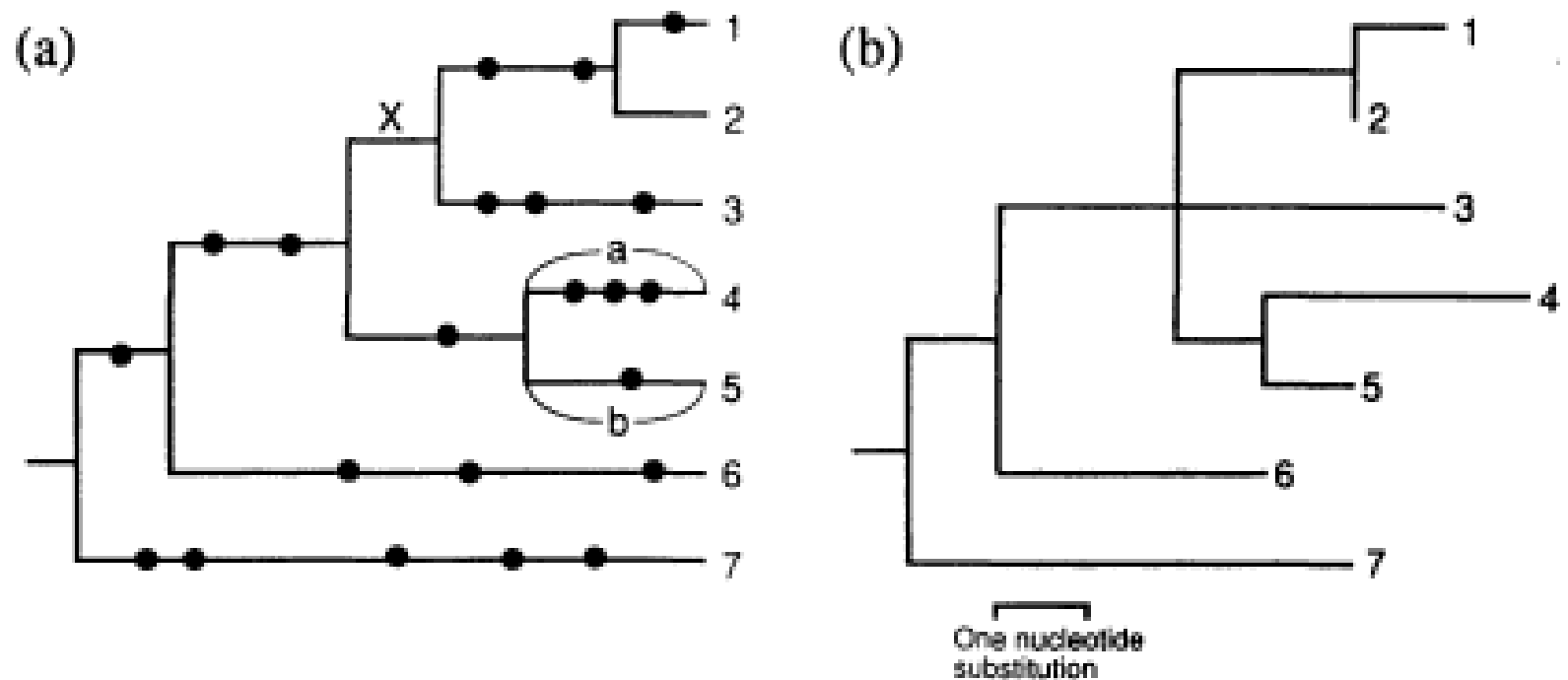
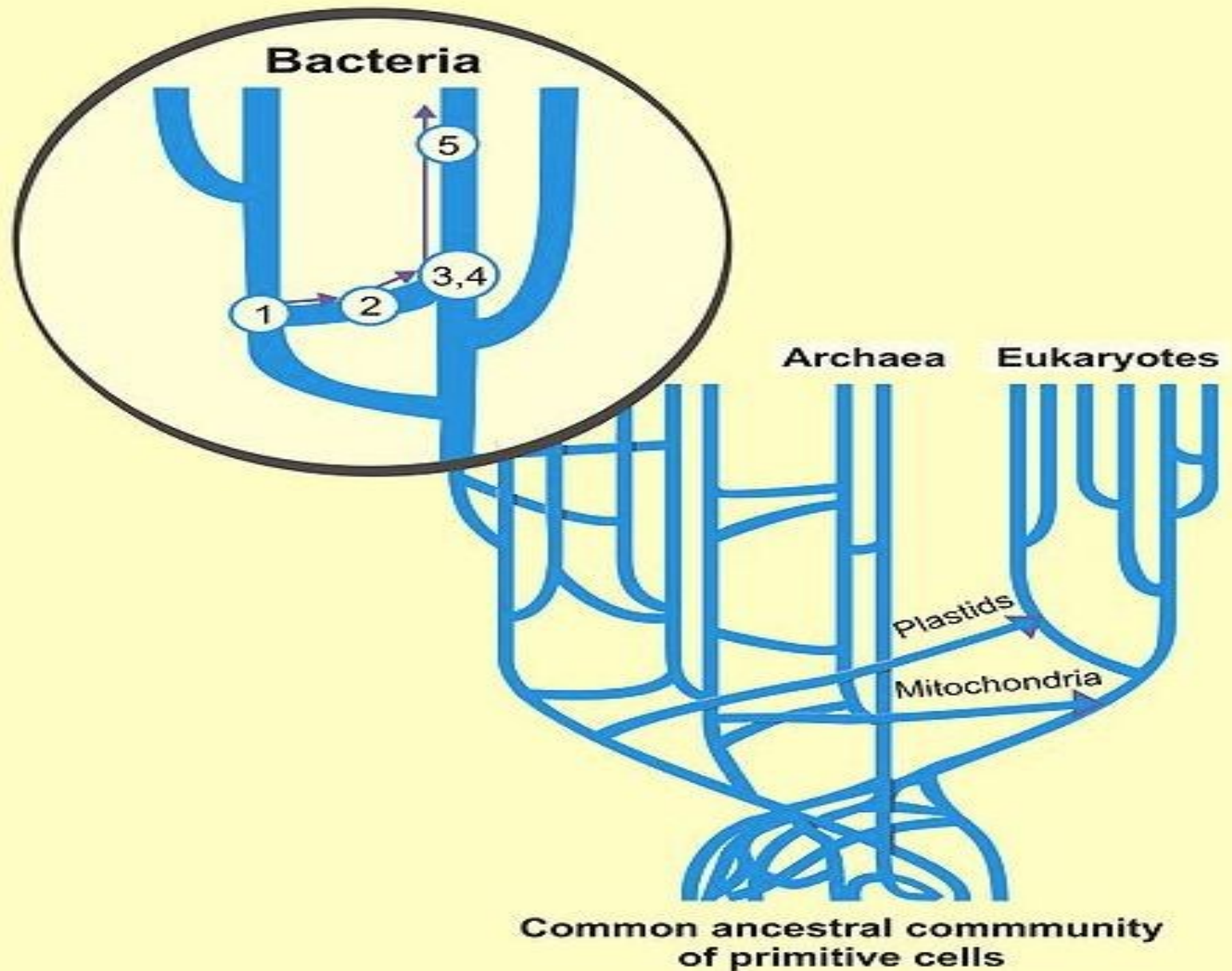


FIG. 5. Examples of the expected gene tree (a) and the corresponding realized gene trees (b). Filled circles on the expected gene tree denote nucleotide substitutions. Because no substitution occurred at branch X of the expected gene tree (a), the corresponding branch does not exist in the realized gene tree (b).

# Challenges Making Trees: Horizontal Gene Transfer



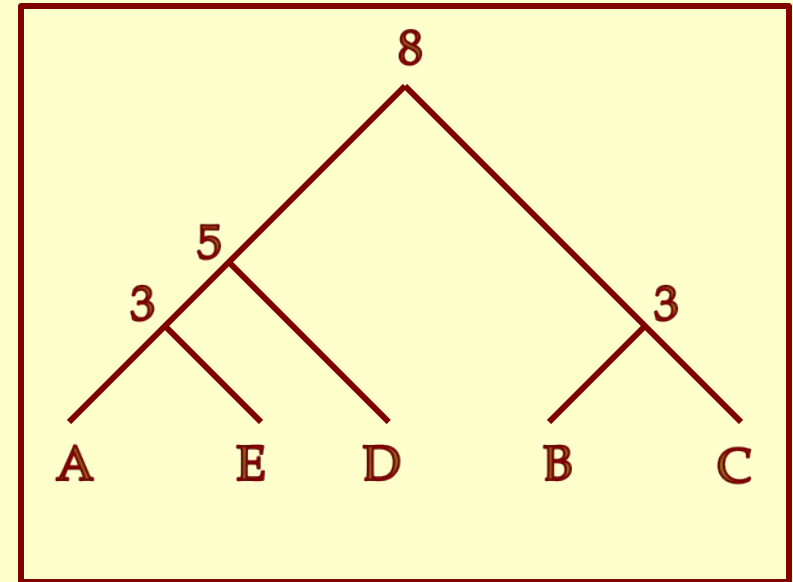
# Maximum Ultrametric Distance Trees

/home/hf/Documents/MA211-DataStructure-Algorithm/20171015—.mp4

Matrix D

	A	B	C	D	E
A	0	8	8	5	3
B		0	3	8	8
C			0	8	8
D				0	5
E					0

Tree T



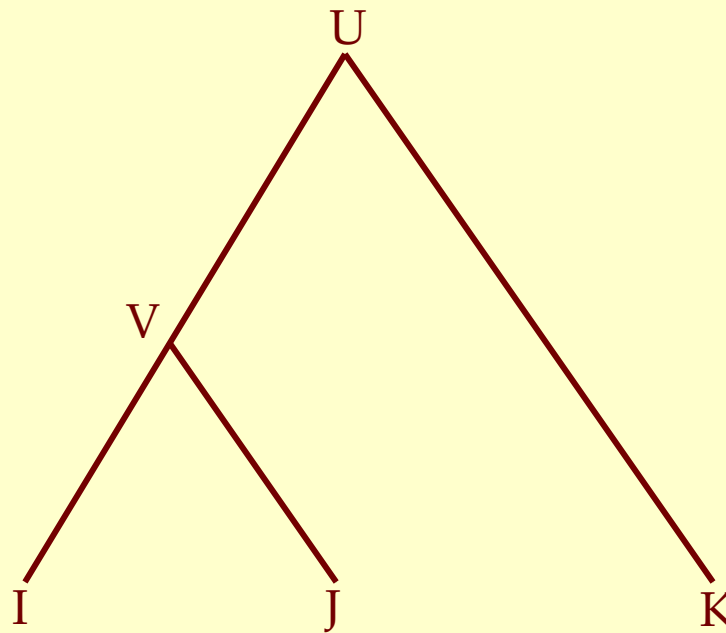
- Matrix D is ultrametric for tree T if:
  - If D is a symmetric  $n$  by  $n$  matrix of distances
  - T contains  $n$  leaves, one from each row or column
  - Each node of T labeled by one entry from D
  - Numbers from root to leaves strictly decrease
  - For any two leaves  $i, j$ ,  $D(i, j)$  labels nearest common ancestor of  $i$  and  $j$  in tree



# Maximum Ultrametric Distance Trees

---

A symmetric matrix  $D$  is ultrametric if and only if for every three leaves  $i$ ,  $j$ , and  $k$ , there is a tie for the maximum distance between  $D(i,j)$ ,  $D(i,k)$  and  $D(j,k)$ .

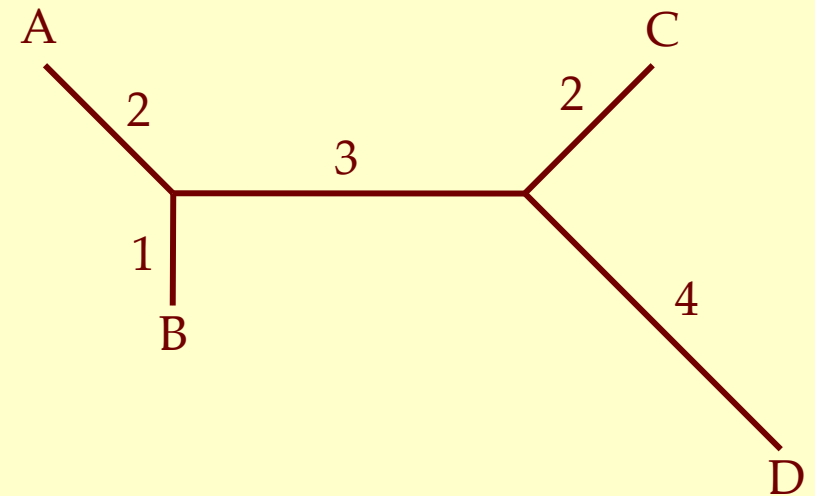


# Additive Distance Trees

Matrix D

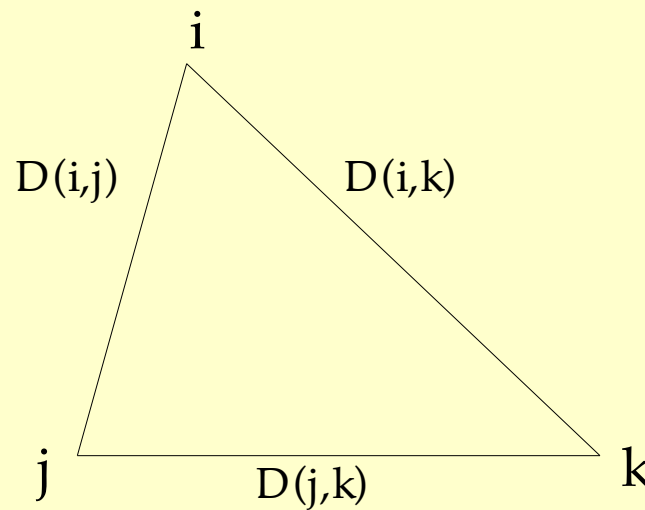
	A	B	C	D
A	0	3	7	9
B		0	6	8
C			0	6
D				0

Tree T

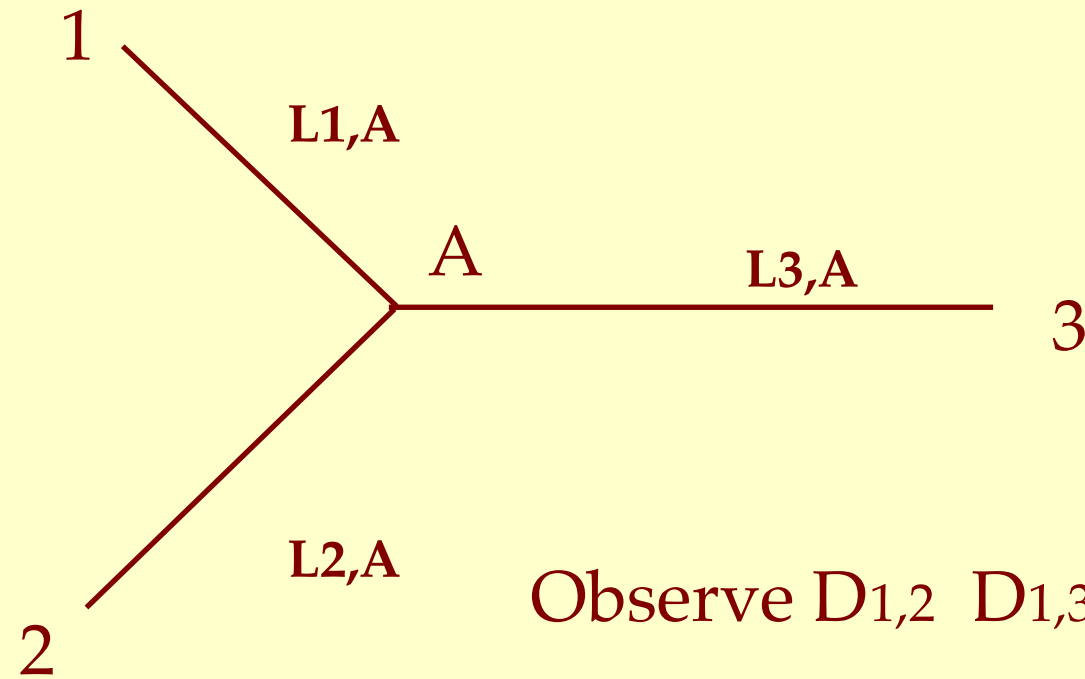


# Distance Metrics Obey the Triangle Inequality

- $D(i,j) \leq D(i,k) + D(j,k)$  for all  $i, j, k$
- (Max Score - Smith-Waterman Score) is a Metric if
  - If Gap-penalty  $\geq 1 + \text{Gap-size} / (n-1)$
  - Assuming match = 1 and mismatch = -1



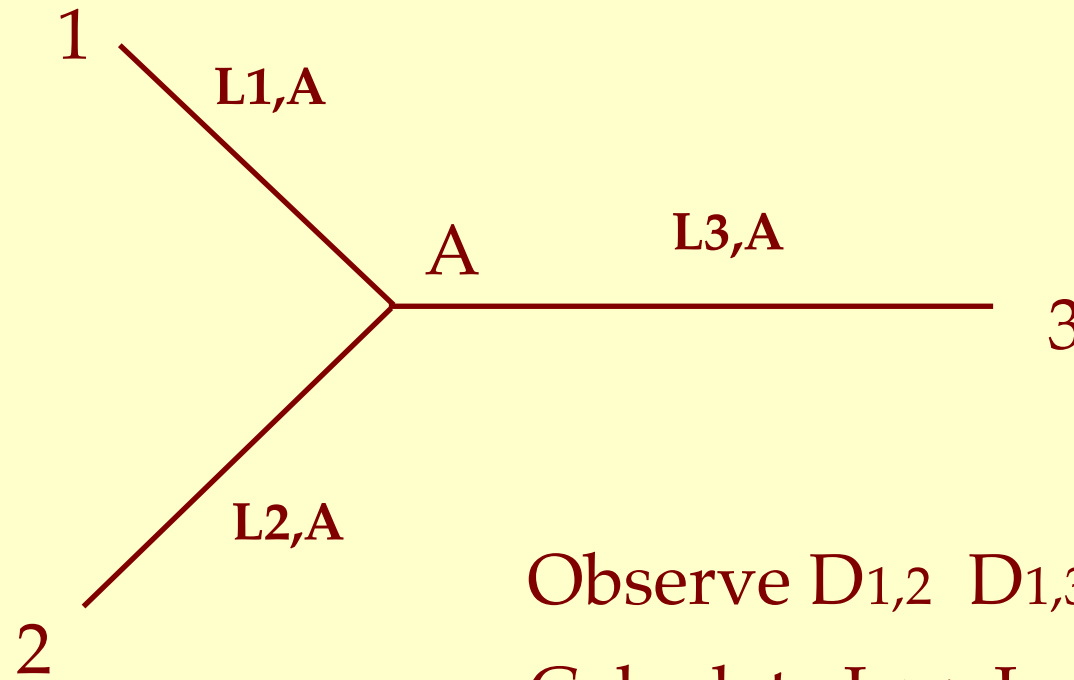
# Three Leaf Tree



Observe  $D_{1,2}$   $D_{1,3}$   $D_{2,3}$

Calculate  $L_{1,A}$   $L_{2,A}$   $L_{3,A}$

# Three Leaf Tree



Observe  $D_{1,2}$   $D_{1,3}$   $D_{2,3}$

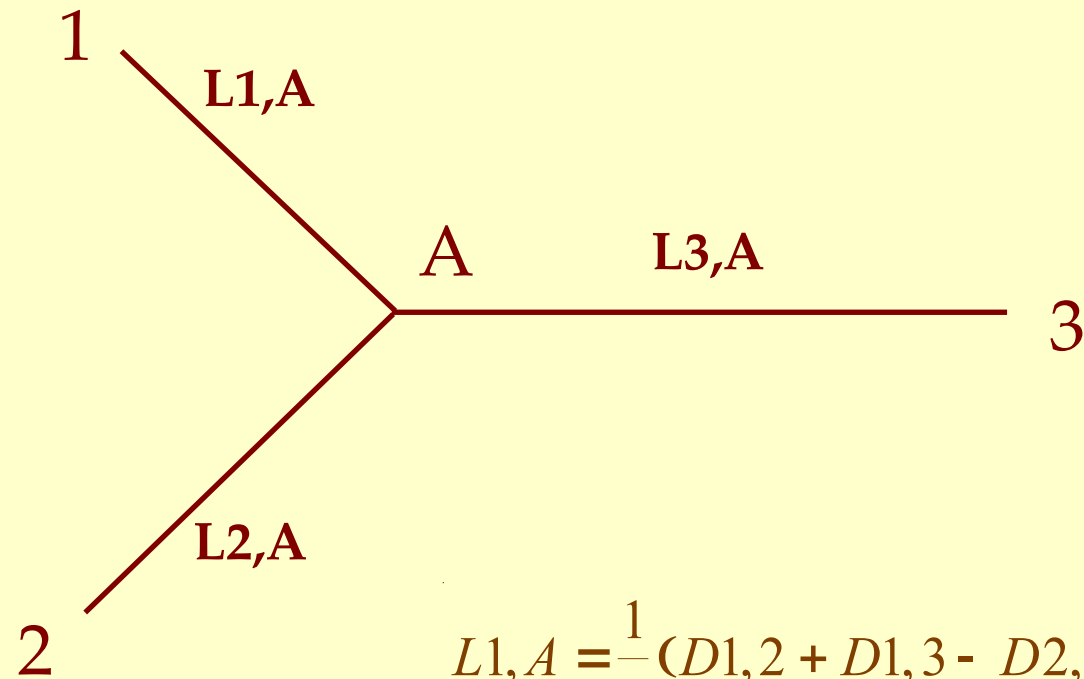
Calculate  $L_{1,A}$   $L_{2,A}$   $L_{3,A}$

$$D_{1,2} = L_{1,A} + L_{2,A}$$

$$D_{1,3} = L_{1,A} + L_{3,A}$$

$$D_{2,3} = L_{2,A} + L_{3,A}$$

# Solution to Three Species Tree

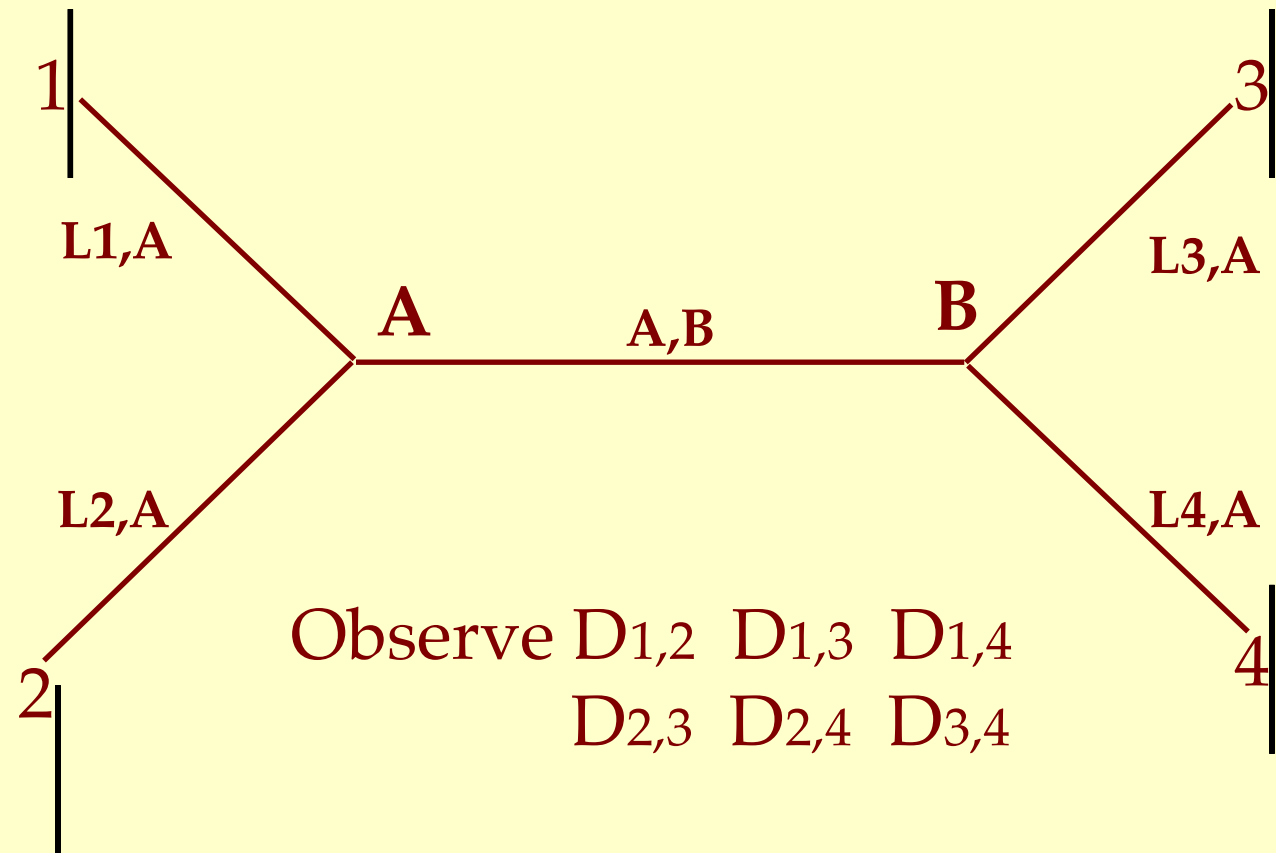


$$L_{1,A} = \frac{1}{2}(D_{1,2} + D_{1,3} - D_{2,3})$$

$$L_{2,A} = \frac{1}{2}(D_{1,2} + D_{2,3} - D_{1,3})$$

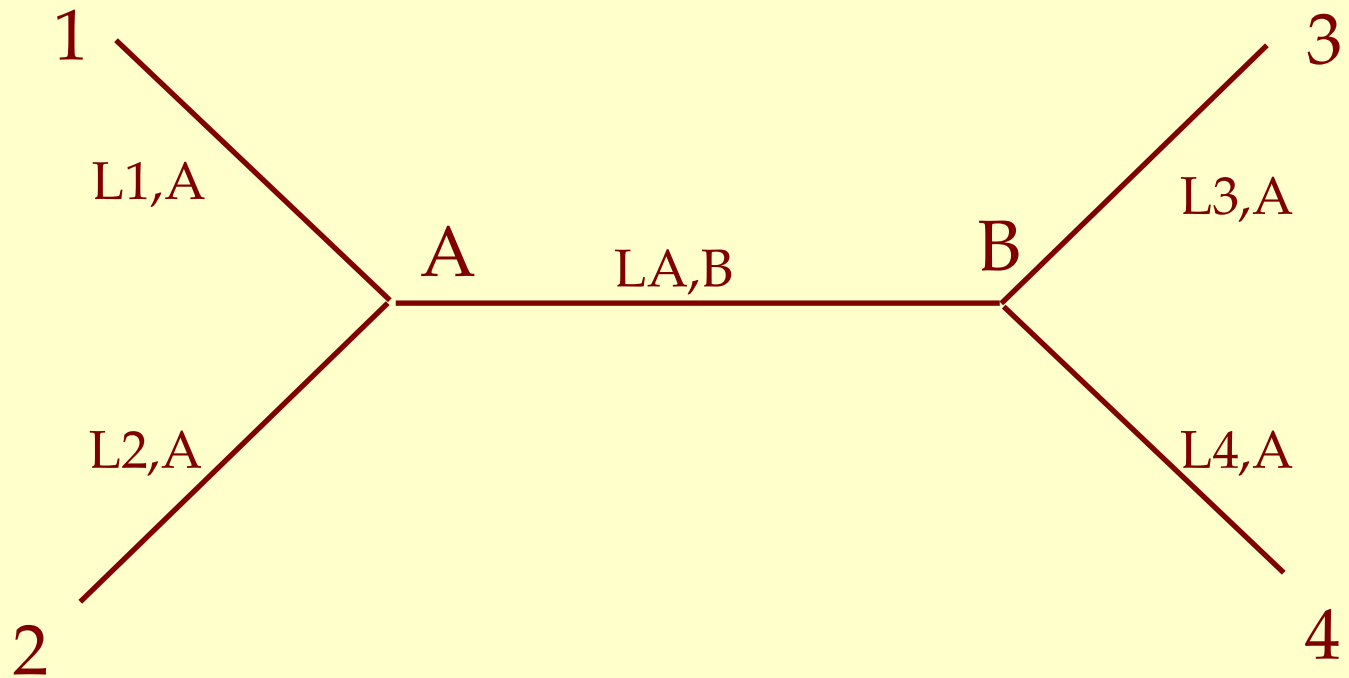
$$L_{3,A} = \frac{1}{2}(D_{1,3} + D_{2,3} - D_{1,2})$$

# Four Species Tree



Calculate  $L_{1,A}$   $L_{2,A}$   $L_{3,B}$   $L_{4,B}$ ,  $L_{A,B}$

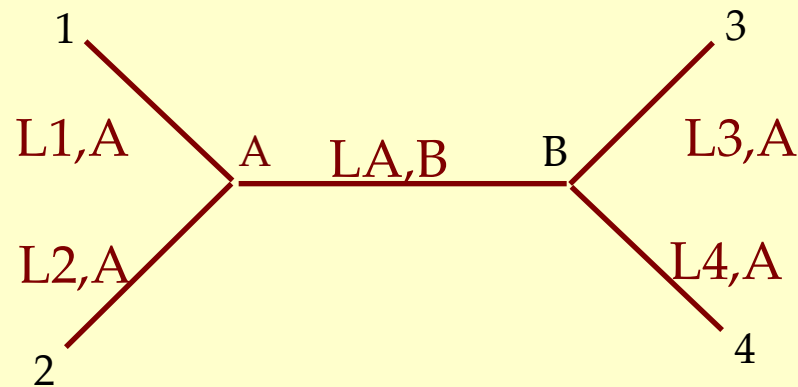
# Four Species Topology



Label species 1, 2, 3, and 4 so that:  
$$D(1,2) + D(3,4) \leq D(1,3) + D(2,4) = D(1,4) + D(2,3)$$



# Solution for Four Species



$$L_{1,A} = 1/4 * (D_{1,3} + D_{1,4} - D_{2,3} - D_{2,4}) + 1/2 * D_{1,2}$$

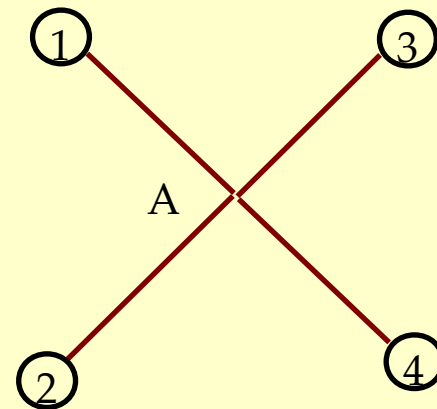
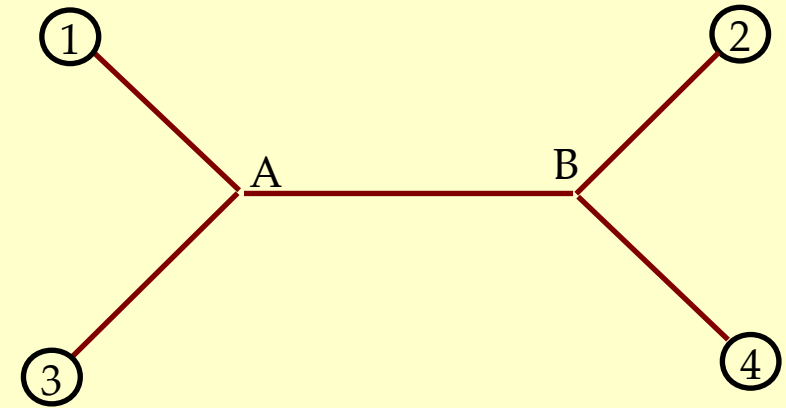
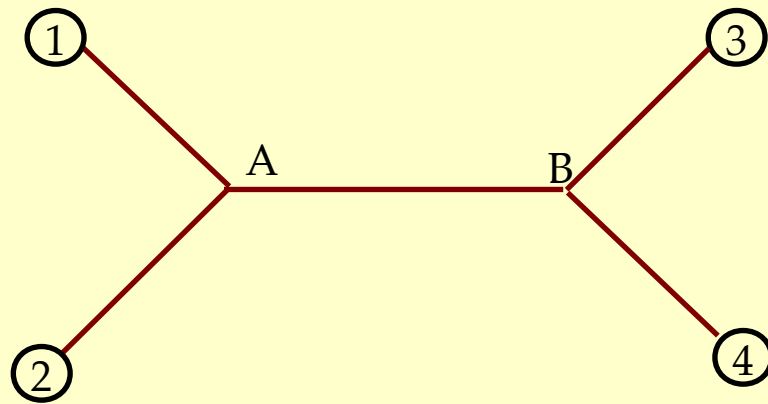
$$L_{2,A} = 1/4 * (D_{2,3} + D_{2,4} - D_{1,3} - D_{1,4}) + 1/2 * D_{1,2}$$

$$L_{B,3} = 1/4 * (D_{1,3} + D_{2,3} - D_{1,4} - D_{2,4}) + 1/2 * D_{3,4}$$

$$L_{B,4} = 1/4 * (D_{1,4} + D_{2,4} - D_{1,3} - D_{2,3}) + 1/2 * D_{3,4}$$

$$L_{A,B} = 1/4 * (D_{1,3} + D_{1,4} + D_{2,3} + D_{2,4}) - 1/2 * (D_{1,2} + D_{3,4})$$

# Four Species => Three Topologies



# Species, Distances, Branches & Topologies

---

Number of Species	Number of Distances	Number of Branches	Number of Topologies
2	1	1	1
3	3	3	1
4	6	5	3
5	10	7	15
6	15	9	105

# Species, Distances, Branches & Topologies

Number of Species	Number of Distances	Number of Branches	Number of Topologies
$n$	$D_n$	$L_n$	$T_n$
$n+1$	$D_{n+n}$	$L_{n+2}$	$L_n * T_n$
•	•	•	•
$n$	$\binom{n}{2} = \frac{n!}{2!(n-2)!}$	$(2n - 3)$	$\prod_{i=1}^{n-2} (2n - 1)$

# Number of Topologies for n Species

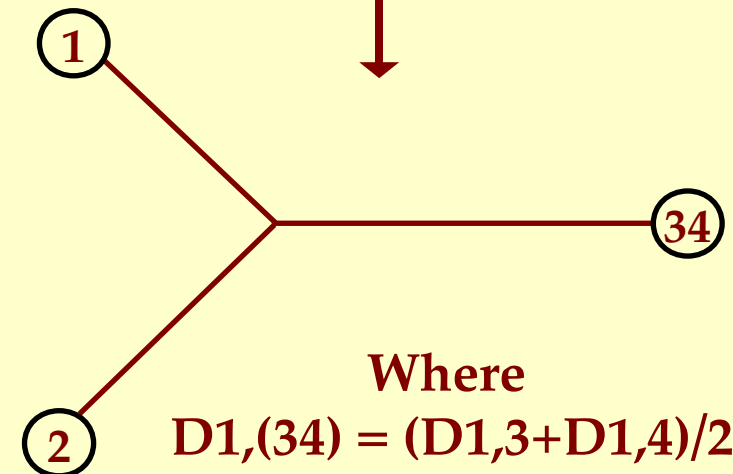
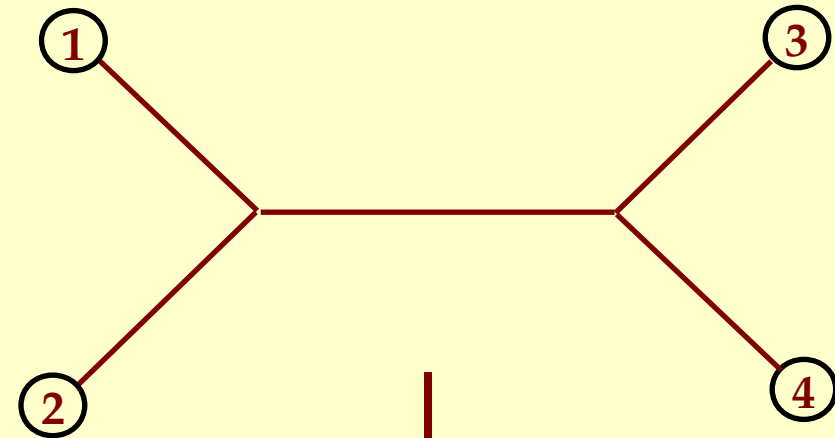
---

n	Tn
3	1
4	3
5	15
6	105
7	945
8	10,395
9	$1.35 \times 10^5$
10	$2.03 \times 10^6$
15	$2.13 \times 10^{14}$
20	$8.20 \times 10^{21}$

# UPGMA: Unweighted Pair Group Method with Arithmetic Average

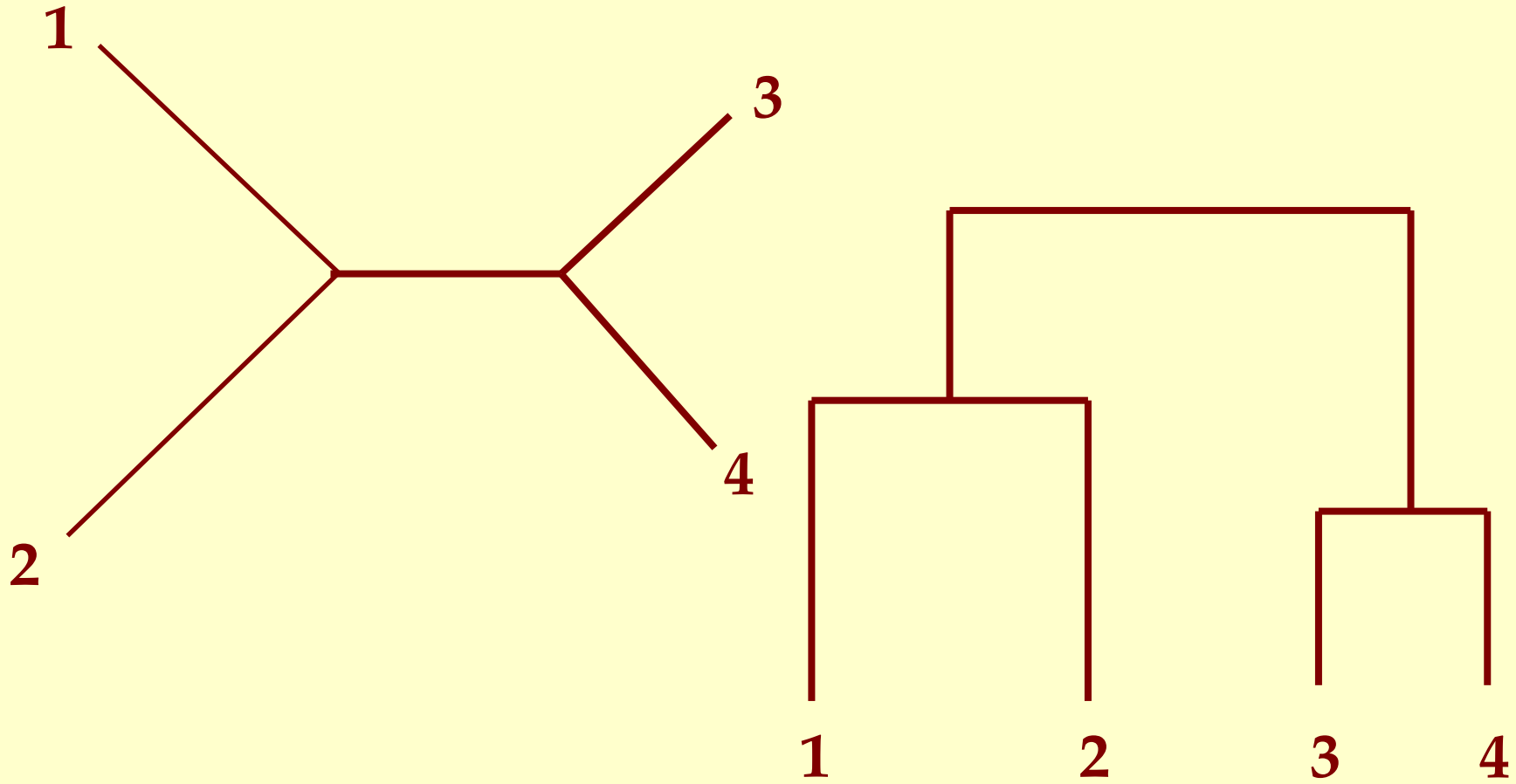
OTU	1	2	3
2	D1,2		
3	D1,3	D2,3	
4	D1,4	D2,4	D3,4

OTU	1	2
2	D1,2	
(34)	D1,(34)	D2,(34)

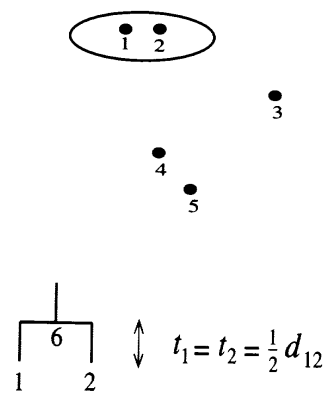


Where  
 $D1,(34) = (D1,3 + D1,4)/2$   
and  
 $D2,(34) = (D2,3 + D2,4)/2$

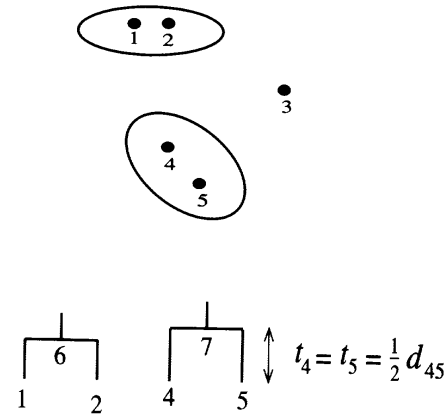
# UPGMA Dendrogram



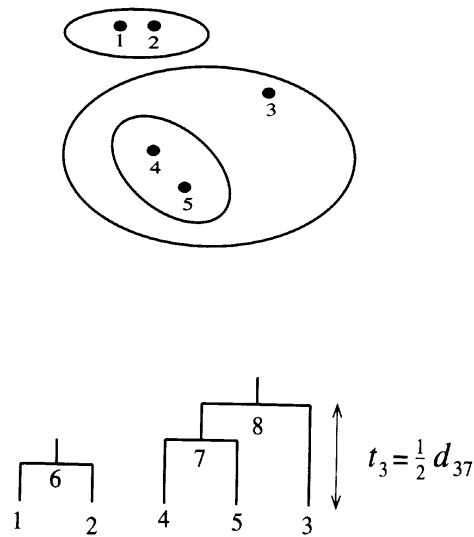
# UPGMA Clustering



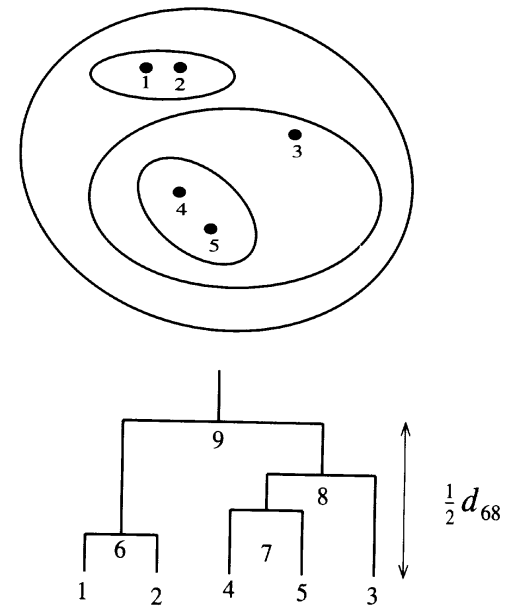
(i)



(ii)



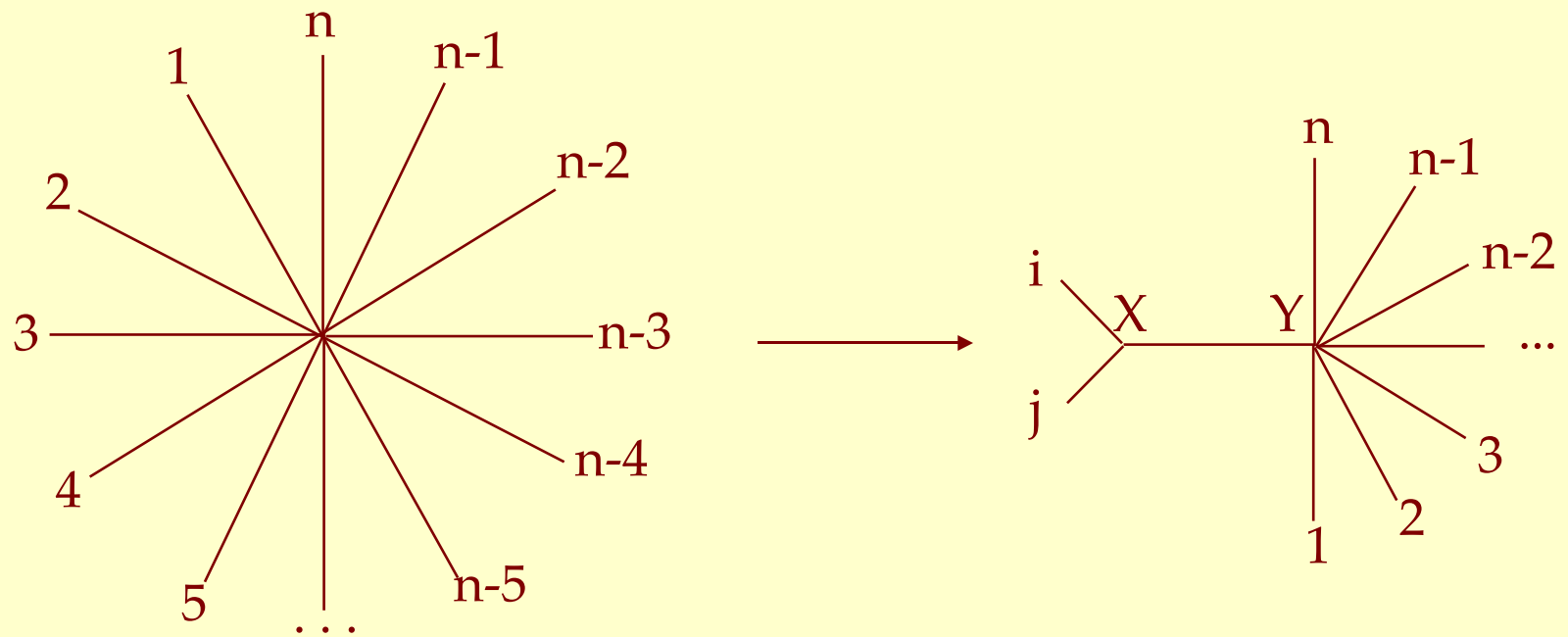
(iii)



(iv)



# Neighbor Joining Method



For starlike tree  $S_0 = Q / (n - 1)$  where  $Q = \sum_{i < j} D_{ij}$

For nearest neighbor tree  $S_{ij} = (B_{iX} + B_{jX}) + B_{XY} + \sum_{k \neq i, j} B_{kY}$

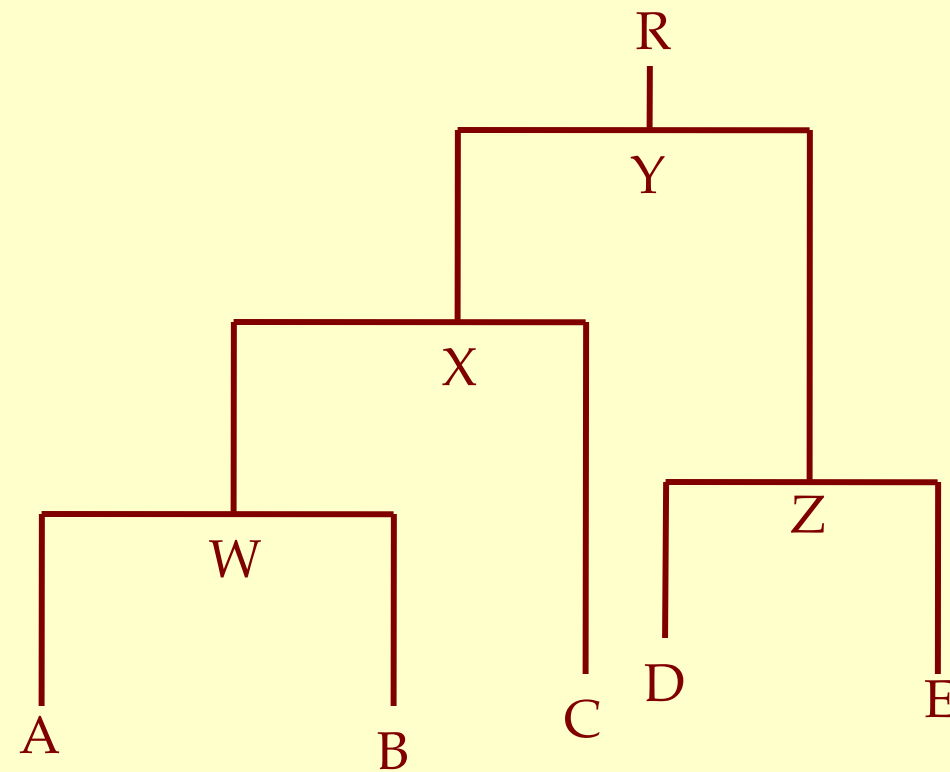
$$\begin{aligned} D_{ij} &= B_{iX} + B_{jX} & D_{ik} &= B_{iX} + B_{XY} + B_{kY} & (k \neq i, j) \\ D_{kl} &= B_{iY} + B_{jY} & D_{jk} &= B_{jX} + B_{XY} + B_{kY} & (k, l \neq i, j) \end{aligned}$$

$$B_{XY} = \frac{Q - (n - 1)D_{ij} - \frac{(n - 1)}{(n - 3)} \sum_{k, l \neq i, j} D_{kl}}{2(n - 2)}$$

$$S_{ij} = \frac{D_{ij}}{2} + \frac{2 \sum_{i < j} D_{ij} - \sum_j D_{ij} - \sum_i D_{ij}}{2(n - 2)}$$

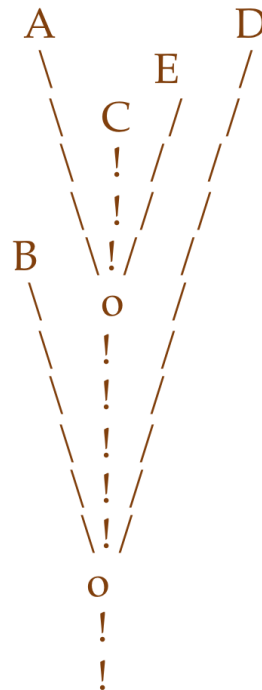
# Nearest Neighbor Dendrogram

---



# New Hampshire Standard Tree

If we have this rooted tree:



then the tree file is represented by the following sequence of printable characters, starting at the beginning of the file:

(B,(A,C,E),D);

(B:6.0,(A:5.0,C:3.0,E:4.0):5.0,D:11.0);

# SeqWeb GrowTree Program

<http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=evolution-prot>

SeqWeb v3.1



	Programs	Managers	Help Topics   Support																																									
<b>Programs</b>	<b>GrowTree</b> <span>?</span>																																											
Comparison	<b>Construct Phylogenetic Trees from Peptide Sequences.</b>																																											
Database Searching	<input type="text" value="Input sequences:"/> <span>Select From: <input type="button" value="Default"/> <input type="button" value="Project"/> <input type="button" value="Local File"/> <input type="button" value="Clipboard"/> <input type="button" value="Database"/></span>																																											
Similarity	<table border="1"> <thead> <tr> <th>Sequence</th> <th>Description</th> <th>Type</th> <th>Length</th> <th>Range</th> </tr> </thead> <tbody> <tr> <td><a href="#">myg_phyca</a></td> <td>myg_phyca</td> <td>P</td> <td>153</td> <td><a href="#">1 .. 153</a></td> </tr> <tr> <td><a href="#">glb5_petma.pep</a></td> <td>ID GLB5_PETMA STANDARD; PRT; 149 AA.</td> <td>P</td> <td>149</td> <td><a href="#">1 .. 149</a></td> </tr> <tr> <td><a href="#">hba_human</a></td> <td>hba_human</td> <td>P</td> <td>141</td> <td><a href="#">1 .. 141</a></td> </tr> <tr> <td><a href="#">hba_horse.pep</a></td> <td>ID HBA_HORSE STANDARD; PRT; 141 AA.</td> <td>P</td> <td>141</td> <td><a href="#">1 .. 141</a></td> </tr> <tr> <td><a href="#">hbb_horse.pep</a></td> <td>ID HBB_HORSE STANDARD; PRT; 146 AA.</td> <td>P</td> <td>146</td> <td><a href="#">1 .. 146</a></td> </tr> <tr> <td><a href="#">lgb1_soybn.pep</a></td> <td>- ID LGB1_SOYBN STANDARD; PRT; 143 AA.</td> <td>P</td> <td>143</td> <td><a href="#">1 .. 143</a></td> </tr> <tr> <td><a href="#">hbb_human</a></td> <td>hbb_human</td> <td>P</td> <td>146</td> <td><a href="#">1 .. 146</a></td> </tr> </tbody> </table>				Sequence	Description	Type	Length	Range	<a href="#">myg_phyca</a>	myg_phyca	P	153	<a href="#">1 .. 153</a>	<a href="#">glb5_petma.pep</a>	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	<a href="#">1 .. 149</a>	<a href="#">hba_human</a>	hba_human	P	141	<a href="#">1 .. 141</a>	<a href="#">hba_horse.pep</a>	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	<a href="#">1 .. 141</a>	<a href="#">hbb_horse.pep</a>	ID HBB_HORSE STANDARD; PRT; 146 AA.	P	146	<a href="#">1 .. 146</a>	<a href="#">lgb1_soybn.pep</a>	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	<a href="#">1 .. 143</a>	<a href="#">hbb_human</a>	hbb_human	P	146	<a href="#">1 .. 146</a>
Sequence	Description	Type	Length	Range																																								
<a href="#">myg_phyca</a>	myg_phyca	P	153	<a href="#">1 .. 153</a>																																								
<a href="#">glb5_petma.pep</a>	ID GLB5_PETMA STANDARD; PRT; 149 AA.	P	149	<a href="#">1 .. 149</a>																																								
<a href="#">hba_human</a>	hba_human	P	141	<a href="#">1 .. 141</a>																																								
<a href="#">hba_horse.pep</a>	ID HBA_HORSE STANDARD; PRT; 141 AA.	P	141	<a href="#">1 .. 141</a>																																								
<a href="#">hbb_horse.pep</a>	ID HBB_HORSE STANDARD; PRT; 146 AA.	P	146	<a href="#">1 .. 146</a>																																								
<a href="#">lgb1_soybn.pep</a>	- ID LGB1_SOYBN STANDARD; PRT; 143 AA.	P	143	<a href="#">1 .. 143</a>																																								
<a href="#">hbb_human</a>	hbb_human	P	146	<a href="#">1 .. 146</a>																																								
Reference	<input type="button" value="Refresh"/> <input type="button" value="Clear"/>																																											
Evolution	<b>Input Parameters:</b>																																											
Mapping	<table border="1"> <tbody> <tr> <td rowspan="3"><a href="#">Distance Correction Method</a></td> <td>uncorrected distance</td> <td><input type="radio"/></td> </tr> <tr> <td>Jukes-Cantor distance</td> <td><input type="radio"/></td> </tr> <tr> <td>Kimura distance</td> <td><input checked="" type="radio"/></td> </tr> <tr> <td rowspan="2"><a href="#">Tree Construction Method</a></td> <td>Neighbor joining</td> <td><input checked="" type="radio"/></td> </tr> <tr> <td>UPGMA</td> <td><input type="radio"/></td> </tr> </tbody> </table>				<a href="#">Distance Correction Method</a>	uncorrected distance	<input type="radio"/>	Jukes-Cantor distance	<input type="radio"/>	Kimura distance	<input checked="" type="radio"/>	<a href="#">Tree Construction Method</a>	Neighbor joining	<input checked="" type="radio"/>	UPGMA	<input type="radio"/>																												
<a href="#">Distance Correction Method</a>	uncorrected distance	<input type="radio"/>																																										
	Jukes-Cantor distance	<input type="radio"/>																																										
	Kimura distance	<input checked="" type="radio"/>																																										
<a href="#">Tree Construction Method</a>	Neighbor joining	<input checked="" type="radio"/>																																										
	UPGMA	<input type="radio"/>																																										
Pattern Recognition																																												
Primer Selection																																												
Protein Analysis																																												
Nucleic Acid																																												
Secondary Structure																																												
Translation																																												
Utilities																																												



# GrowTree Parameters

## Input Parameters:

<a href="#">Distance Correction Method</a>	uncorrected distance	<input type="radio"/>
	Jukes-Cantor distance	<input type="radio"/>
	Kimura distance	<input checked="" type="radio"/>
<a href="#">Tree Construction Method</a>	Neighbor joining	<input checked="" type="radio"/>
	UPGMA	<input type="radio"/>
<p>Select a sequence comparison matrix. This matrix determines how matches and mismatches are scored. The default penalties for gap creation and extension are given after each matrix name.</p>		
<a href="#">Scoring Matrix</a>	<input type="text" value="blosum62"/>	
<a href="#">Set gap creation penalty</a>	<input type="text" value="8"/>	
<a href="#">Set gap extension penalty</a>	<input type="text" value="2"/>	
<p>Limit the maximum input sequence range only when needed. Setting a higher limit allows you to align longer sequences while setting a lower limit allows you to add more and longer gaps to each sequence.</p>		
<a href="#">Maximum input sequence range</a>	<input type="text" value="5000"/> (range 1 thru 7000)	
<p>Limit the maximum number of gaps only when needed. Setting a higher limit allows you to add more and longer gaps to each sequence while setting a lower limit allows you to align a greater number of sequences.</p>		
<a href="#">Maximum number of gap characters ('.' and '~') added to any sequence</a>	<input type="text" value="2000"/> (range 0 thru 7000)	
<a href="#">Consider partial matches between degenerate symbols for uncorrected or Jukes-Cantor distance</a>	<input type="checkbox"/>	
<a href="#">Gap weight for uncorrected or Jukes-Cantor distance</a>	<input type="text" value="0.0"/> (range 0.0 thru 2.0)	
<a href="#">Report negative branch lengths as negative (instead of zero)</a>	<input type="checkbox"/>	
<a href="#">Display Tree As:</a>	phylogram (branch lengths proportional to distance)	<input checked="" type="radio"/>
	cladogram (all branches the same length)	<input type="radio"/>

Run

Reset

## SeqWeb v3.1

**Evolutionary Analysis Results**

## Genetic Distances

Calculated over: 9 to 157

Correction method: Kimura protein distance

Distances are: estimated number of substitutions per 100 amino acids

Symmatrix version 1

Number of matrices: 1

//

Matrix 1, dimension: 7

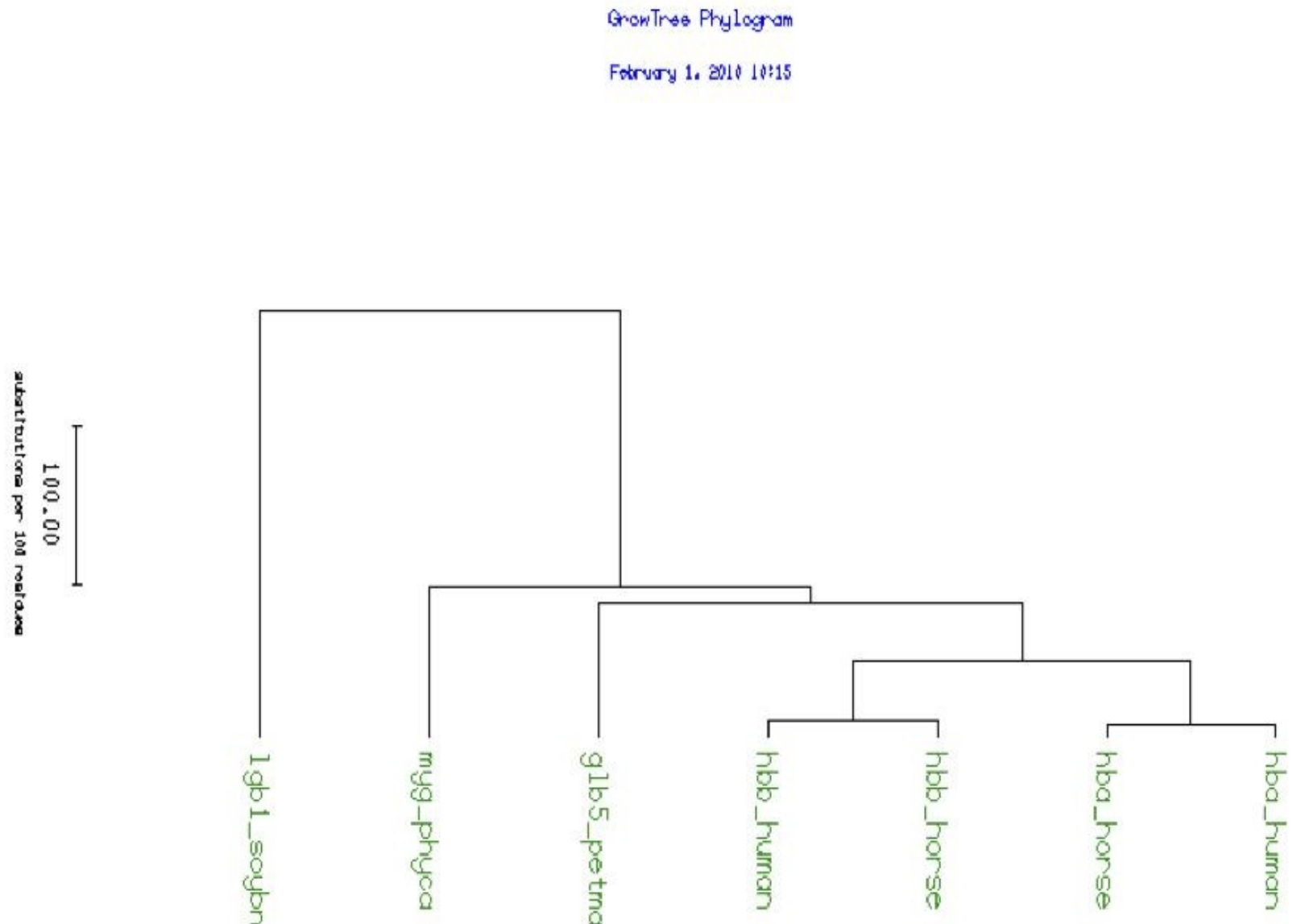
Key for column and row indices:

- 1 hba\_human
- 2 hba\_horse
- 3 hbb\_horse
- 4 hbb\_human
- 5 glb5\_petma
- 6 myg\_phyca
- 7 lgb1\_soybn

Matrix 1: Part 1

	1	2	3	4	5	6	7
1	0.00	13.39	95.79	93.49	134.46	173.98	540.37
2		0.00	91.25	95.79	134.46	179.53	540.37
3			0.00	18.90	211.52	180.02	257.93
4				0.00	196.68	180.02	287.05
5					0.00	213.42	336.95
6						0.00	999.99
7							0.00

# GrowTree Phylogram (UPGMA)



# GrowTree Alignment

Symbol comparison table: share\_matrix:blosum62.cmp CompCheck: 11

Gapweight: 8  
GapLengthWeight: 2

Pileup MSF: 165 Type: P February 1, 2010 10:15 Check: 6593 ..

Name: hba_human	Len: 165	Check: 1231	Weight: 1.00
Name: hba_horse	Len: 165	Check: 2167	Weight: 1.00
Name: hbb_horse	Len: 165	Check: 9310	Weight: 1.00
Name: hbb_human	Len: 165	Check: 208	Weight: 1.00
Name: glb5_petma	Len: 165	Check: 2079	Weight: 1.00
Name: myg_phyca	Len: 165	Check: 4320	Weight: 1.00
Name: lgb1_soybn	Len: 165	Check: 7278	Weight: 1.00

//

hba_human	1	~~~~~v	lspadktnvk	aawgkvgaha	geygaealer	mflsfpttk	5
hba_horse		~~~~~V	LSAADKTNVK	AAWSKVGGHA	GEYGAEALER	MFLGFPTTK	
hbb_horse		~~~~~VQ	LSGEEKA AVL	ALWDKV..NE	EEVVGGEALGR	LLVVPWTQ	
hbb_human		~~~~~Vh	ltpeeksavt	alwgkv..nv	devggealgr	llvypwtq	
glb5_petma		PIVDTGSVAP	LSAAEKT KIR	SAWAPVYSTY	ETSGVDILVK	FFTSTPAAQ	
myg_phyca		~~~~~v	lsegewqlvl	hwakveadv	aghgqdilir	lfkshpetl	
lgb1_soybn		~~~~~ga	ftekqealvs	ssfeafkani	pqysvvfyns	ilekapaak	

hba_human	51	yfp hf .dlsh	.....gsaqv	kghgkkvada	ltnavahvdd	mpnalsals	10
hba_horse		YFPHF.DLSH	.....GSAQV	KAHGKKVGDA	LT LAVGHLDD	LPGALS NLS	
hbb_horse		FFDSFGDLSN	PGAVMGNPKV	KAHGKKVLHS	FGEGVHHLDN	LKGTFAALS	
hbb_human		ffesfgdlst	pdavmgnpkv	kahgkkvlga	fsdglahldn	lkgtfatls	
glb5_petma		FFPKFKGLTT	ADQLKKSADV	RWHAERIINA	VNDAVASMDD	TEKMMSMKLR	
myg_phyca		kfd r f k h l k t	eaemkasedl	kkhgvtvlta	lg...ailkk	kghheaelk	
lgb1_soybn		lfsflan...	.gvdptnpl	tghaeklfal	vrdsagql.k	tngtvvada	

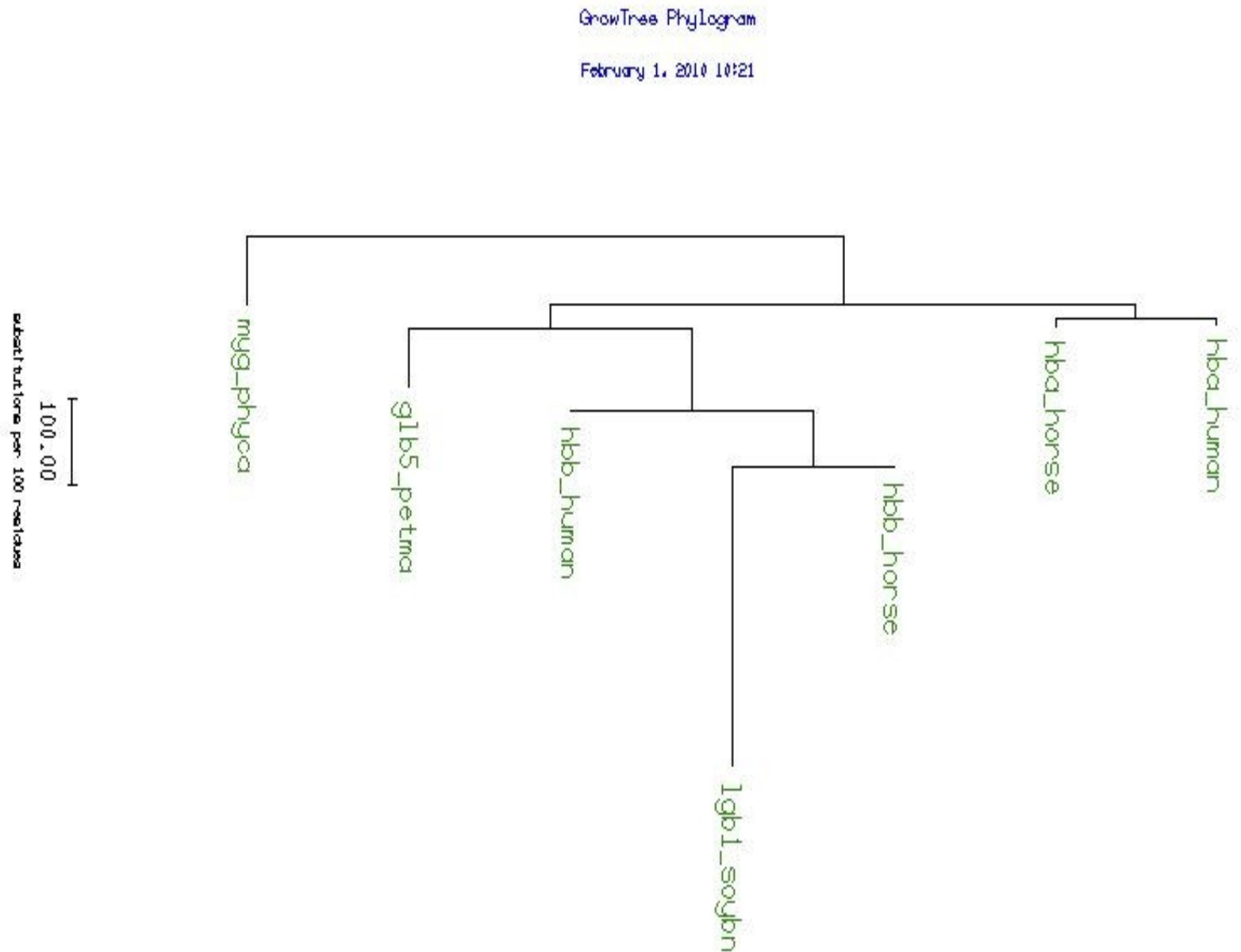
hba_human	101	l...hahklr	vdpvnfklls	hc llvtlaah	lpaeftpavh	asl d k f l a s	15
hba_horse		L...HAHKLR	VDPVNFKLLS	HCLLSTLAVH	LPNDFTPAVH	ASLDKFLSS	
hbb_horse		L...HCDKLH	VDPENFRLLG	NVLVWVLARH	FGKDFTPELQ	ASYQKV VAG	
hbb_human		l...hcdklh	vdpenfrllg	nv lvcvlahh	fgkeftppvq	aayqkvvmag	
glb5_petma		LSGKHAKSFQ	VDPQYFKVLA	AVIADTVA..	.....AGD	AGFEKLM SM	
myg_phyca		laqshatk h k	ipikylefis	ea i i h v l h s r	hpgdfgadaq	gamnkalel	
lgb1_soybn		lvsihaqkav	tdpq.fvvvk	ea l l k t i k e a	v g g n w s d e l s	sawevayde	

hba_human	151	stvltskyr~	~~~~~	165
hba_horse		STVLTSKYR~	~~~~~	
hbb_horse		ANALAHKYH~	~~~~~	
hbb_human		analahkyh~	~~~~~	
glb5_petma		CILLRSAY~	~~~~~	
myg_phyca		rkdiaa kyke	lgyqq	
lgb1_soybn		aaaikka~	~~~~~	



# GrowTree Neighbor Joining Tree

<http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=evolution-prot>



# GrowTree VegF Input

<http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=evolution-prot>

## GrowTree

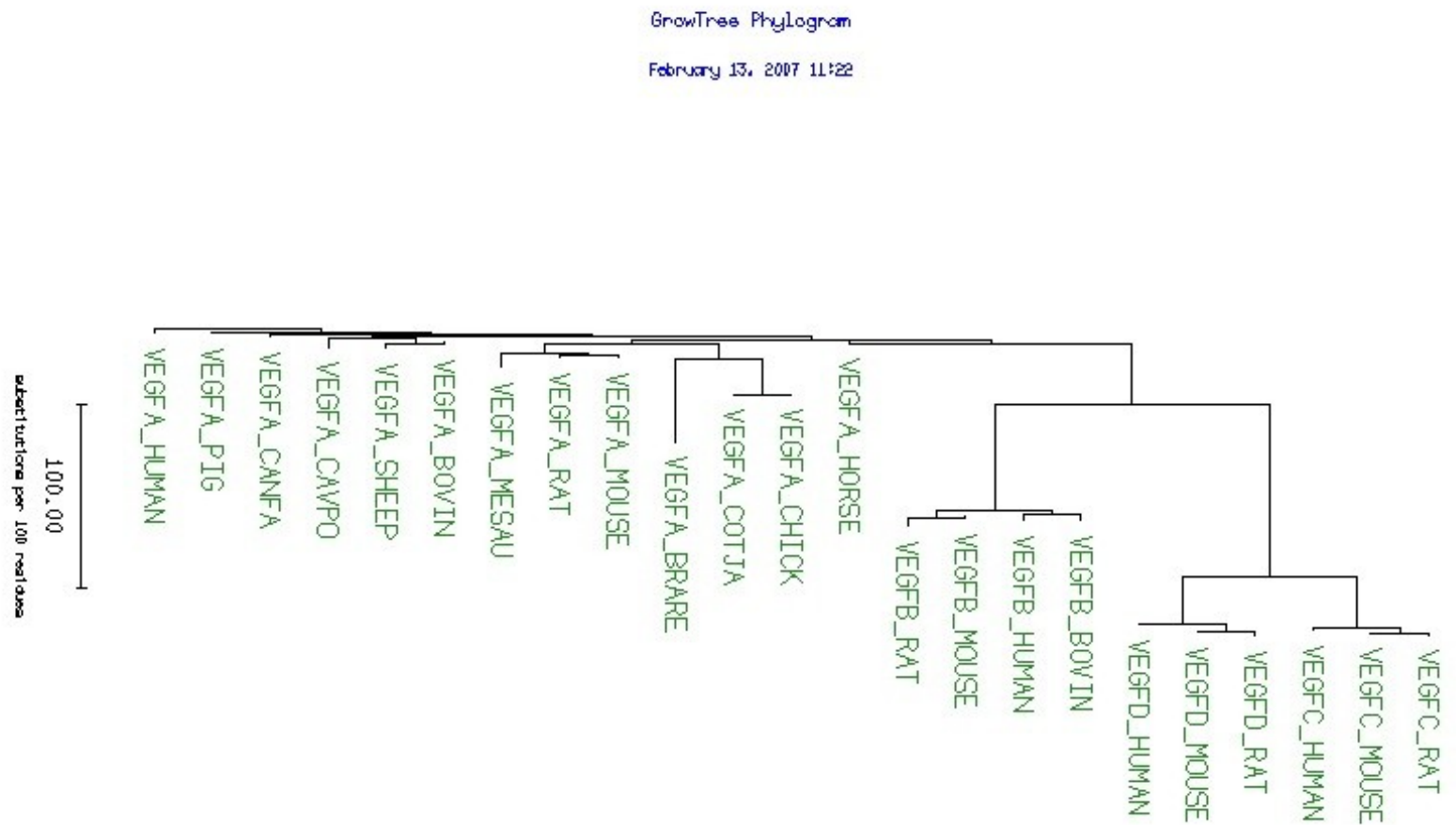
### Construct Phylogenetic Trees from Peptide Sequences.

Input sequences:

Select From:

Sequence	Description	Type	Length	Range
<a href="#">VEGFA_CHICK.ssf</a>	VEGFA_CHICK 216 aa 01-JAN-1970	P	216	<a href="#">1 .. 216</a>
<a href="#">VEGFA_MOUSE.ssf</a>	VEGFA_MOUSE 214 aa 01-JAN-1970	P	214	<a href="#">1 .. 214</a>
<a href="#">VEGFA_BRARE.ssf</a>	VEGFA_BRARE 188 aa 01-JAN-1970	P	188	<a href="#">1 .. 188</a>
<a href="#">VEGFA_RAT.ssf</a>	VEGFA_RAT 214 aa 01-JAN-1970	P	214	<a href="#">1 .. 214</a>
<a href="#">VEGFD_RAT.ssf</a>	VEGFD_RAT 326 aa 01-JAN-1970	P	326	<a href="#">1 .. 326</a>
<a href="#">VEGFA_MESAU.ssf</a>	VEGFA_MESAU 190 aa 01-JAN-1970	P	190	<a href="#">1 .. 190</a>
<a href="#">VEGFB_MOUSE.ssf</a>	VEGFB_MOUSE 207 aa 01-JAN-1970	P	207	<a href="#">1 .. 207</a>
<a href="#">VEGFA_CANFA.ssf</a>	VEGFA_CANFA 214 aa 01-JAN-1970	P	214	<a href="#">1 .. 214</a>
<a href="#">VEGFB_BOVIN.ssf</a>	VEGFB_BOVIN 207 aa 01-JAN-1970	P	207	<a href="#">1 .. 207</a>
<a href="#">VEGFB_RAT.ssf</a>	VEGFB_RAT 207 aa 01-JAN-1970	P	207	<a href="#">1 .. 207</a>
<a href="#">VEGFD_MOUSE.ssf</a>	VEGFD_MOUSE 358 aa 01-JAN-1970	P	358	<a href="#">1 .. 358</a>
<a href="#">VEGFA_SHEEP.ssf</a>	VEGFA_SHEEP 146 aa 01-JAN-1970	P	146	<a href="#">1 .. 146</a>
<a href="#">VEGFC_RAT.ssf</a>	VEGFC_RAT 415 aa 01-JAN-1970	P	415	<a href="#">1 .. 415</a>
<a href="#">VEGFA_BOVIN.ssf</a>	VEGFA_BOVIN 190 aa 01-JAN-1970	P	190	<a href="#">1 .. 190</a>
<a href="#">VEGFA_HORSE.ssf</a>	VEGFA_HORSE 190 aa 01-JAN-1970	P	190	<a href="#">1 .. 190</a>
<a href="#">VEGFA_CAVPO.ssf</a>	VEGFA_CAVPO 164 aa 01-JAN-1970	P	164	<a href="#">1 .. 164</a>
<a href="#">VEGFA_COTJA.ssf</a>	VEGFA_COTJA 216 aa 01-JAN-1970	P	216	<a href="#">1 .. 216</a>
<a href="#">VEGFC_HUMAN.ssf</a>	VEGFC_HUMAN 419 aa 01-JAN-1970	P	419	<a href="#">1 .. 419</a>
<a href="#">VEGFD_HUMAN.ssf</a>	VEGFD_HUMAN 354 aa 01-JAN-1970	P	354	<a href="#">1 .. 354</a>
<a href="#">VEGFC_MOUSE.ssf</a>	VEGFC_MOUSE 415 aa 01-JAN-1970	P	415	<a href="#">1 .. 415</a>
<a href="#">VEGFA_PIG.ssf</a>	VEGFA_PIG 190 aa 01-JAN-1970	P	190	<a href="#">1 .. 190</a>
<a href="#">VEGFB_HUMAN.ssf</a>	VEGFB_HUMAN 207 aa 01-JAN-1970	P	207	<a href="#">1 .. 207</a>
<a href="#">VEGFA_HUMAN.ssf</a>	VEGFA_HUMAN 232 aa 01-JAN-1970	P	232	<a href="#">1 .. 232</a>

# GrowTree VegF Neighbor Joining Tree



# VegF Growth Factors

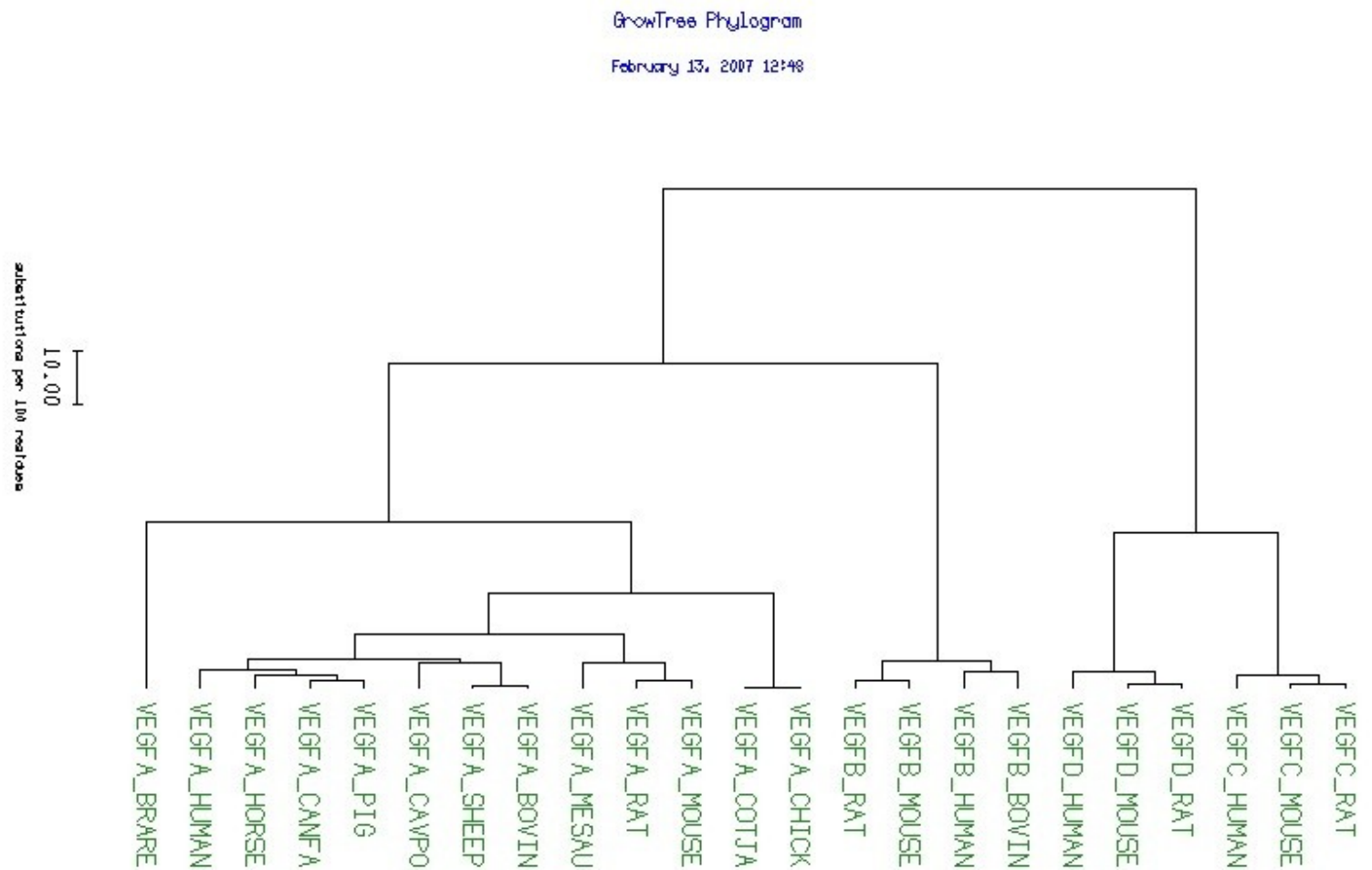
[http://en.wikipedia.org/wiki/Vascular\\_endothelial\\_growth\\_factor](http://en.wikipedia.org/wiki/Vascular_endothelial_growth_factor)

Comparison	
Type	Function
VEGF-A	<ul style="list-style-type: none"><li>▪ Angiogenesis<ul style="list-style-type: none"><li>▪ ↑ Migration of endothelial cells</li><li>▪ ↑ mitosis of endothelial cells</li><li>▪ ↑ Methane monooxygenase activity</li><li>▪ ↑ <math>\alpha v \beta 3</math> activity</li><li>▪ creation of blood vessel lumen</li><li>▪ creates fenestrations</li></ul></li><li>▪ Chemotactic for macrophages and granulocytes</li><li>▪ Vasodilation (indirectly by NO release)</li></ul>
VEGF-B	Embryonic angiogenesis
VEGF-C	Lymphangiogenesis
VEGF-D	Needed for the development of lymphatic vasculature surrounding lung bronchioles
PlGF	Important for Vasculogenesis, Also needed for angiogenesis during ischemia, inflammation, wound healing, and cancer.



# GrowTree VegF UPGMA Tree

<http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=evolution-prot>



# GrowTree VegF Alignment

<http://seqweb.stanford.edu:81/gcg-bin/analysis.cgi?program=evolution-prot>

```

101                                     150
VEGFC_RAT MRTGDTVKLA AAHYNTEILK SIDNEWRTQ CMPREVCIDV GKEFGAATNT
VEGFC_MOUSE TRTGDSVKFA AAHYNTEILK SIDNEWRTQ CMPREVCIDV GKEFGAATNT
VEGFC_HUMAN SRTEETIKFA AAHYNTEILK SIDNEWRTQ CMPREVCIDV GKEFGVATNT
VEGFD_RAT RST...RFA ATFYDTETLK VIDEEWQRTQ CSPRETCVEV ASELGKTTNT
VEGFD_MOUSE RST...RFA ATFYDTETLK VIDEEWQRTQ CSPRETCVEV ASELGKTTNT
VEGFD_HUMAN RST...RFA ATFYDIETLK VIDEEWQRTQ CSPRETCVEV ASELGKSTNT
VEGFB_BOVIN .AQAPVSQPD APGHQKKVVS WID.VYARAT CQPREVVVPL NMELMGTVAK
VEGFB_HUMAN .AQAPVSQPD APGHQKKVVS WID.VYTRAT CQPREVVVPL TVELMGTVAK
VEGFB_MOUSE .TQAPVSQFD GPSHQKKVVP WID.VYARAT CQPREVVVPL SMELMGVVK
VEGFB_RAT .TQAPVSQFD GPSHQKKVVS WID.VYARAT CQPREVVVPL SMELMGVVK
VEGFA_CHICK LSKAAPALGD GERKPNEVIK FLE.VYERSF CRTIETLVDI FQEYPDEVEY
VEGFA_COTJA LSKAAPALGD GERKPNEVIK FLE.VYERSF CRTIETLVDI FQEYPDEVEY
VEGFA_MOUSE WSQAAPTTE. GEQKSHEVIK FMD.VYQRSY CRPIETLVDI FQEYPDEIEY
VEGFA_RAT WSQAAPTTE. GEQKAHEVVK FMD.VYQRSY CRPIETLVDI FQEYPDEIEY
VEGFA_MESAU WSQAAPTTE. GEQKAHVVE FMD.VYRRSY CHPIETLVDI FQEYPDEIEY
VEGFA_BOVIN WSQAAPMAE. GGQKPHEVVK FMD.VYQRSF CRPIETLVDI FQEYPDEIEF
VEGFA_PIG WSQAAPMAE. GDQKPHEVVK FMD.VYQRSY CRPIETLVDI FQEYPDEIEY
VEGFA_HORSE WSQAAPMAE. GEHKTHEVVK FMD.VYQRSY CRPIETLVDI FQEYPDEIEY
VEGFA_CAVPO ~~~~~APMAE. GEQKPREEVK FMD.VYKRSY CRPIETLVDI FQEYPDEIEY
VEGFA_CANFA WSQAAPMA. G GEHKPHEVVK FMD.VYQRSY CRPIETLVDI FQEYPDEIEY
VEGFA_HUMAN WSQAAPMAEG GGQNHHEVVK FMD.VYQRSY CHPIETLVDI FQEYPDEIEY
VEGFA_SHEEP WSQAAPMAEG G.QKPHEVMK FMD.VYQRSF CRPIETLVDI FQEYPDEIEF
VEGFA_BRARE ...AAHIPKE GGSKNDVIP FMD.VYKSA CKTRELLVDI IQEYPDEIEH

151                                     200
VEGFC_RAT FFKPPCVSVY RCGGCCNSEG LQCMNTSTGY LSKTLFEITV PLSQGPVPVT
VEGFC_MOUSE FFKPPCVSVY RCGGCCNSEG LQCMNTSTGY LSKTLFEITV PLSQGPVPVT
VEGFC_HUMAN FFKPPCVSVY RCGGCCNSEG LQCMNTSTSY LSKTLFEITV PLSQGPVPVT
VEGFD_RAT FFKPPCVNVF RCGGCCNEES VMCMNTSTSY ISKQLFEISV PLTSVPELVP
VEGFD_MOUSE FFKPPCVNVF RCGGCCNEEG VMCMNTSTSY ISKQLFEISV PLTSVPELVP
VEGFD_HUMAN FFKPPCVNVF RCGGCCNEES LICMNTSTSY ISKQLFEISV PLTSVPELVP
VEGFB_BOVIN QLVPSCVTVQ RCGGCCPDDG LECVPTGQH Q VRMILMIQ. YPSS..QLGE
VEGFB_HUMAN QLVPSCVTVQ RCGGCCPDDG LECVPTGQH Q VRMILMIR. YPSS..QLGE
VEGFB_MOUSE QLVPSCVTVQ RCGGCCPDDG LECVPTGQH Q VRMILMIQ. YPSS..QLGE
VEGFB_RAT QLVPSCVTVQ RCGGCCPDDG LECVPTGQH Q VRMILMIQ. YPSS..QLGE
VEGFA_CHICK IFRPSCVPLM RCAGCCGDEG LECVPDVYN VTMEIARIKP HQSQ..HIAH
VEGFA_COTJA IFRPSCVPLM RCAGCCGDEG LECVPDVYN VTMEIARIKP HQSQ..HIAH
VEGFA_MOUSE IFKPSCVPLM RCAGCCNDEA LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_RAT IFKPSCVPLM RCAGCCNDEA LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_MESAU IFKPSCVPLM RCAGCCSDEA LECVPTSESN ITMQIMRVKP HQSQ..HIGE
VEGFA_BOVIN IFKPSCVPLM RCGGCCNDES LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_PIG IFKPSCVPLM RCGGCCNDEG LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_HORSE IFKPSCVPLM RCGGCCNDEG LECVPTAEFN ITMQIMRIKP HQSQ..HIGE
VEGFA_CAVPO IFKPSCVPLM RCGGCCNDES LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_CANFA IFKPSCVPLM RCGGCCNDEG LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_HUMAN IFKPSCVPLM RCGGCCNDEG LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_SHEEP IFKPSCVPLM RCGGCCNDES LECVPTSESN ITMQIMRIKP HQSQ..HIGE
VEGFA_BRARE TYIPSCVPLM RCAGCCNDEA LECVPTETRN VTMEVLRVKQ RVSQ..HNFQ
```