

Reconstructing Phylogenetic Trees

September 18th, 2008

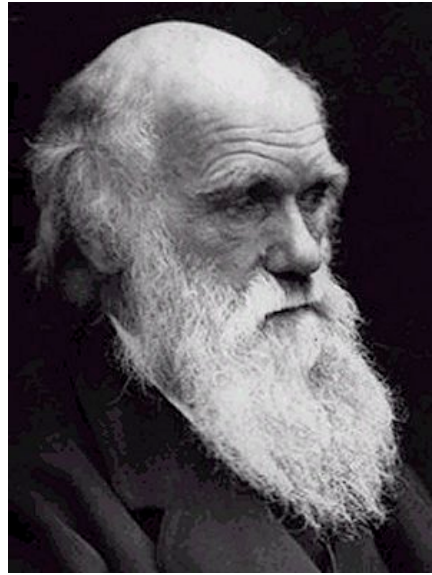


UNIVERSITY *of* ROCHESTER

Systematic Methods Through Time



Carolus Linneaus



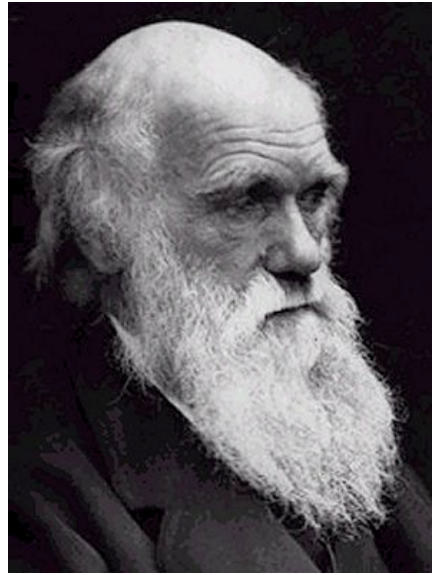
Charles Darwin



Systematic Methods Through Time



Carolus Linneaus

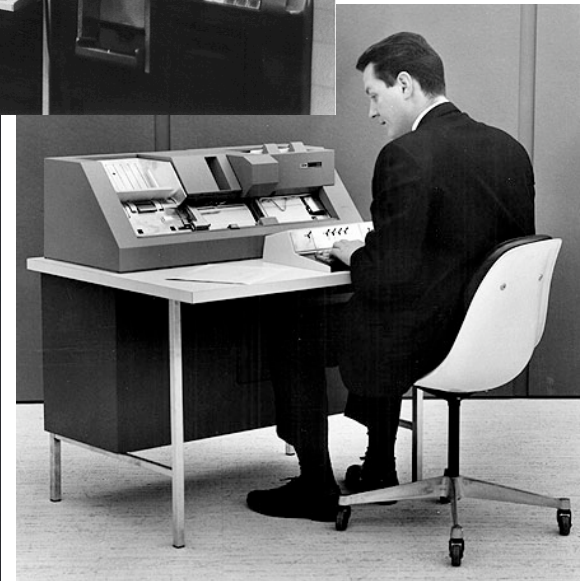
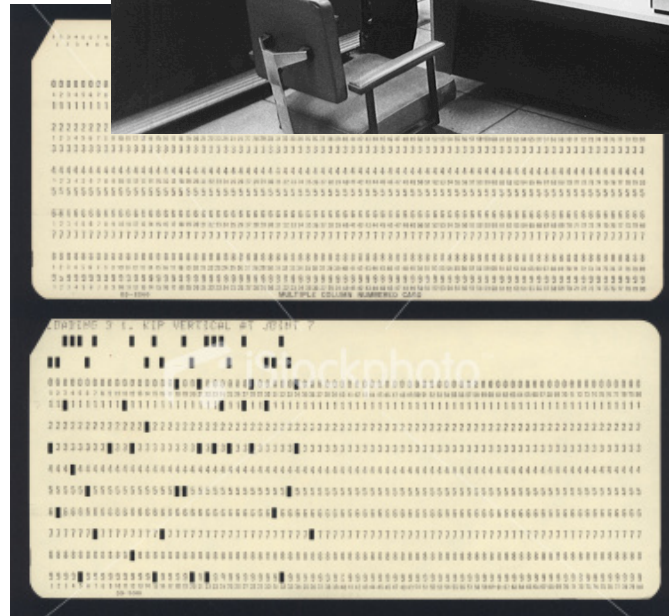


Charles Darwin



Computing Revolution

- Late 1950s & 60s.
- Growing availability of core computing facilities.



<http://www1.istockphoto.com/>

http://www-03.ibm.com/ibm/history/exhibits/vintage/vintage_4506VV4002.html



UNIVERSITY of ROCHESTER

A QUANTITATIVE APPROACH TO A PROBLEM
IN CLASSIFICATION ¹

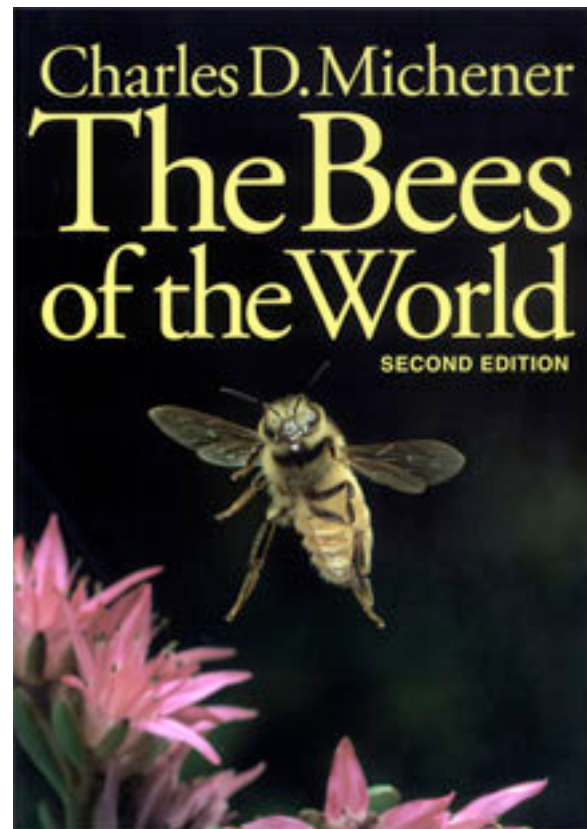
CHARLES D. MICHENER AND ROBERT R. SOKAL ²

Department of Entomology, University of Kansas, Lawrence

Received October 10, 1956



Charles D. Michener



Robert R. Sokal



Tree Reconstruction I: Intro. & Distance Measures

- The challenge of tree reconstruction.
- Phenetics and an introduction to tree reconstruction methods.
- Discrete v. distance measures.
- Clustering v. optimality searches.
- Tree building algorithms.



How Many Trees?

Taxa	Unrooted Trees	Rooted Trees
4	3	15
8	10,395	135,135
10	2,027,025	34,459,425
22	3×10^{23}	
50	3×10^{74} *	

* More trees than there are atoms in the universe.



Reconstructing Trees

- The challenge of tree reconstruction.
 - **Lots of possibilities.**
- Phenetics and an introduction to tree reconstruction methods.
 - Discrete v. distance measures.
 - Clustering v. optimality searches.
 - Tree building algorithms.



Discrete Data

discrete
sites

	1	2	3	4	5	6	7
1	T	T	A	T	T	A	A
2	A	A	T	T	T	A	A
3	A	A	A	A	A	T	A
4	A	A	A	A	A	A	T

sequences

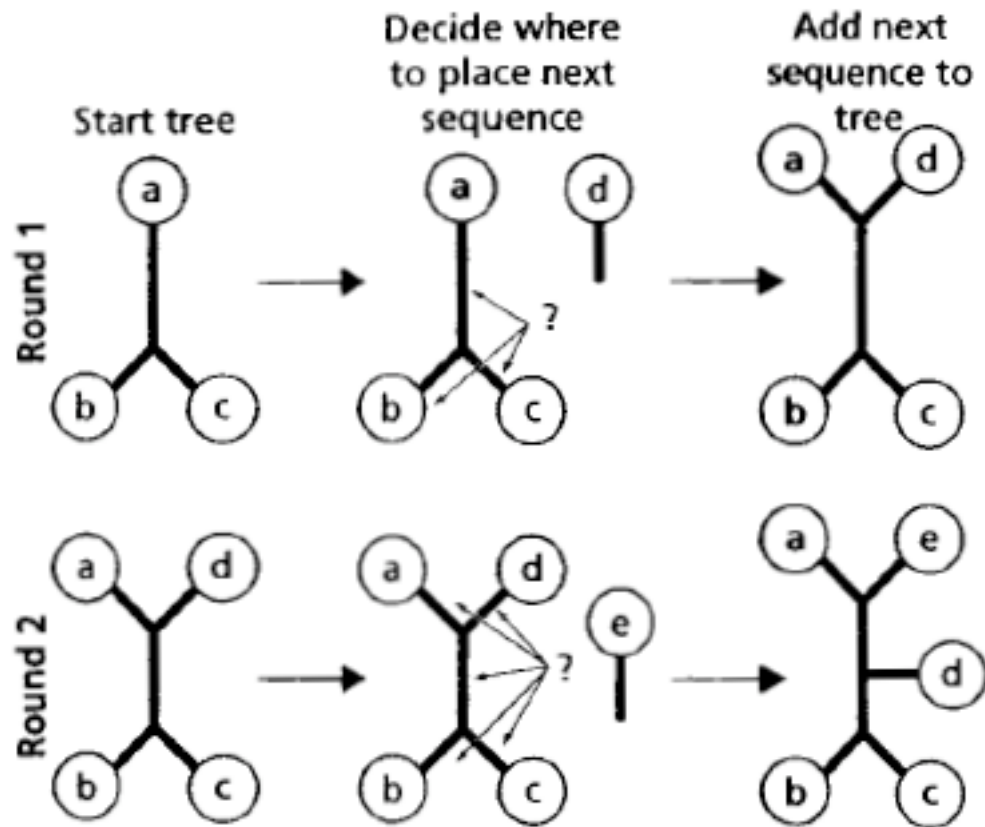


Discrete Data

		sequences						
		sites						
		1	2	3	4	5	6	7
sequences	1	T	T	A	T	T	A	A
	2	A	A	T	T	T	A	A
	3	A	A	A	A	A	T	A
	4	A	A	A	A	A	A	T

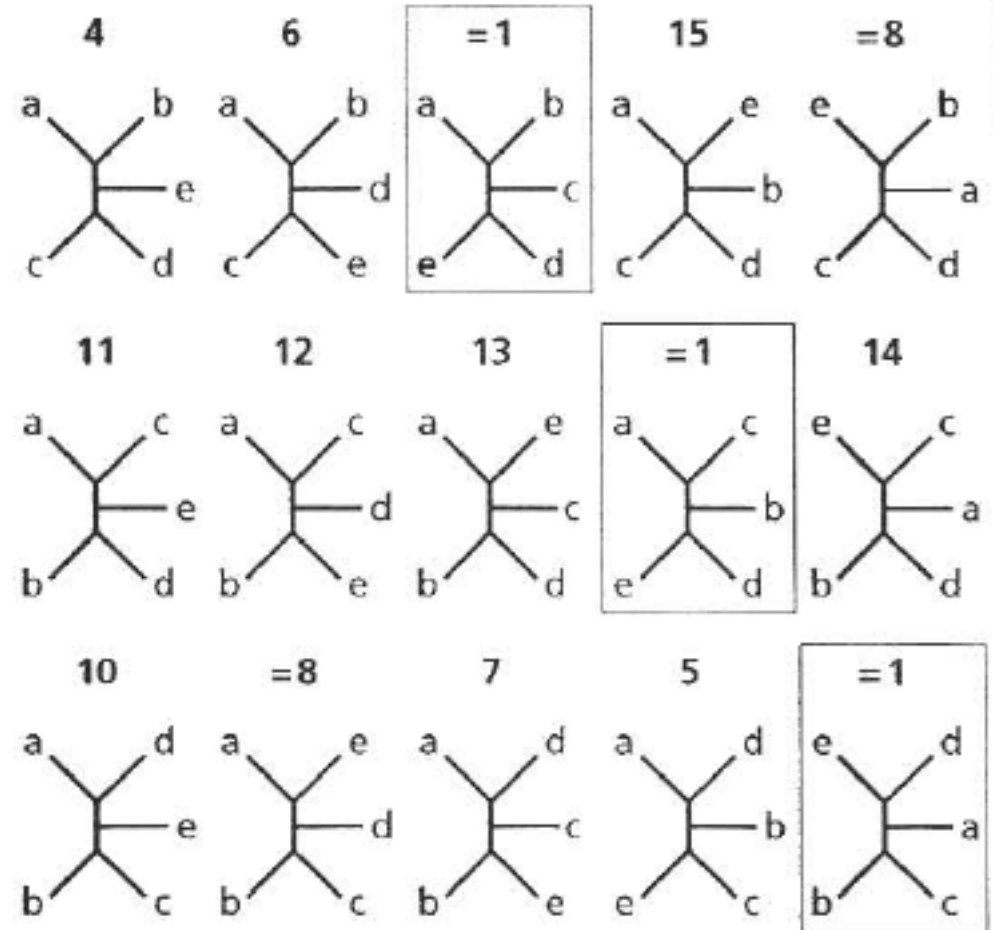
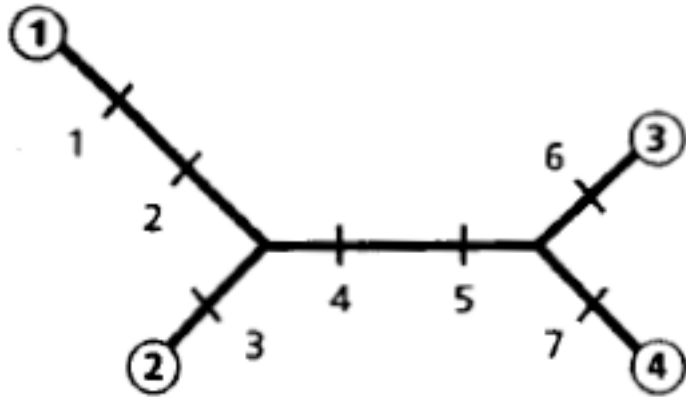


Clustering Methods



Optimality Criterion

Parsimony



NP-Completeness

- Non-deterministic polynomial.
- Impossible to guarantee optimal tree for even relatively modest number of sequences.
- Use of heuristic methods.



Available Methods

		Type of data	
		Distances	Nucleotide sites
Tree-building method	Clustering algorithm	UPGMA Neighbour joining	
	Optimality criterion	Minimum evolution	Maximum parsimony Maximum likelihood



Distance Clustering Methods

- The phenetic approach.
- Two common algorithms for tree reconstruction.
- UPGMA & Neighbor joining.



Phenetics



- Also called **numerical taxonomy** because of emphasis on data.
- Relationships inferred based on overall similarity.



Phenetics



UPGMA

- **UPGMA** - Unweighted pair group method with arithmetic means (Sokal & Michener 1958).
- Remarkably simple and straightforward.
- Can be used with many types of distances (molecular, morphological, etc.).

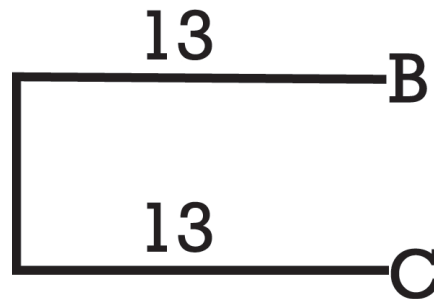


UPGMA Algorithm

	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0



UPGMA Algorithm



	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0



UPGMA Algorithm

	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0



	A	B/C	D	E
A	0	40	51	50
B/C	40	0	38	36.5
D	51	38	0	44
E	50	36.5	44	0



UPGMA Algorithm

	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0

$A \text{ to } B/C = (A \text{ to } B + A \text{ to } C)/2$ $40 = (32 + 48)/2$

	A	B/C	D	E
A	0	40	51	50
B/C	40	0	38	36.5
D	51	38	0	44
E	50	36.5	44	0



UPGMA Algorithm

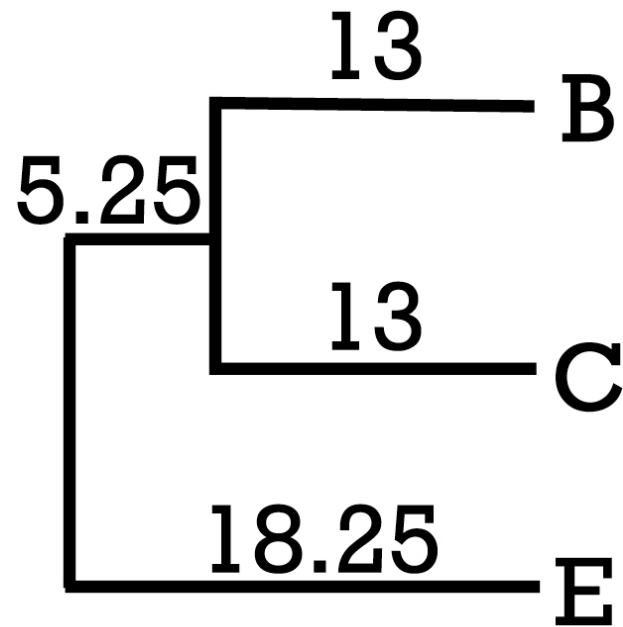
	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0

$A \text{ to } B/C = (A \text{ to } B + A \text{ to } C)/2$ $40 = (32 + 48)/2$

	A	B/C	D	E
A	0	40	51	50
B/C	40	0	38	36.5
D	51	38	0	44
E	50	36.5	44	0



UPGMA Algorithm

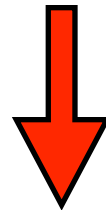


	A	B/C	D	E
A	0	40	51	50
B/C	40	0	38	36.5
D	51	38	0	44
E	50	36.5	44	0



UPGMA Algorithm

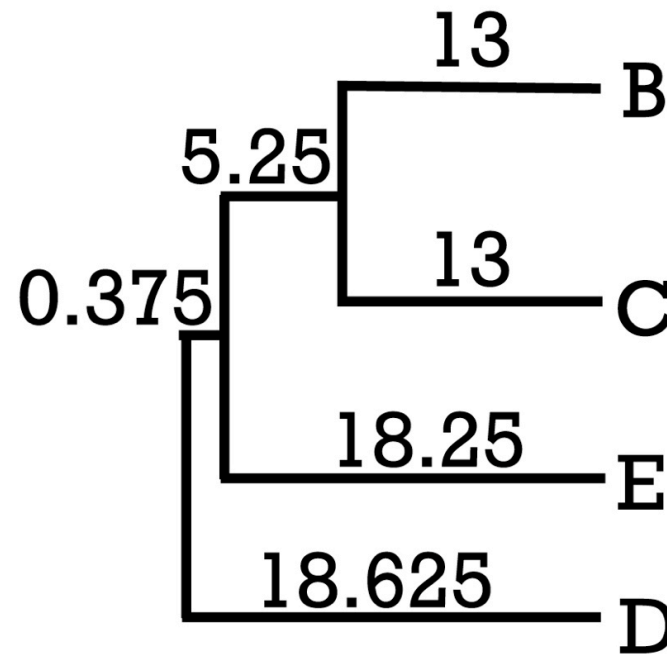
	A	B/C	D	E
A	0	40	51	50
B/C	40	0	38	36.5
D	51	38	0	44
E	50	36.5	44	0



	A	B/C/E	D
A	0	45	51
B/C/E	45	0	37.25
D	51	37.25	0



UPGMA Algorithm



	A	B/C/E	D
A	0	45	51
B/C/E	45	0	37.25
D	51	37.25	0



UPGMA Algorithm

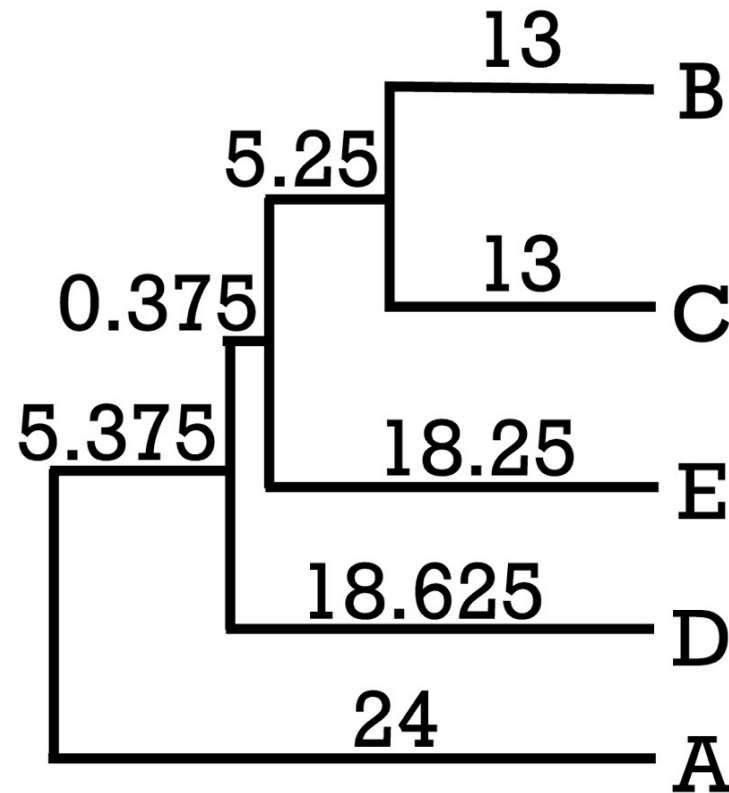
	A	B/C/E	D
A	0	45	51
B/C/E	45	0	37.25
D	51	37.25	0



	A	B/C/E/D
A	0	48
B/C/E/D	48	0



UPGMA Algorithm



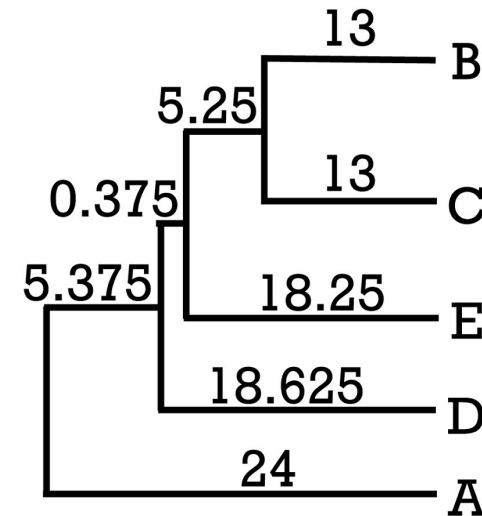
	A	B/C/E/D
A	0	48
B/C/E/D	48	0



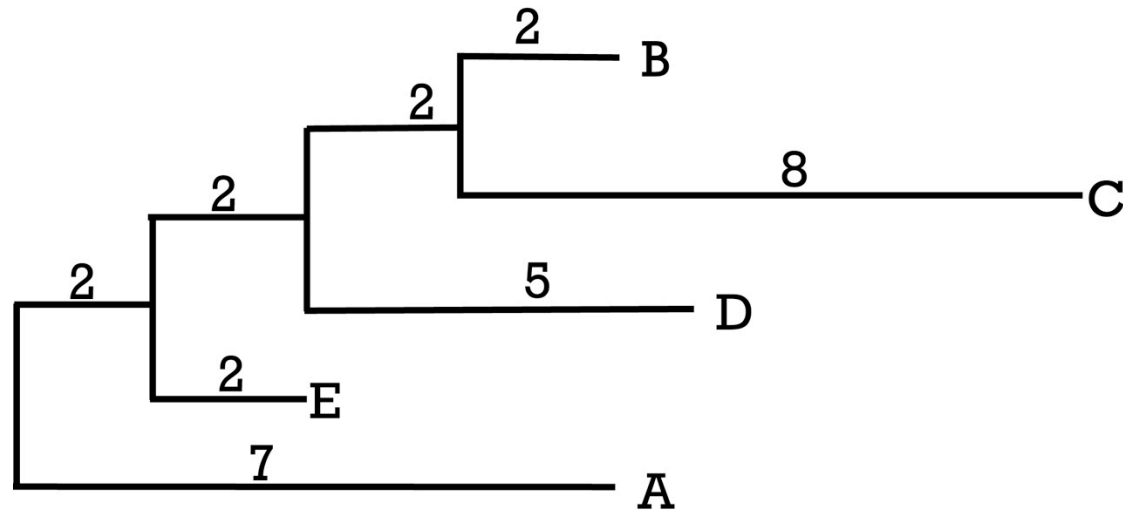
Distance Methods

- Unweighted pair group means (Sokal & Michener 1958).
 - Easy to understand and implement even on large datasets.
 - Assumes molecular clock and prone to error if this is not the case.

	A	B	C	D	E
A	0	32	48	51	50
B	32	0	26	34	29
C	48	26	0	42	44
D	51	34	42	0	44
E	50	29	44	44	0



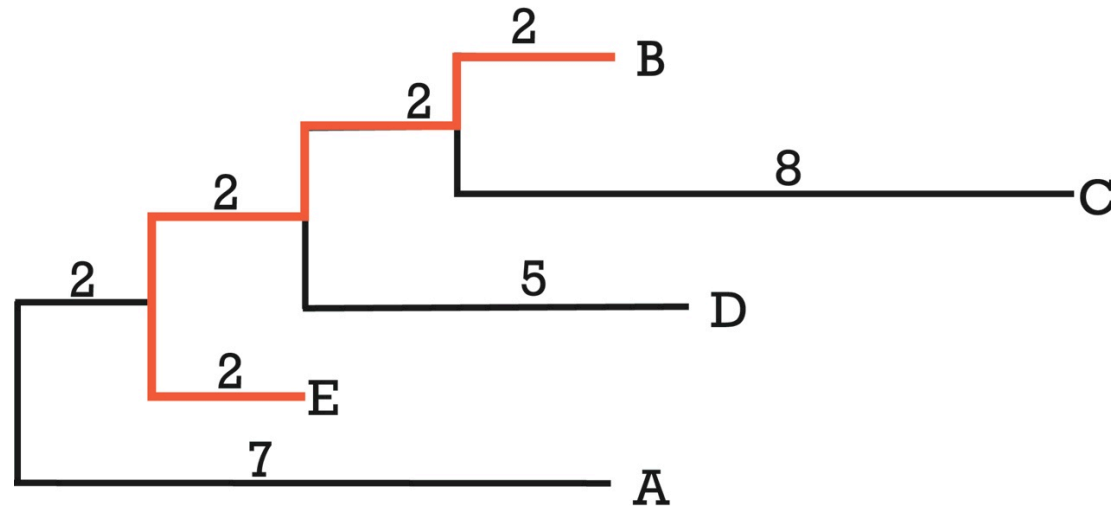
UPGMA Problems



	A	B	C	D	E
A	0	15	21	16	11
B	15	0	10	9	8
C	21	10	0	15	14
D	16	9	15	0	9
E	11	8	14	9	0



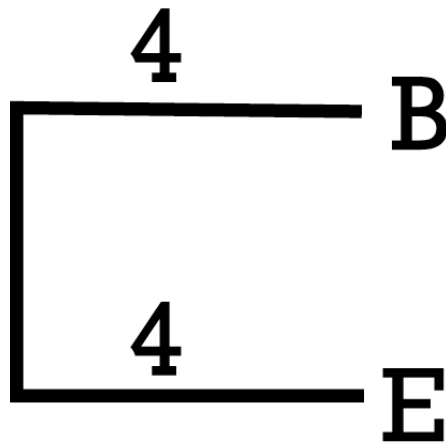
UPGMA Problems



	A	B	C	D	E
A	0	15	21	16	11
B	15	0	10	9	8
C	21	10	0	15	14
D	16	9	15	0	9
E	11	8	14	9	0



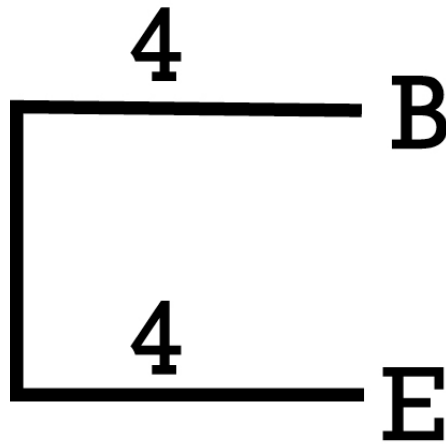
UPGMA Problems



	A	B	C	D	E
A	0	15	21	16	11
B	15	0	10	9	8
C	21	10	0	15	14
D	16	9	15	0	9
E	11	8	14	9	0



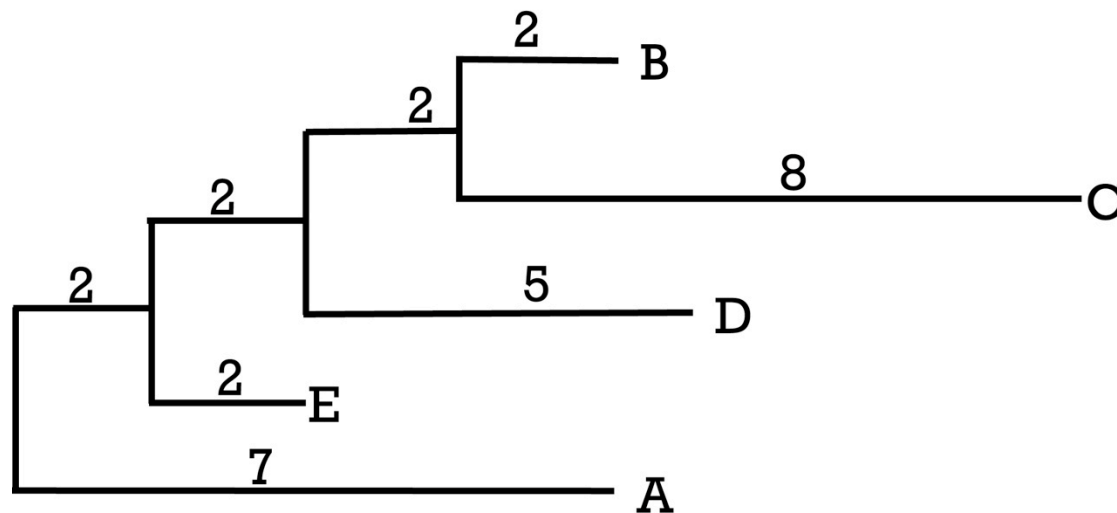
UPGMA Problems



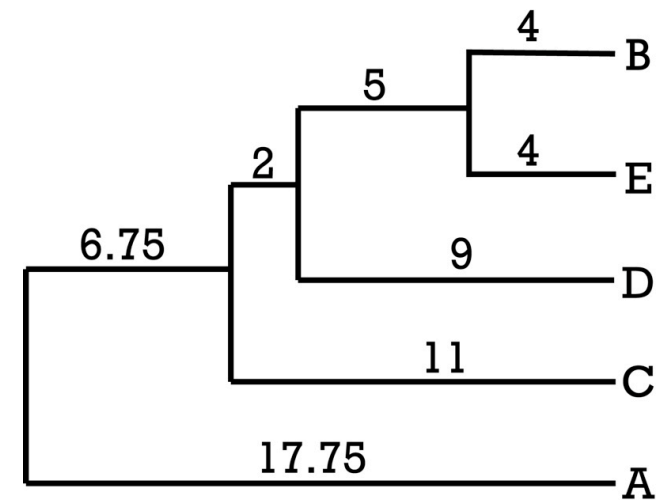
	A	B/E	C	D
A	0	13	21	16
B/E	13	0	12	9
C	21	10	0	15
D	16	9	15	0



UPGMA Problems



Real Tree

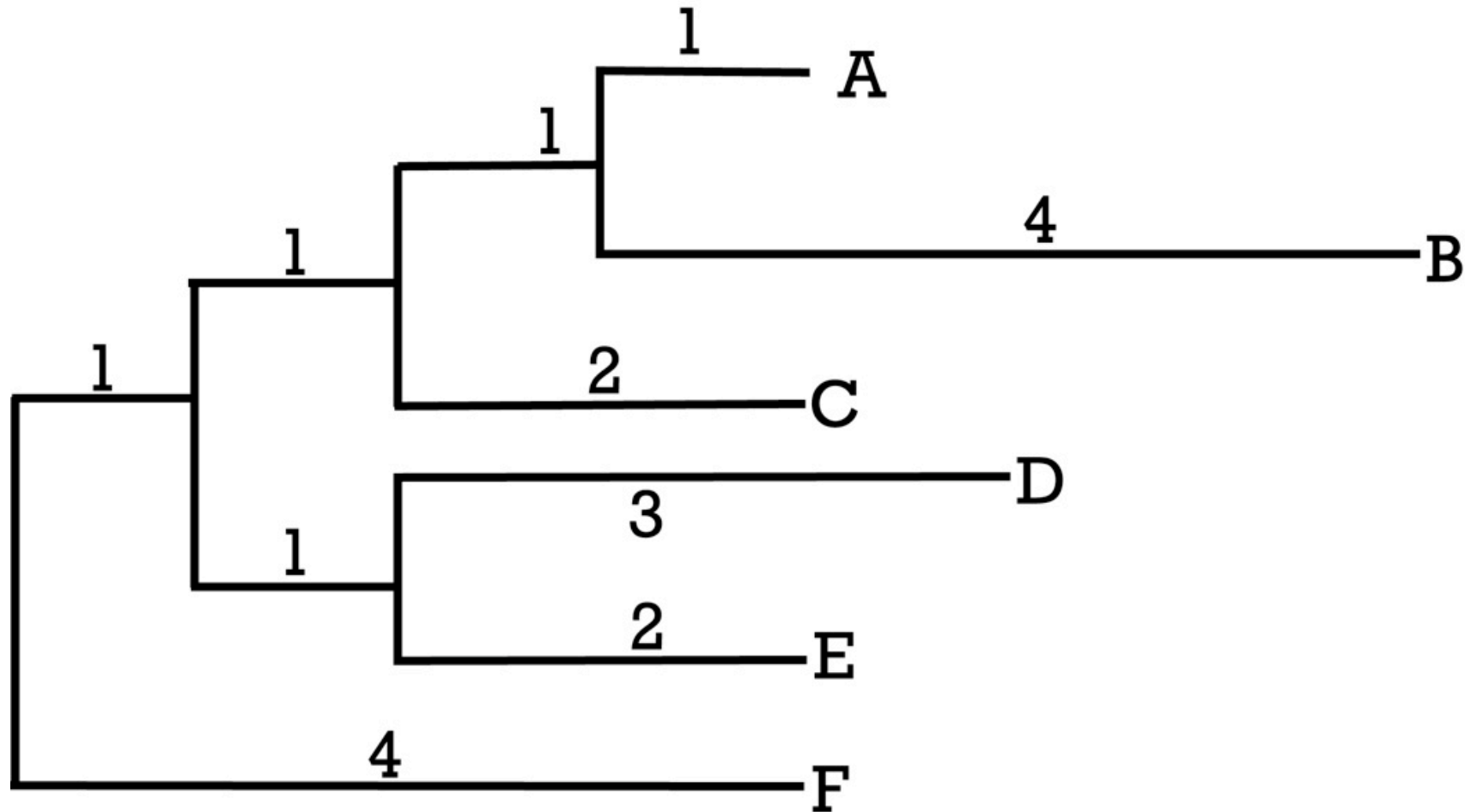


UPGMA Tree



Neighbor-Joining

Saitou & Nei 1987



<http://www.icp.be/~opperd/private/neighbor.html>



UNIVERSITY of ROCHESTER

Advantages of Distance Clustering Methods for Contemporary Systematics

- Computationally easy & quick.
- Typically give a single tree.
- Provide starting point for more sophisticated analyses.



Disadvantages of Distance Clustering Methods & Phenetics for Contemporary Systematics

- Problems with distance measures.
- Problems with clustering algorithms.
- Problems with character choice.



Problems with Phenetics



Problems with Phenetics



Types of Similarity

Homology v. Analogy

*Similar due to
inheritance*



*Similar due to...
uh...other factors*

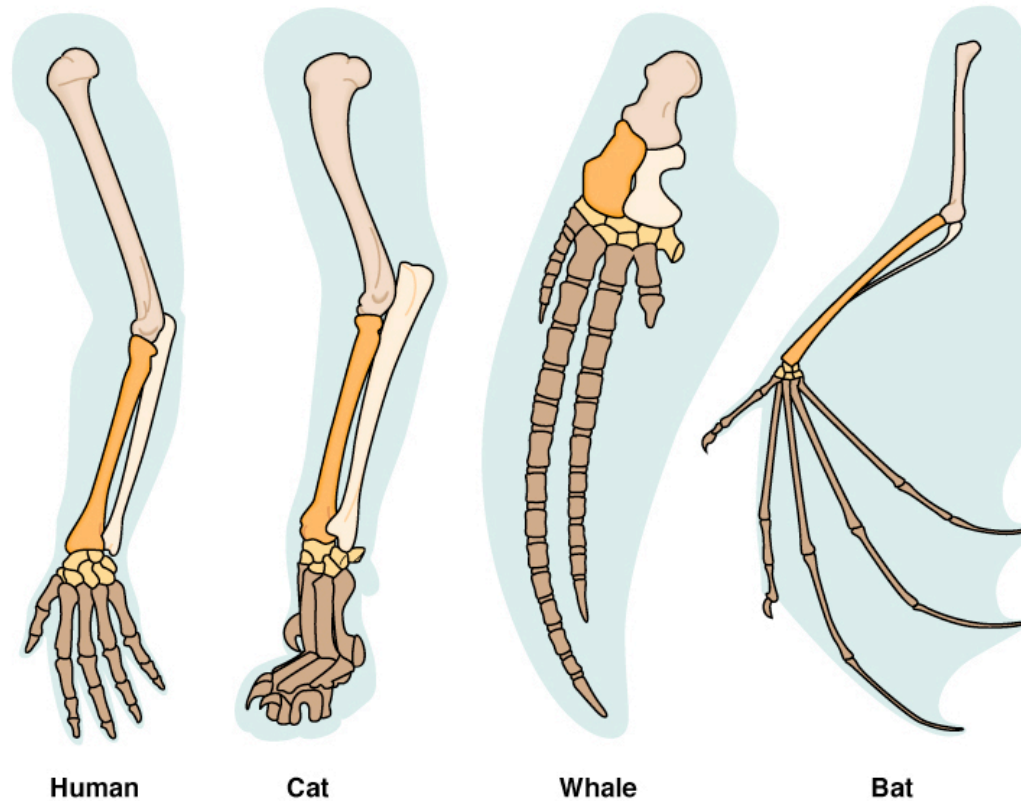


<http://evolution.berkeley.edu/evolibrary/>



Homology

- Features shared due to inheritance from a common ancestor



©1999 Addison Wesley Longman, Inc.

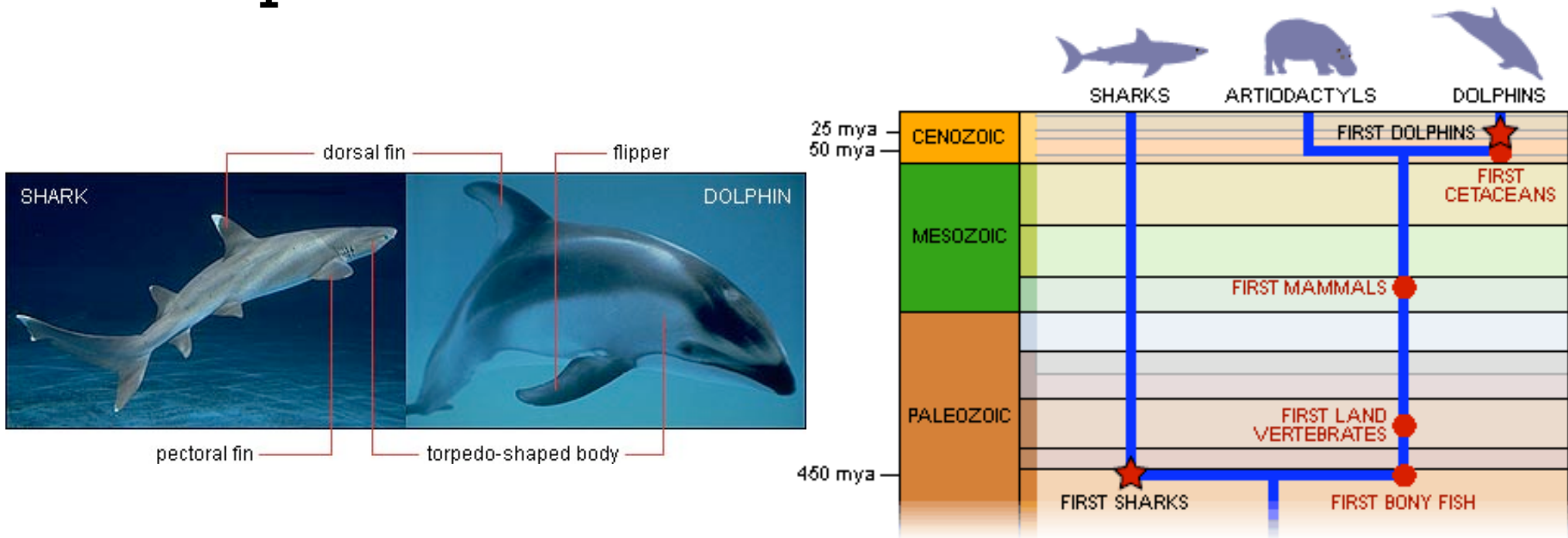
Vertebrate Forelimbs

<http://www.mun.ca/biology/scarr/>



Analogies

- Features shared due convergence or parallelism.

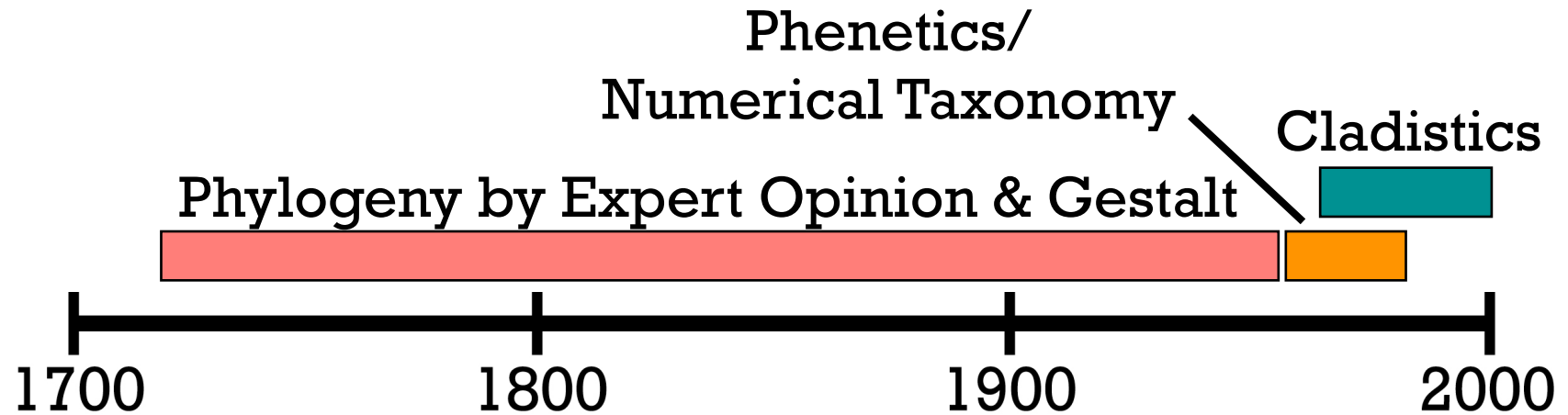


Swimming body form of vertebrates

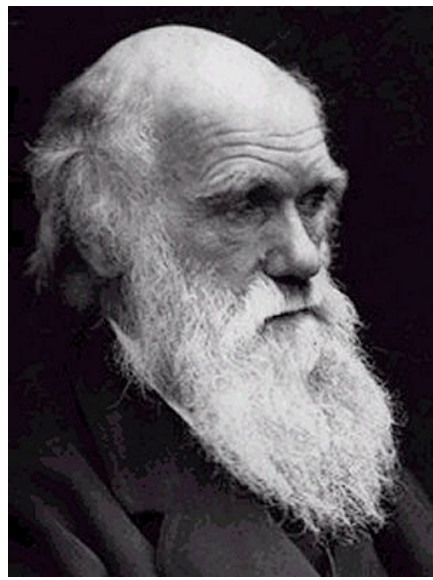
<http://evolution.berkeley.edu/evolibrary/>



Timeline



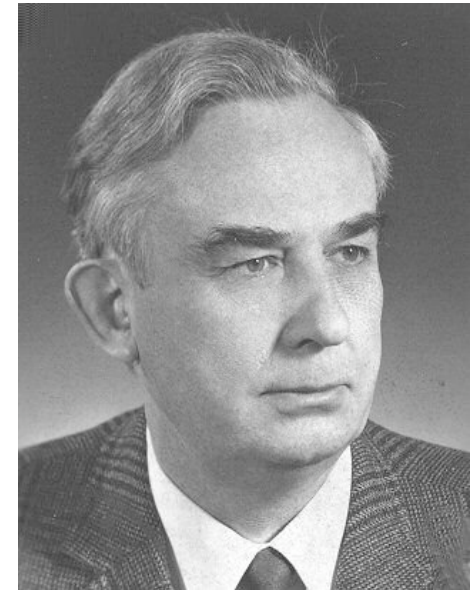
Carolus Linneaus



Charles Darwin



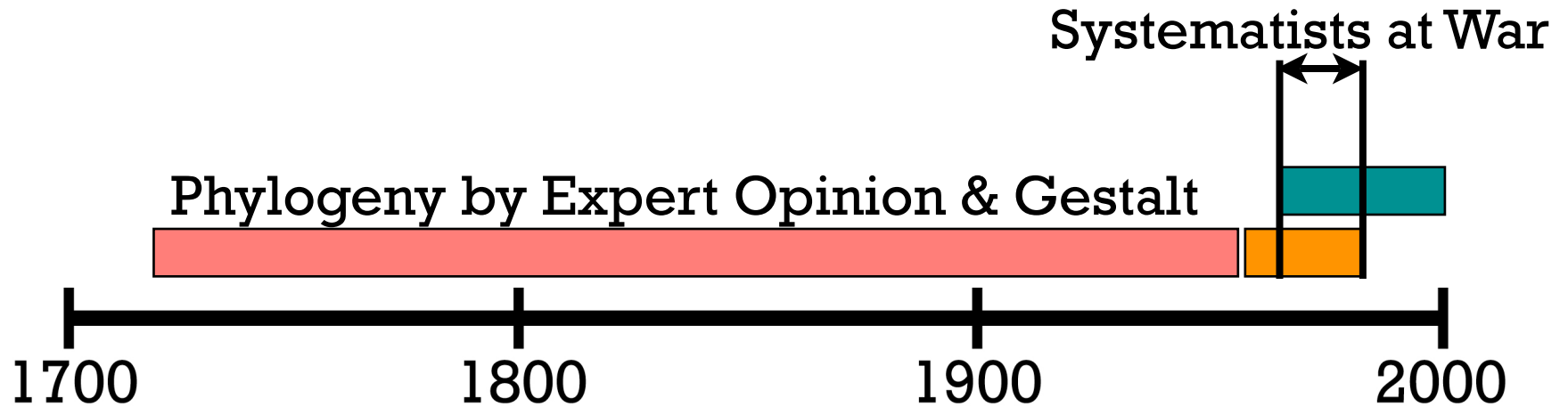
Robert R.



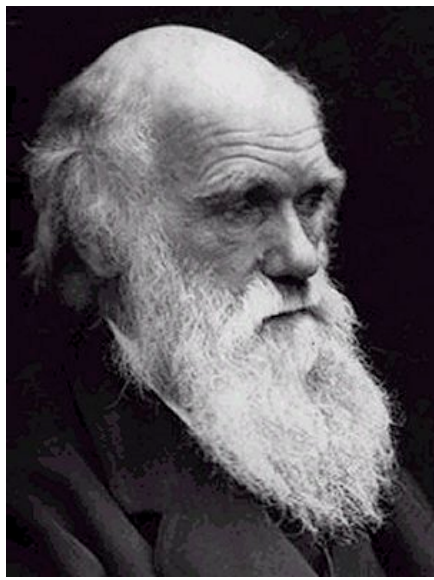
Willi Hennig



Timeline



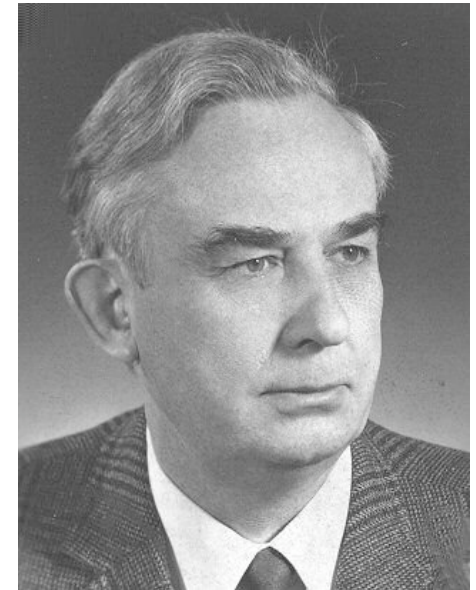
Carolus Linneaus



Charles Darwin



Robert R.

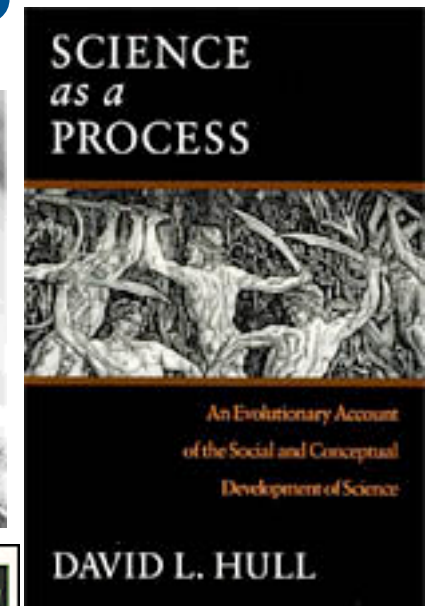


Willi Hennig



War of the Systematists

- Intense battles in the 1960s and 1970s.
- Considered in detail by the philosopher David Hull.



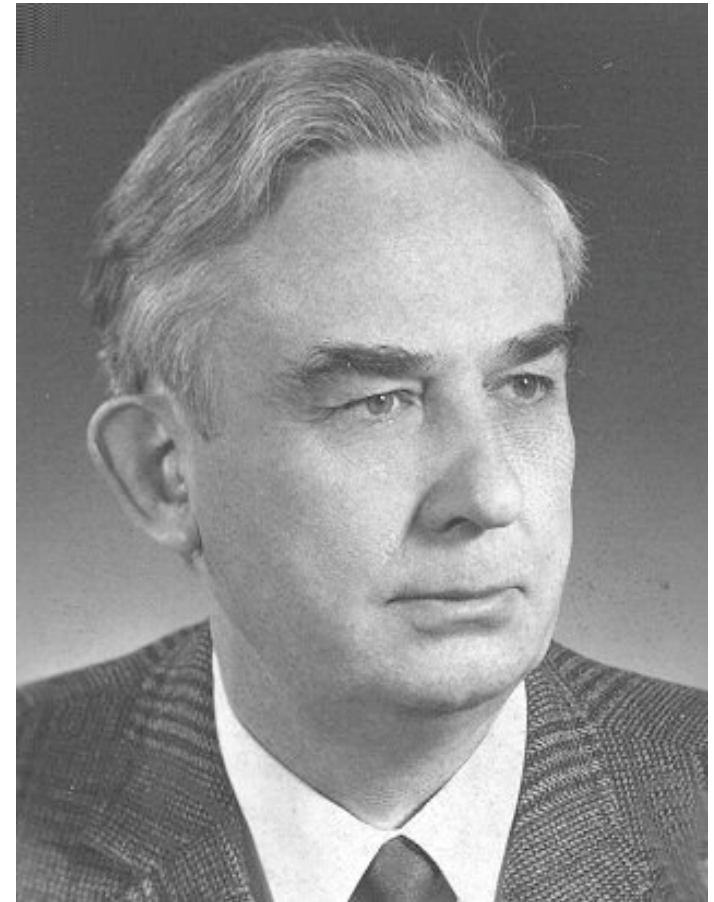
The Cladistic Revolution

- **Cladistics:** Hierarchical classification of species based on evolutionary ancestry.



The Cladistic Revolution

- Willi Hennig - the founder of cladistics.
- Phylogenetic Systematics
 - 1950 (German edition).
 - 1966 (English translation).

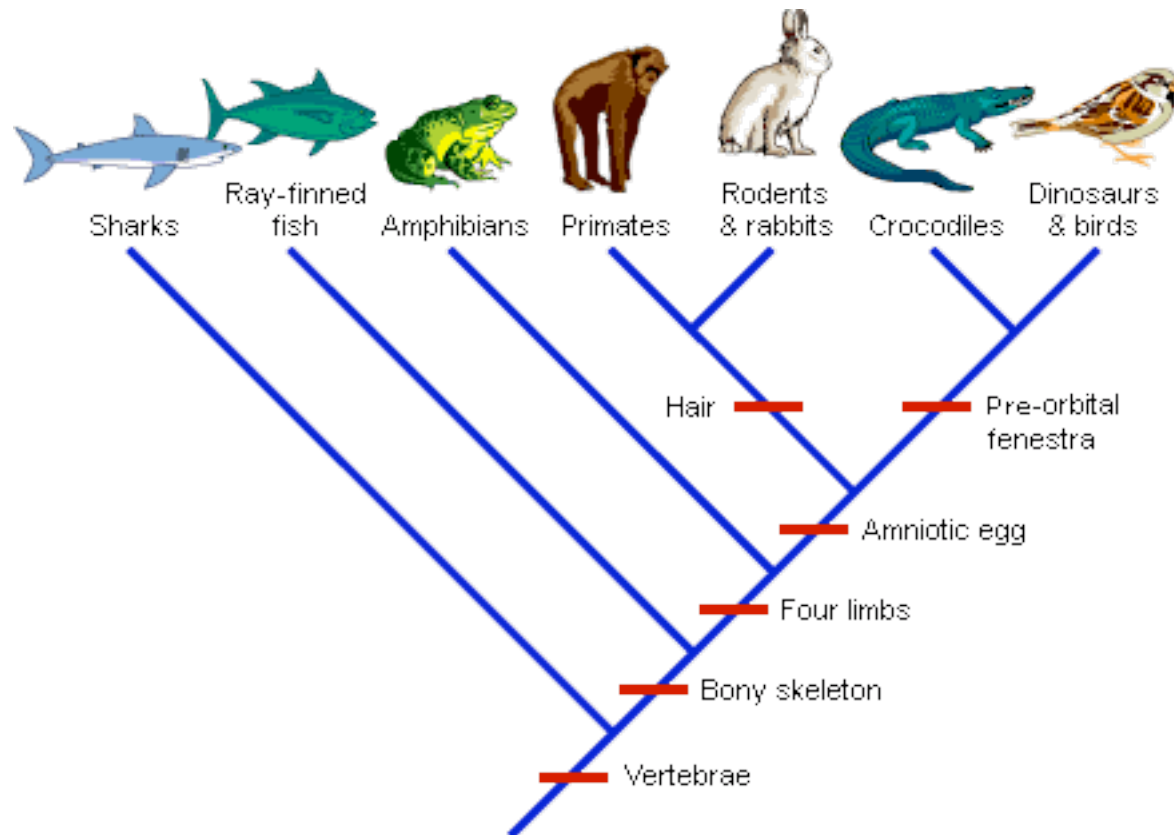


Willi Hennig



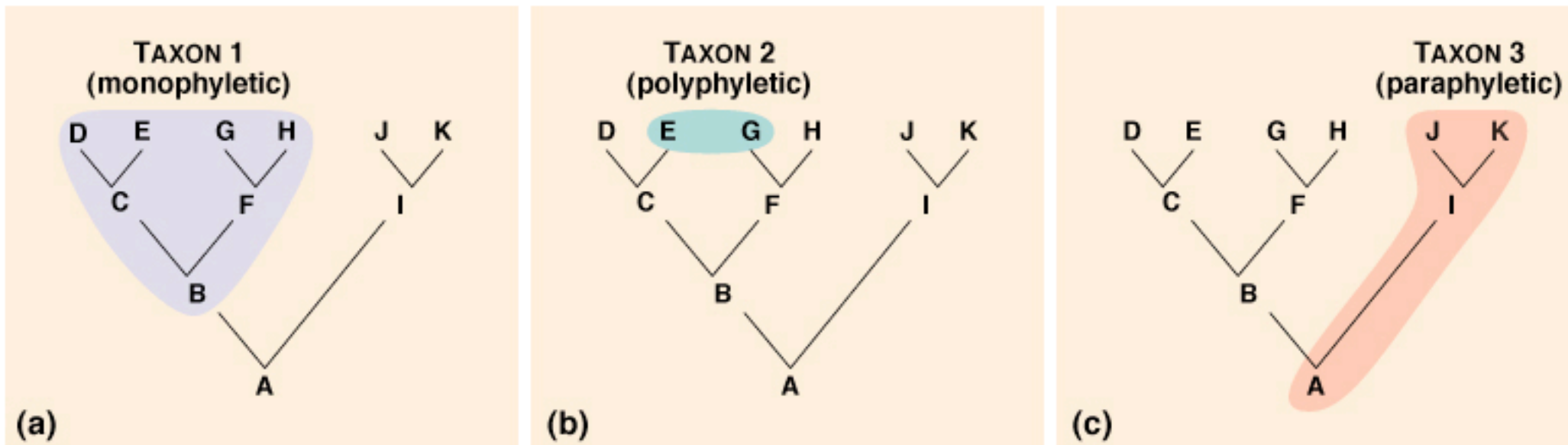
The Cladogram

- **Cladogram:** A branching diagram that depicts a set of hypothesized evolutionary relationships (i.e., a phylogeny).



Clades & Cladists

- **Clade:** A monophyletic group.



©1999 Addison Wesley Longman, Inc.



Clades & Cladists

- **Cladist:** One who uses cladistics to infer phylogenetic relationships.



Arnold Kluge



Me



Joel Cracraft



Vicki Funk



Steve Poe



Kevin De Queiroz

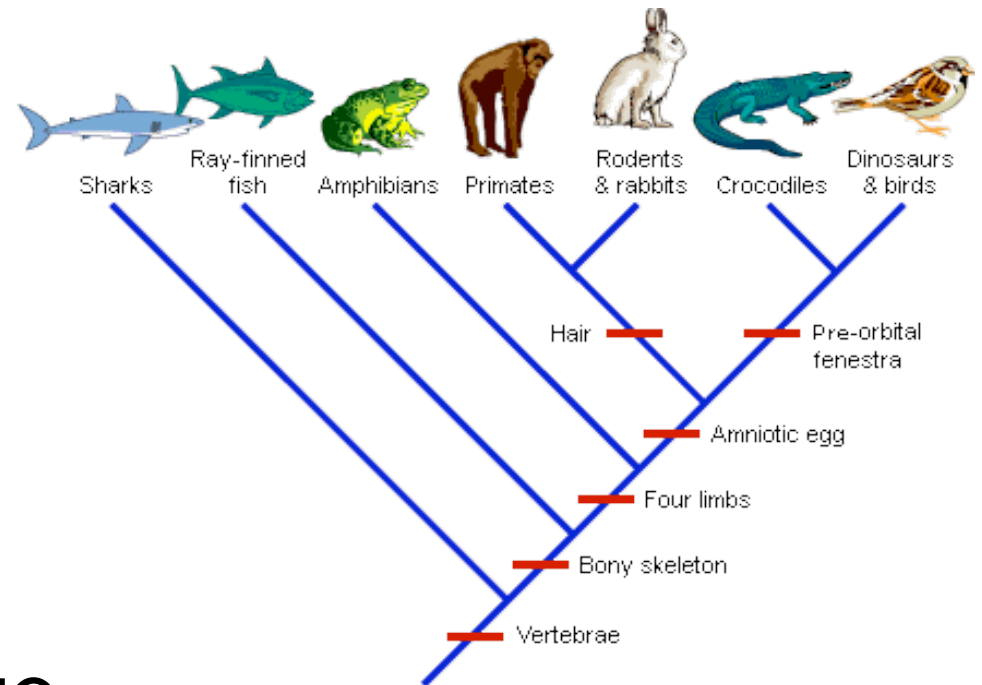


Norman Platnick

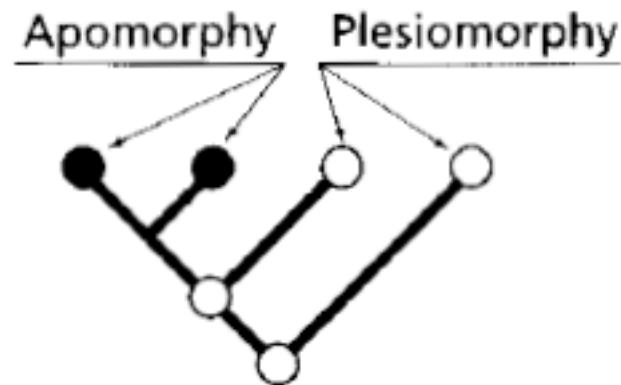


Cladistic Principles

- I. Use characters with special evolutionary significance.
- II. Reconstruction of phylogeny based on most parsimonious reconstruction of these characters.



Cladistic Terms for Characters



- Instead of being 'primitive' or 'advanced', characters are 'plesiomorphic' and 'apomorphic.'
- **Apomorphy:** A derived character.
- **Plesiomorphy:** An ancestral or primitive character.



Cladistic Terms for Characters

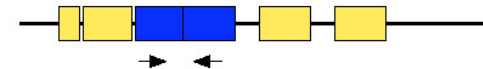
Synapomorphy



Chloroplast gene rpl2 (with intron) and flanking genes



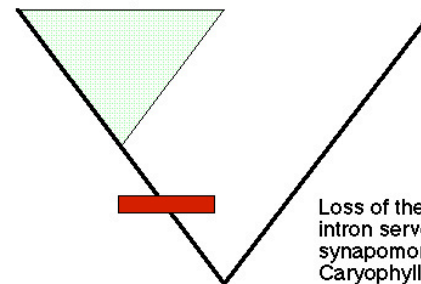
Chloroplast gene rpl2 (without intron) and flanking genes



Agarose Gel Electrophoresis separates PCR fragments based on size



Caryophyllales



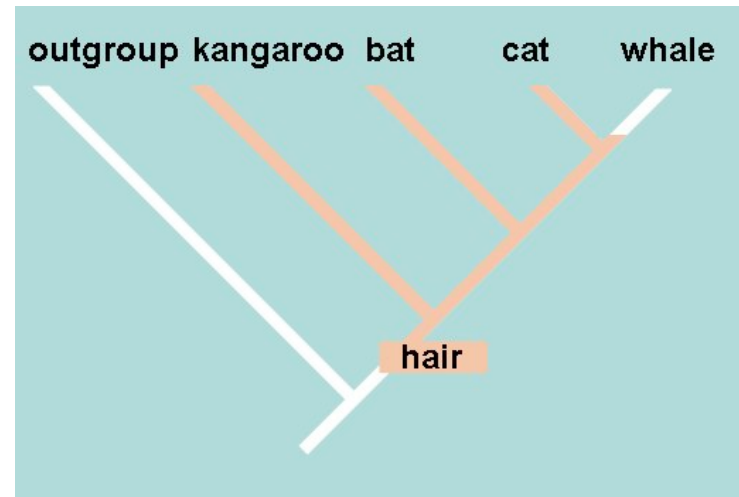
Loss of the rpl2
intron serves as a
synapomorphy for
Caryophyllales

- **Synapomorphy:** An apomorphy shared among taxa due to common ancestry.



Cladistic Terms for Characters

Synapomorphy



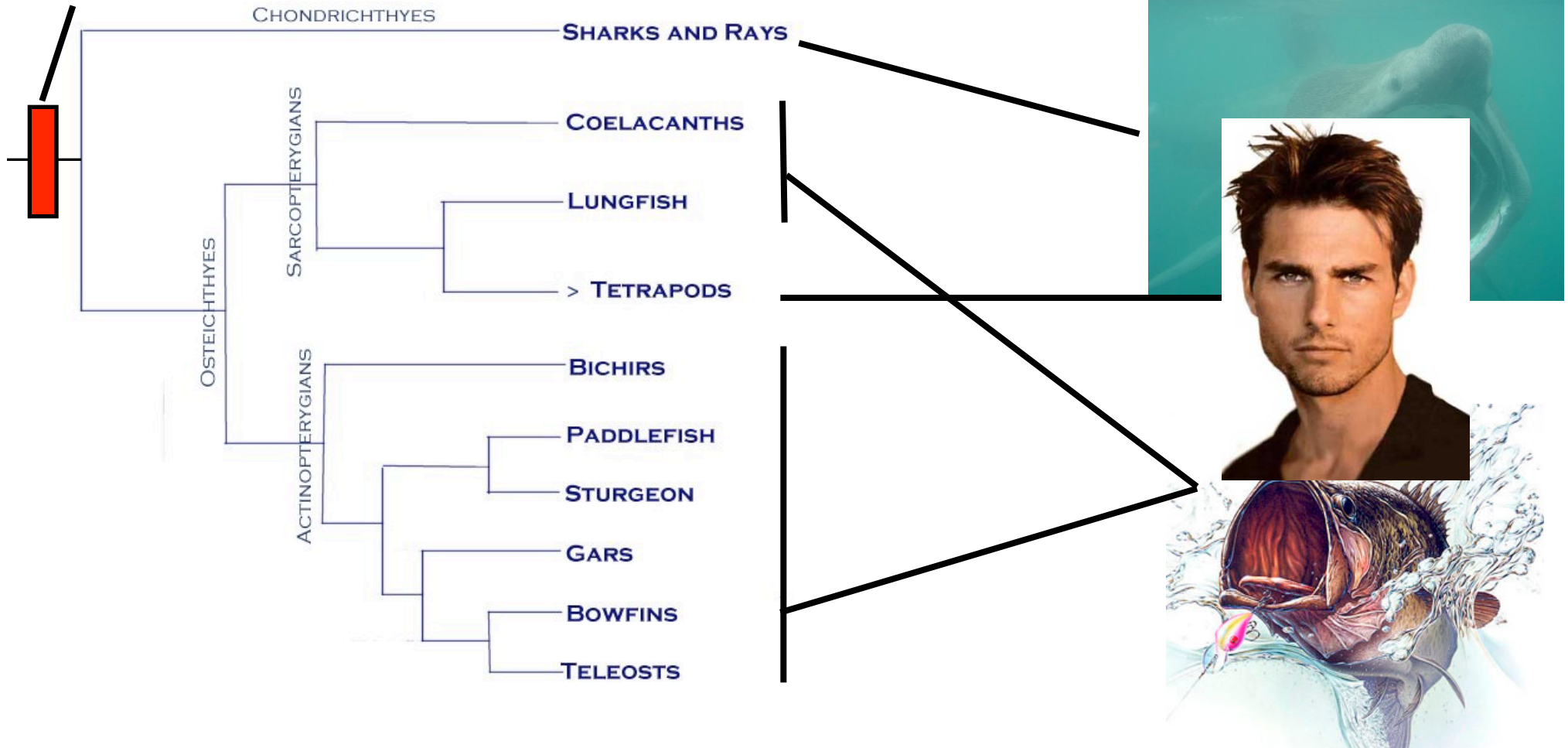
http://www.mun.ca/biology/scarr/Hair_symplesiomorphy.htm

- **Synapomorphy:** An apomorphy shared among taxa due to common ancestry.
- **Symplesiomorphy:** A plesiomorphy shared among taxa, but which predates their common ancestor.



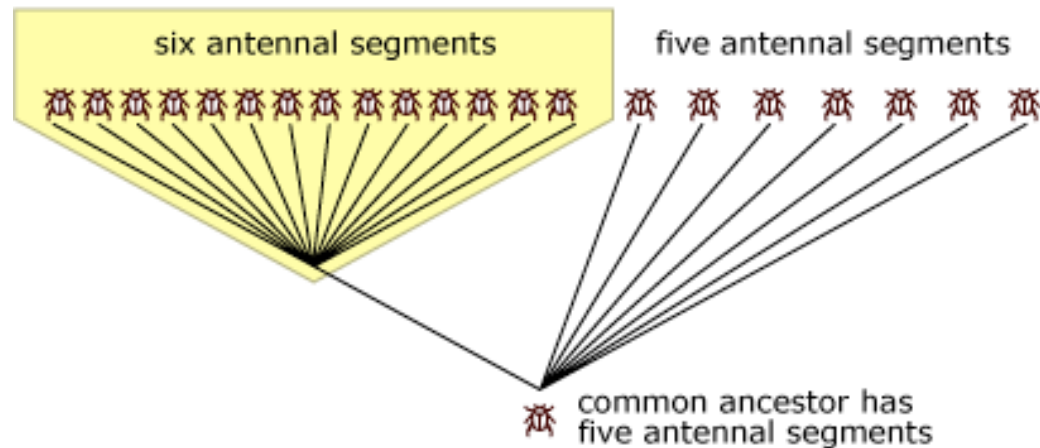
Classification with Sympleiomorphy

Gill Breathing

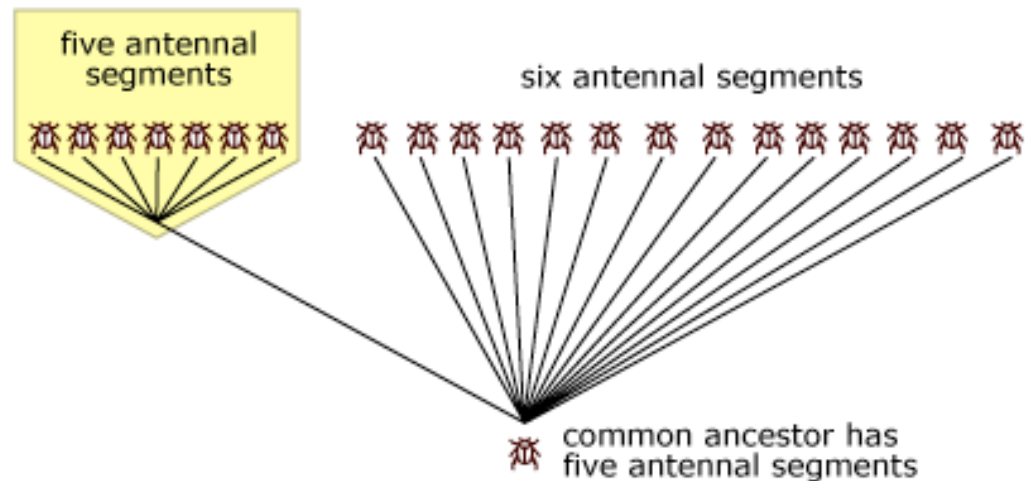


Cladistic Terms

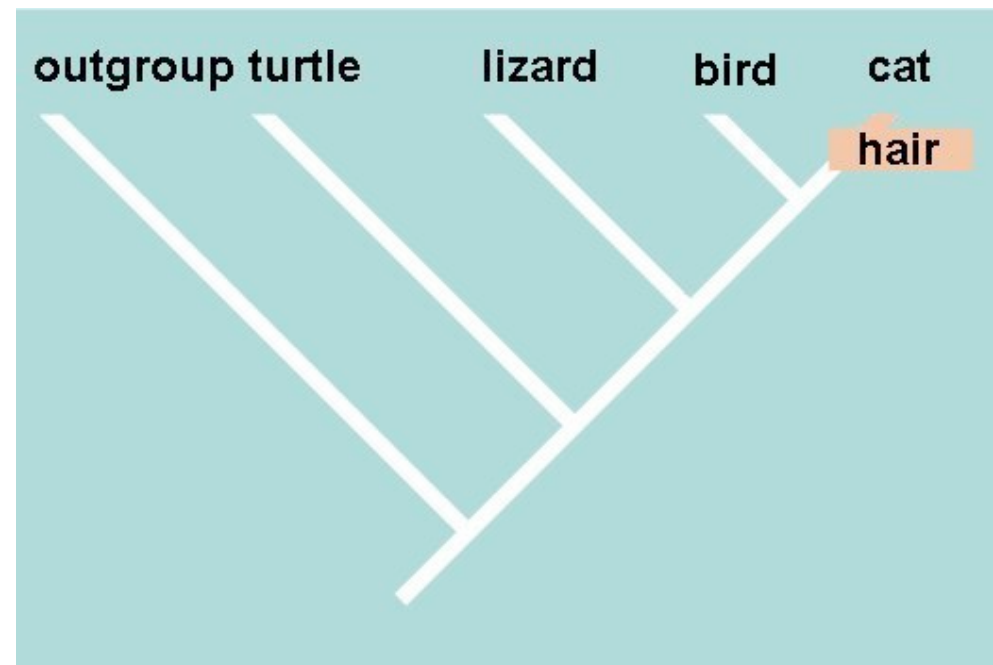
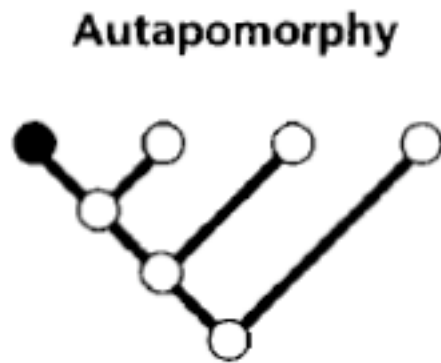
Correct: grouped by synapomorphy



Incorrect: grouped by symplesiomorphy



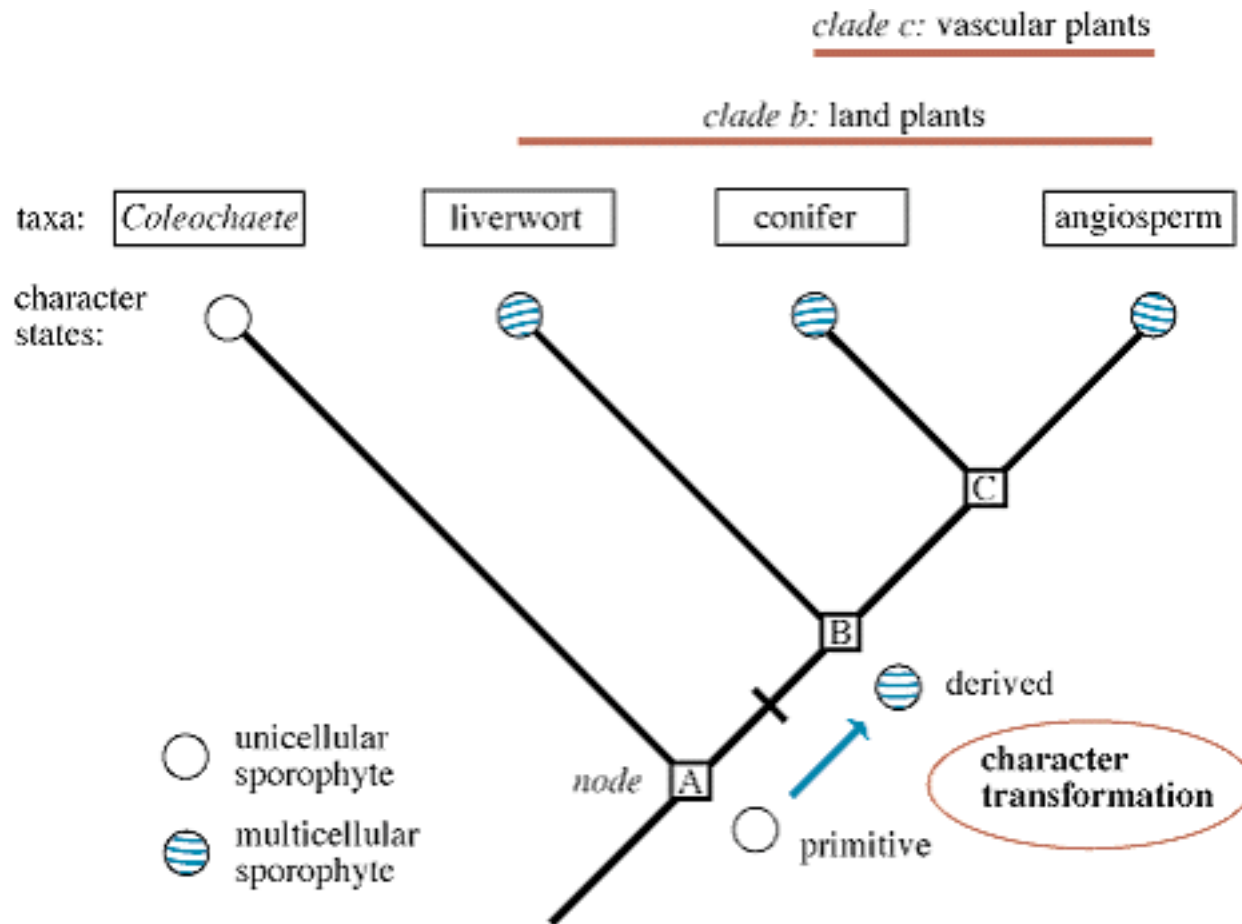
Cladistic Terms



- **Autapomorphy:** A derived trait that is unique to one group.



Hierarchy of Cladistic Characters



<http://www.ucmp.berkeley.edu/IB181/VPL/Phylo/Phylo2.html>



Identifying Synapomorphies

- Similar fundamental structure.
- Same relation to surrounding characters.
- Similar development.
- Product of natural selection?



Methods

		Type of data	
		Distances	Nucleotide sites
Tree-building method	Clustering algorithm	UPGMA Neighbour joining	
	Optimality criterion	Minimum evolution	Maximum parsimony Maximum likelihood



Methods

		Type of data	
		Distances	Nucleotide sites
Tree-building method	Clustering algorithm	UPGMA Neighbour joining	
	Optimality criterion	Minimum evolution	Maximum parsimony Maximum likelihood



Parsimony

- Preference for the tree with the “minimum net amount of evolution” (Edwards and Cavalli-Sforza 1963)
- Seen as an extension of Occam’s razor.
 - *entia non sunt multiplicanda praeter necessitatem*
 - “entities should not be multiplied beyond necessity” or “all things being equal, the simplest solution tends to be the best one.”



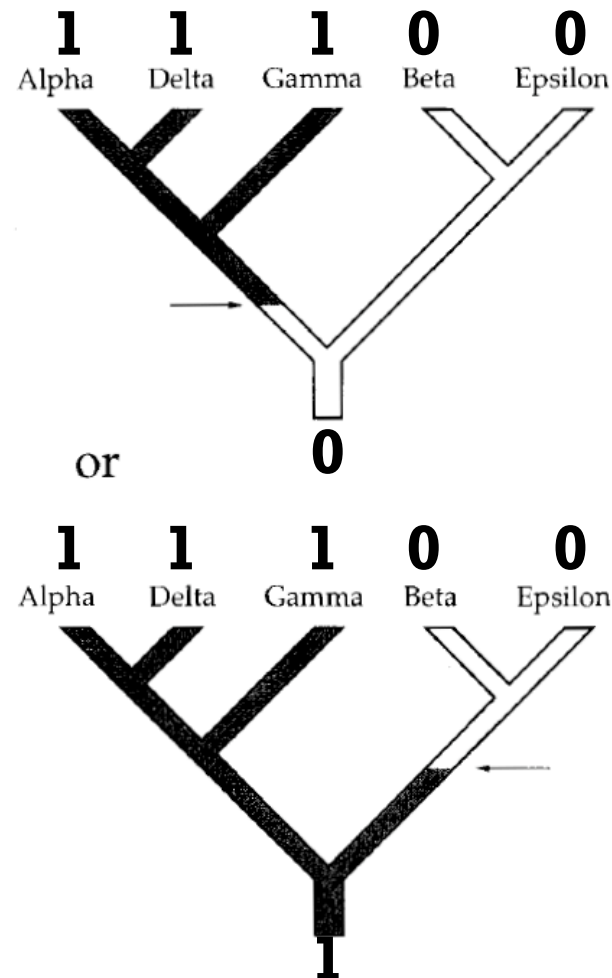
Parsimony Reconstruction

	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Eps	0	0	1	1	1	0



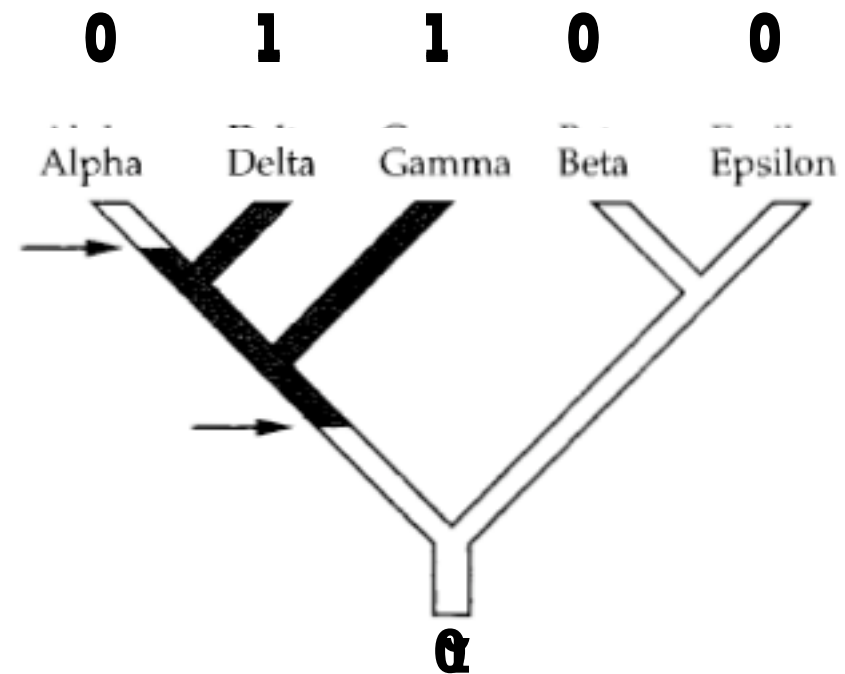
Parsimony Reconstruction

	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Eps	0	0	1	1	1	0

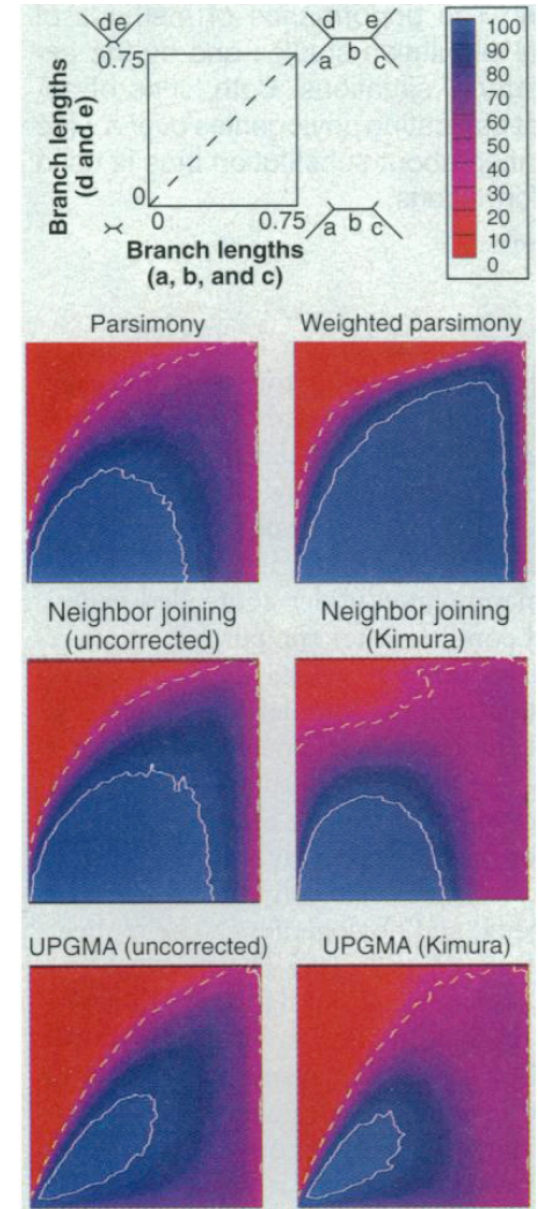
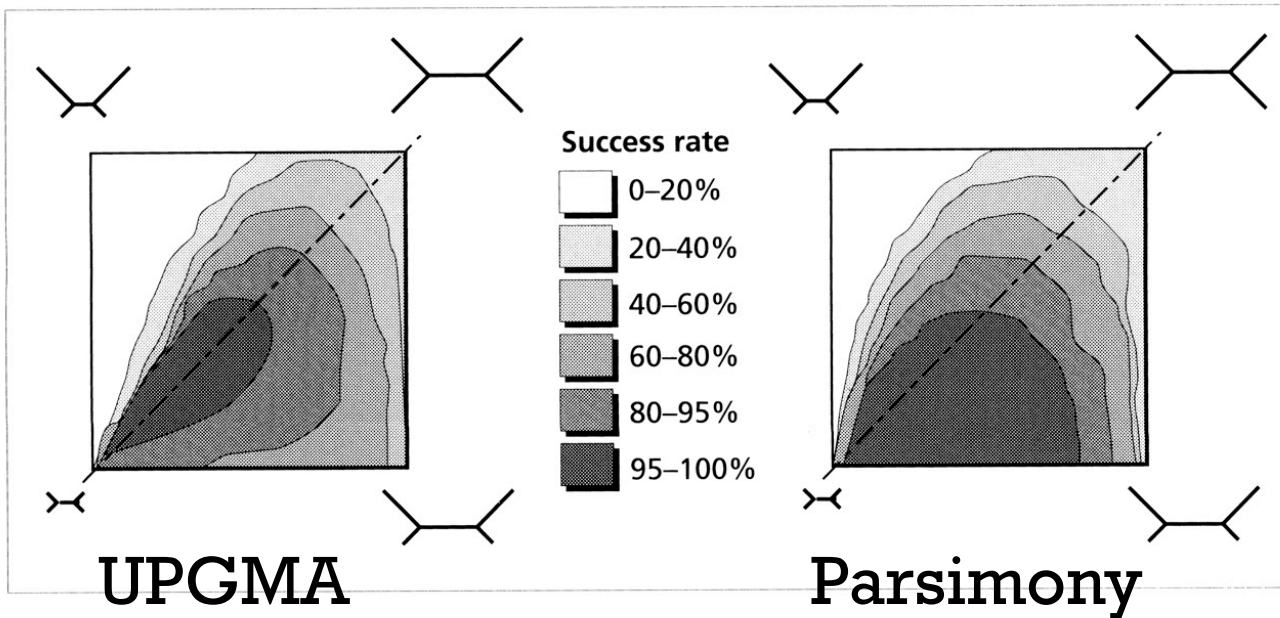


Parsimony Reconstruction

	1	2	3	4	5	6
Alpha	1	0	0	1	1	0
Beta	0	0	1	0	0	0
Gamma	1	1	0	0	0	0
Delta	1	1	0	1	1	1
Eps	0	0	1	1	1	0



Parsimony v. Distance



Cladistics Today

- The Willi Hennig Society.
- The journal *Cladistics*.
- Cladistics and Cladists with a capital 'C.'

