

February, 2002

Genetics 453

Evolutionary Genetics

Molecular Evolution

Joe Felsenstein

Department of Genome Sciences
University of Washington, Seattle

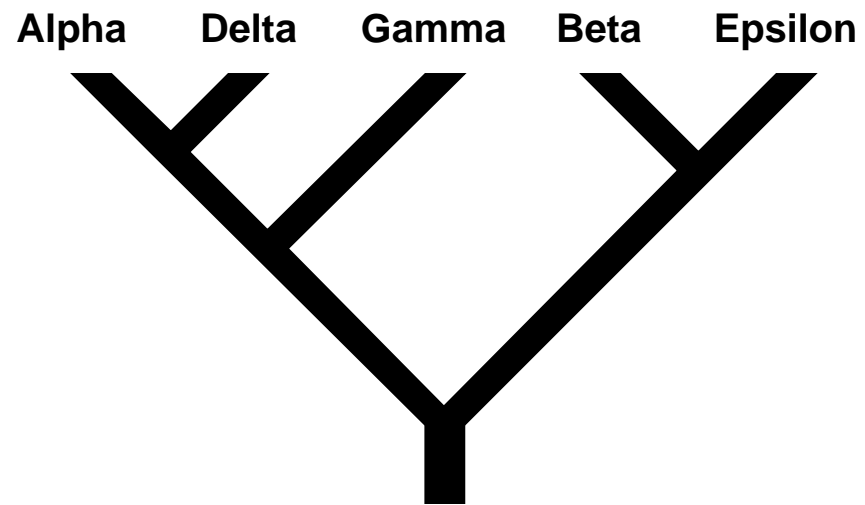
email: joe@genetics.washington.edu

A data example for phylogeny inference

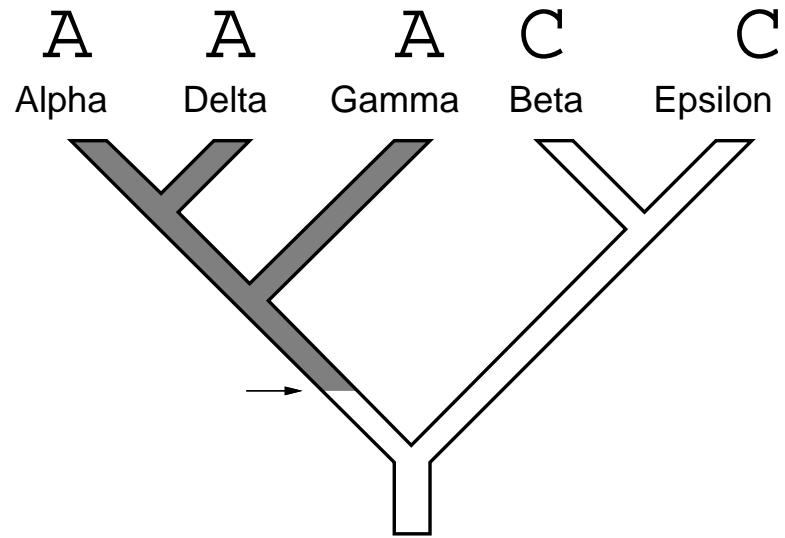
Five DNA sequences, for some gene in an imaginary group of species whose names are Alpha, Beta, Gamma, Delta, and Epsilon:

species	site					
	1	2	3	4	5	6
Alpha	A	T	G	A	G	C
Beta	C	T	C	T	A	C
Gamma	A	G	G	T	A	C
Delta	A	G	G	A	G	T
Epsilon	C	T	C	A	G	C

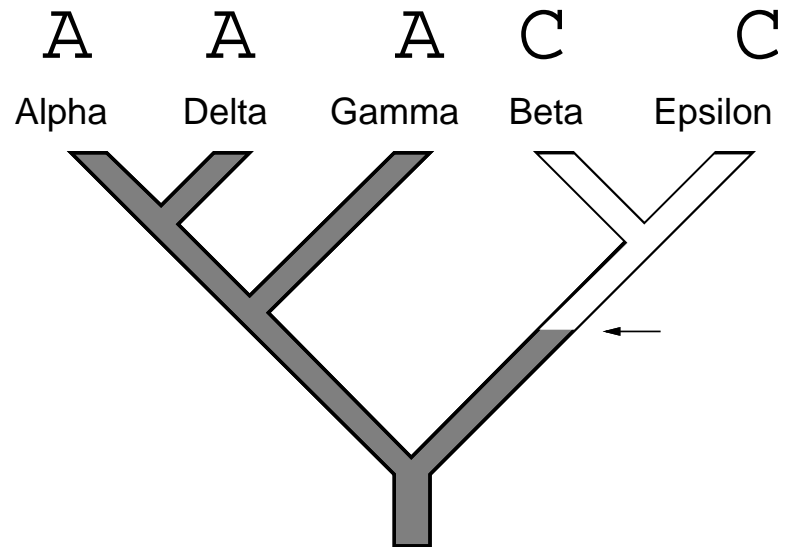
Tree being analyzed for parsimony example



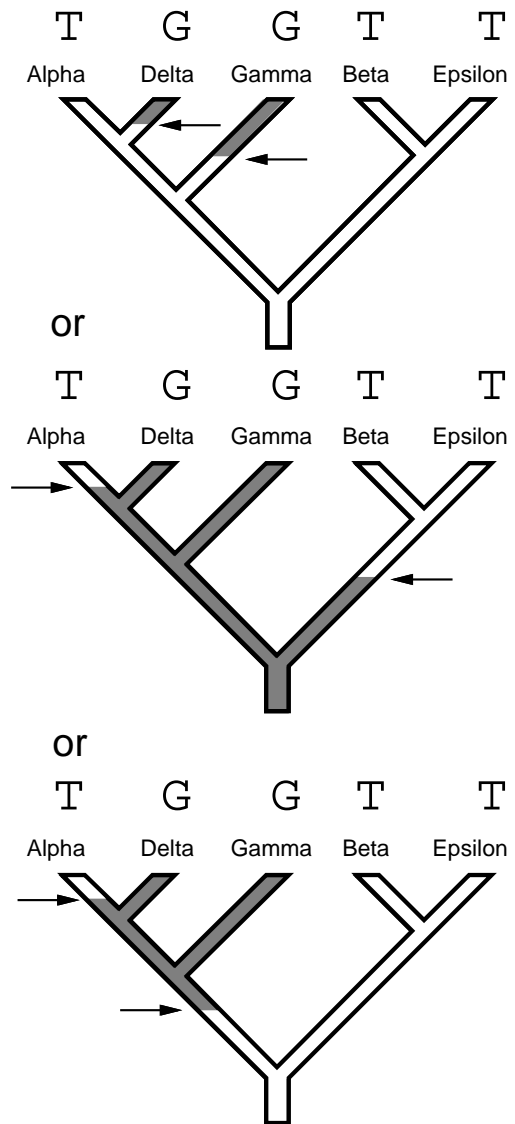
Parsimony example, site 1



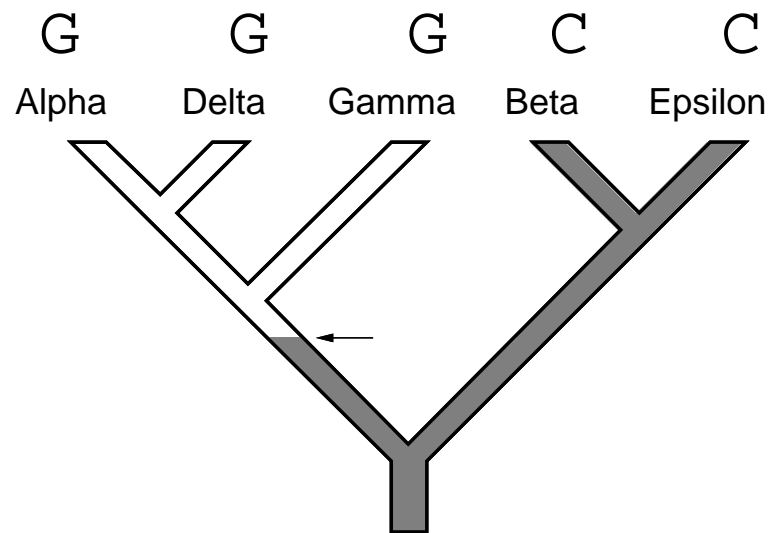
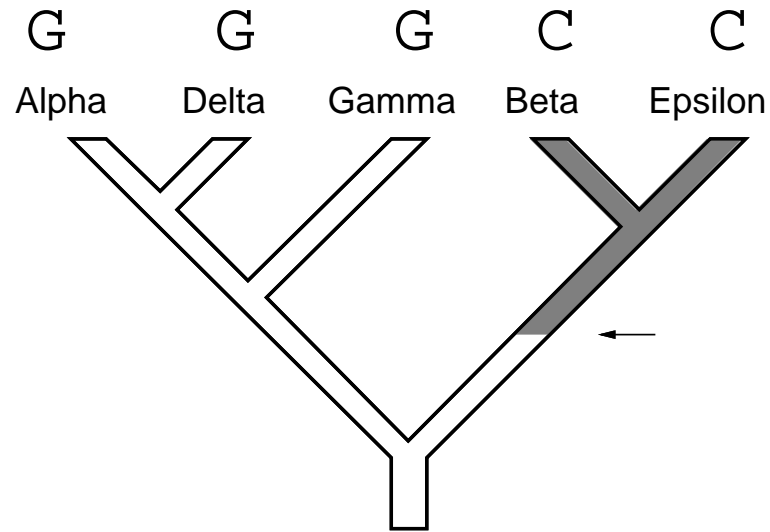
or



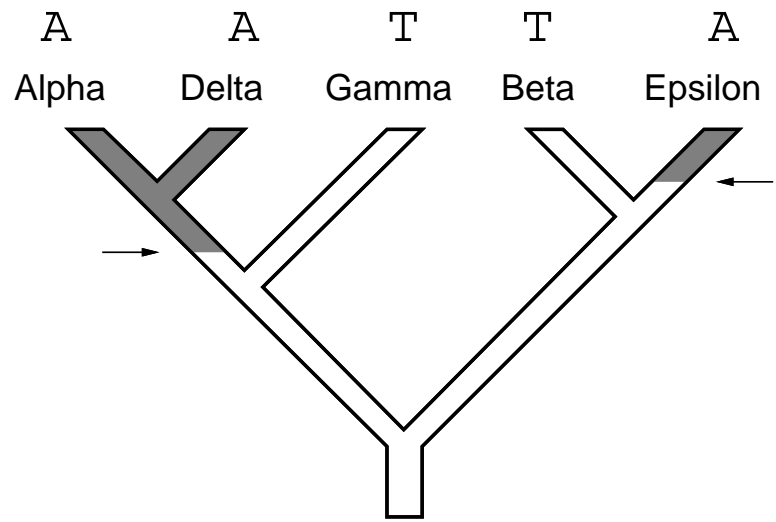
Parsimony example, site 2



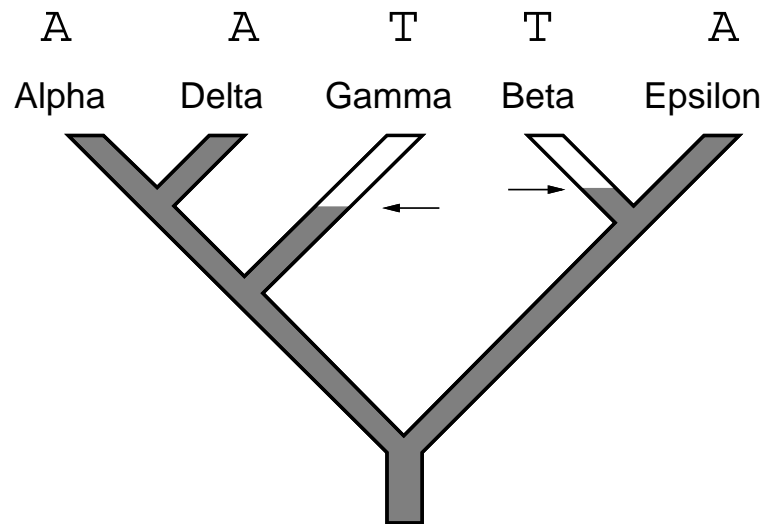
Parsimony example, site 3



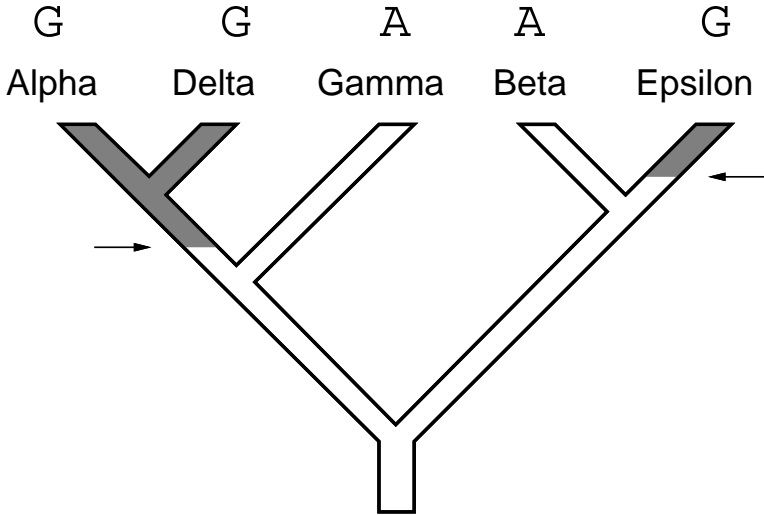
Parsimony example, site 4



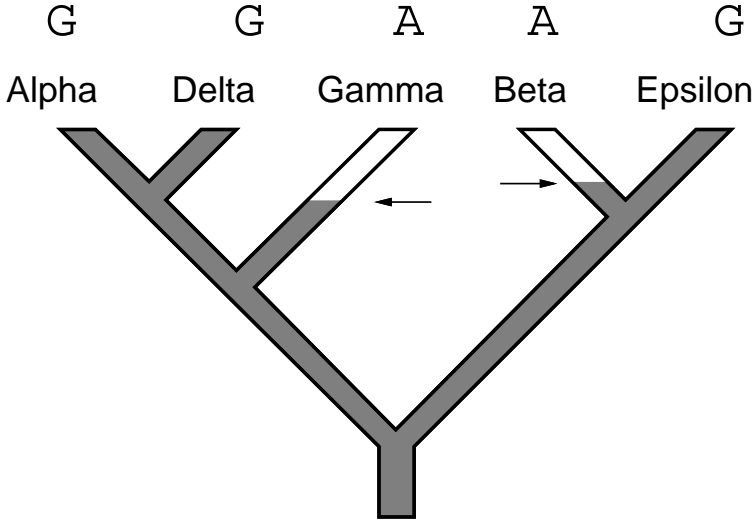
or



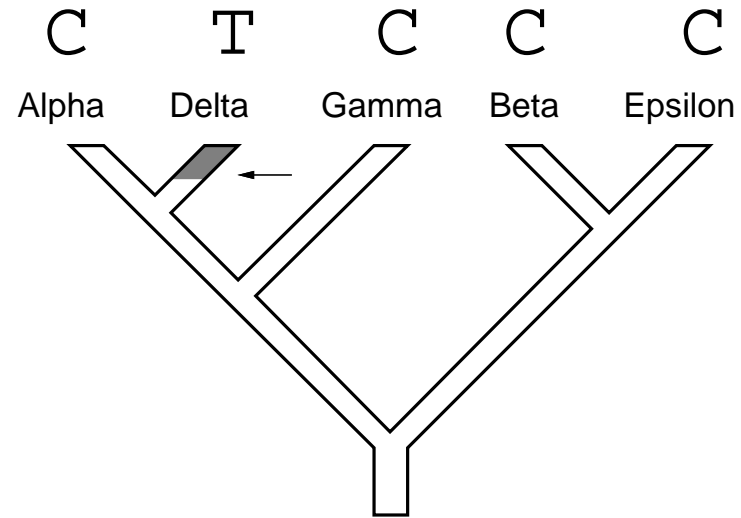
Parsimony example, site 5



or

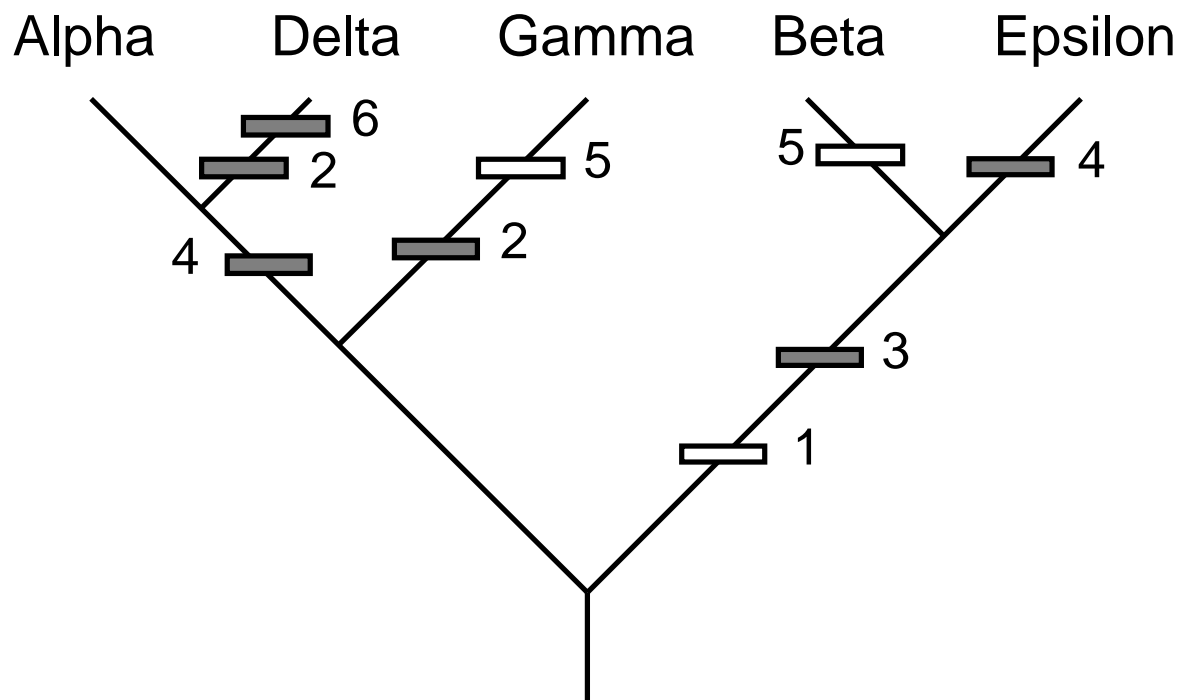


Parsimony example, site 6



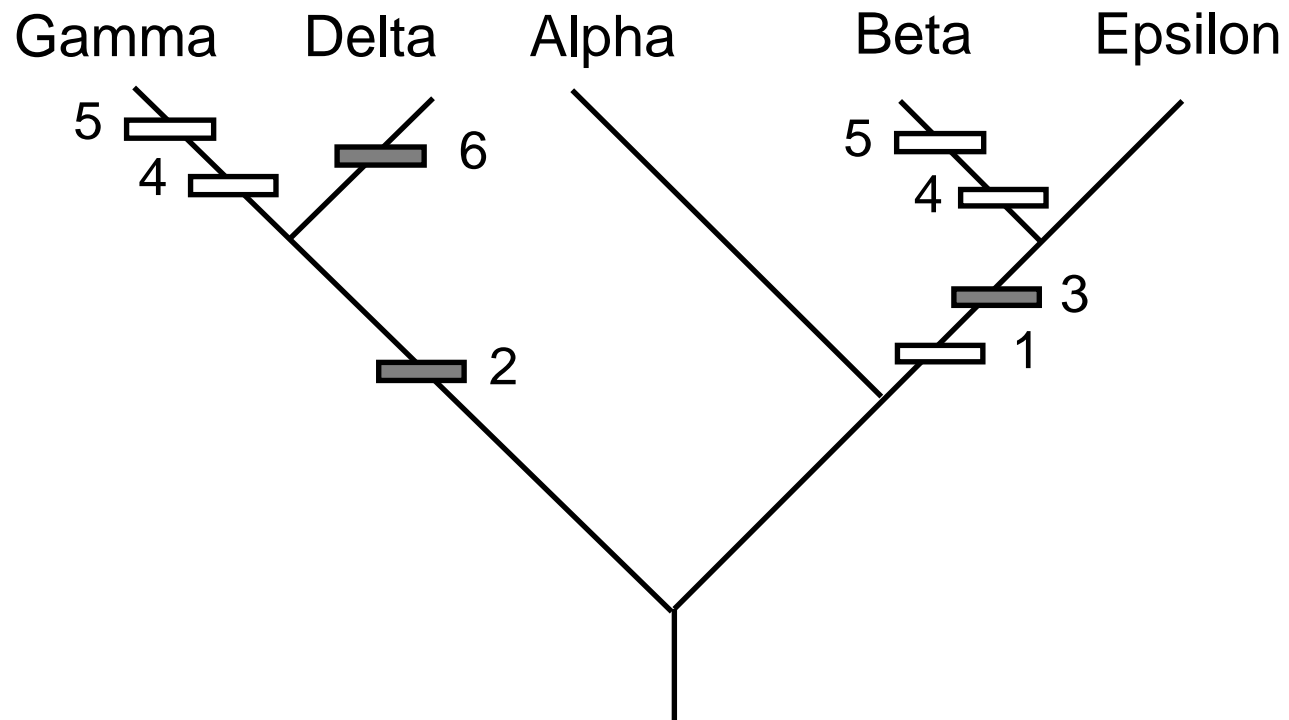
Parsimony example, all changes on tree

showing one of their possible placements



Parsimony example, most parsimonious tree

with one possible placement of the changes

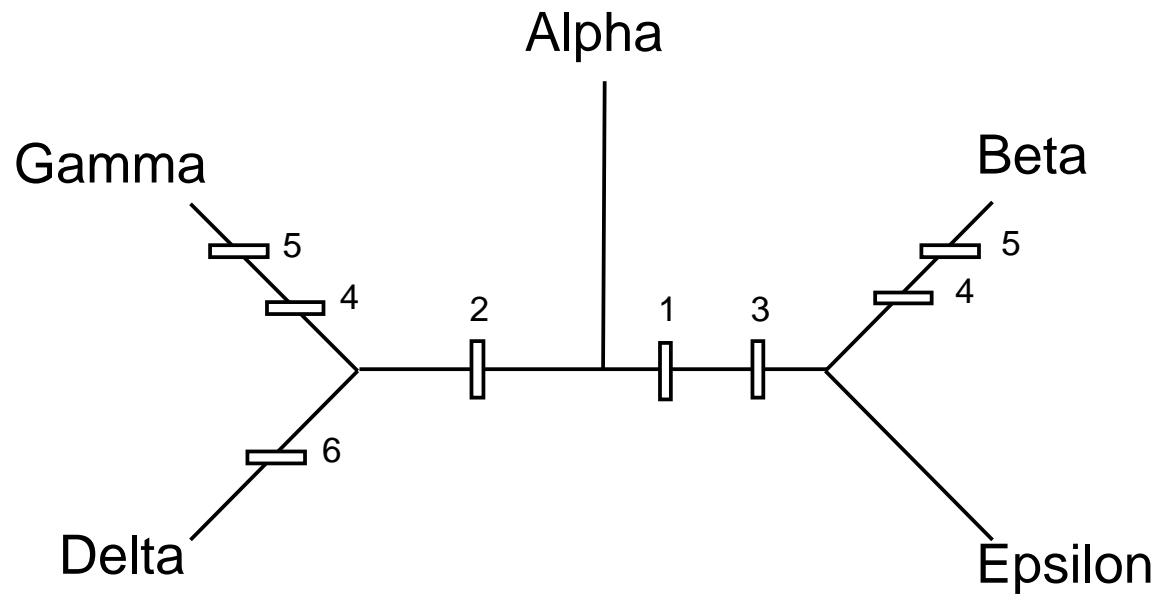


Parsimony example, most parsimonious tree

shown as an unrooted tree

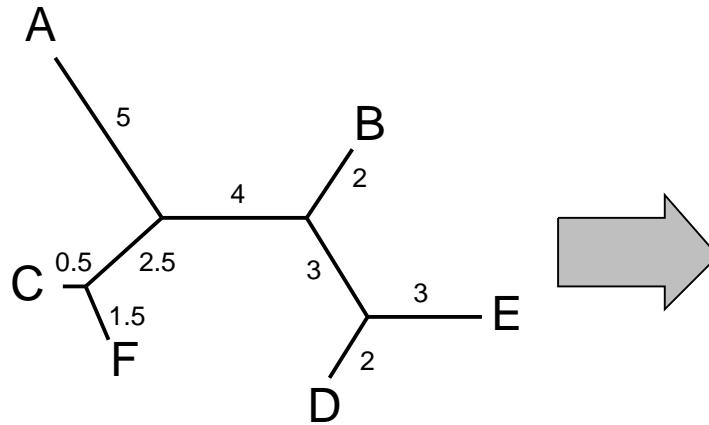
root can be anywhere

changes can occur in either direction



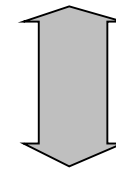
Distance Matrix Methods

Each possible tree (with branch lengths) predict pairwise distances



	A	B	C	D	E	F
A	0	11	8	14	15	9
B	11	0	9	7	8	10
C	8	9	0	13	14	2
D	14	7	13	0	5	13
E	15	8	14	5	0	14
F	9	10	2	13	14	0

Find the one which comes closest to predicting the observed pairwise distances



	A	B	C	D	E	F
A	0	10	9	12	16	9
B	10	0	10	6	9	9
C	9	10	0	10	15	2
D	12	6	10	0	6	13
E	16	9	15	6	0	15
F	9	9	2	13	15	0

Turbeville. J. McC., Schulz, J .R. and R. A. Raff. 1994. Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Molecular Biology and Evolution* **11**: 648-655.

Xenopus	?TACCTGGTTGATCCTGCCAGTAG-CATATGCTTGTCTCAAAGATTAAGCCATGCACGTG
Sebastol	????????????????????????????????AG-CATATGCTTGTCTCAAAGATTAAGCCATGCAAGTC
Latimeri	?TACCTGGTTGATCCTGCCAGTAG-CATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Squalus	????????????????????????????????AG-CATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Myxine	??CCCTGGTTGATCCTGCCAGCCG-CATATGCTTGTCTCAAAGACTAAGCCATGCATGTC
Petromyz	???CCTGGTTGATCCTGCCAGTAG-CATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Branch	???CCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCACGTG
Styela	??ATCTGGTTGATCCTGCCAGTAGTGATATGCTTGTCTCAAAGATTAAGCCATGCAGGTG
Herdman	?TATCTGGTTGATCCTGCCAGTAGTGATATGCTTGTCTCAA-GATTAAGCCATGCAGGTG
Saccogl	??ACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Ophiophol	??ACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTG
Strongyl	??ACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Placopec	CAACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Limicol	?TATCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC
Eurypelm	?TACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC

Tenebrio

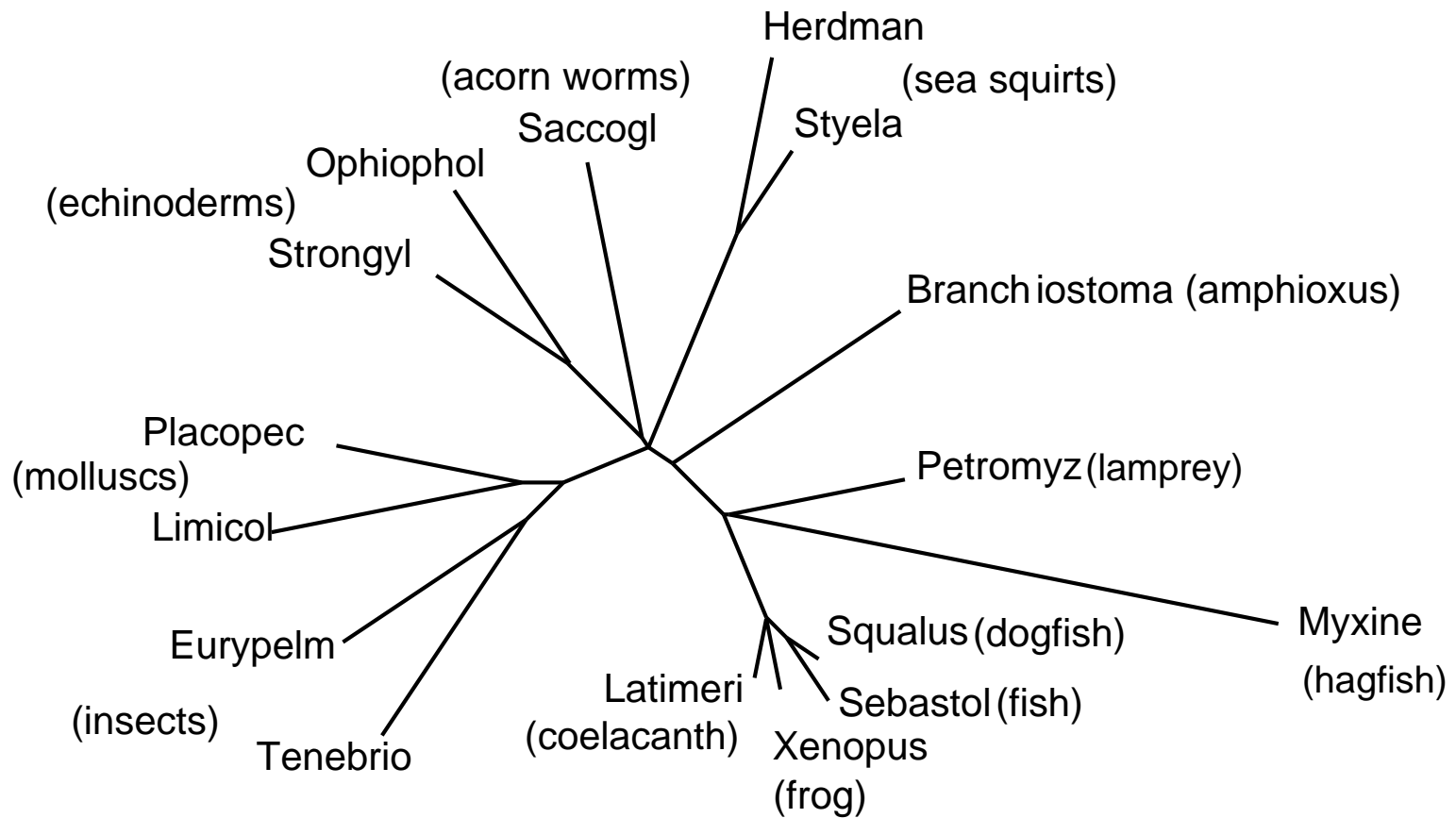
?TCCCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTC

TAAGTACGCACGGCCGGTACA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TAAGTACACACGGCCGGTACA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TAAGTACAAACGGTGCGTACA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TAAGTACACACGGCCGGTACA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TAAGTGCGAACGGACAGACCC-GTTGAGACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TGCGTGCAAACGGATGTTACA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
CAAGTTTAAACTG-TCT--CA-G-TGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
CAAGTACGAGTTG-TCGTAAA-G-CGAAACTGCGAA-TGGCTCATTAAATCAGTCTTGGT
T-----CGAGTTG-TCGTAAA-G-TGAAACGTCGAAACGGCTCATTAAATCAGTCTTGGT
TAAGTACGAGCCT-CGGTACACGGTGAAACTGCGAA-TGGCTCATTAAATCAGTTATGGT
TTAGTACAAGCTTGTA-T-CAAG-CGAAACTGCGGA-TGGCTCATTAAATCAGTCATGGT
TAAGTACAAGCTCGTC-T-CGA--CGAAACTGCGGA-TGGCTCATTAAATCAGTTATGGT
TAAGTACATACTTTTTG-ATG-G-TGAAACCGCGAA-TGGCTCATTAAATCAGTTATGGT
TAAGTTCACACTGTCT-CACG-G-TGAAACCGCGAA-TGGCTCATTAAATCAGTCGAGGT
TAAGTACATGCCTCCT-TAAG-G-CGAAACCGCGAA-TGGCTCATTAAATCAGTTATGGT
TCAGTACAAGCCGAAT-TAAG-G-TGAAACCGCGAA-AGGCTCATTAAATCAGTTATGGT

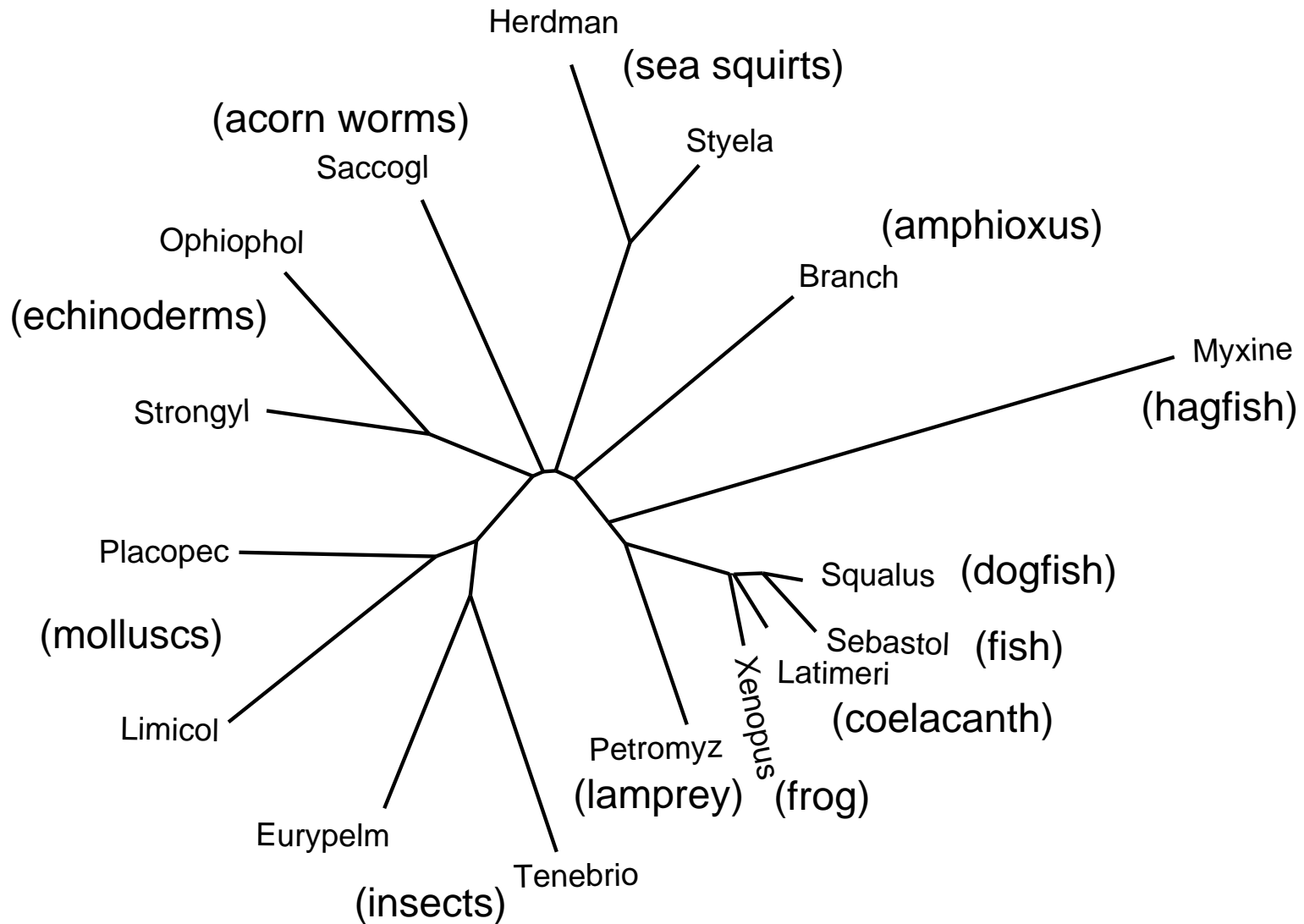
TCCTTTGATCGCTCCATCTGTTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCCTTTGATCGCTCTCA-CGTTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC
TCCTTTGATCGCTCCAA-CGTTACTCGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCCTTTGATCGCTCCAAACGTTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC
CTCTTTGAGCGCTCGGAC-GTTACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC
TCCTTTGATCGCTCGACC-GTTGCTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCCTTTGATCG-TCACATCCT-ACATGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TTATTTGGTC--TCGAAAGCT-AAGTGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TTATTTGGTC-TTGTGAGCG--AAGT-GATAACTGTGGGC--TCTAGAGCTAATACATGC
TCCTTTGATCGTTACCCCTT--ACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC
TCCTTGGAACGAGTTGCCCT--ACATGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCATTGGATCGAGTCCCCCG-ACATGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCCTTAGATCGTACGATCCT--ACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC
TCCTTAGATGACACGATCCT--ACTTGGATAACTGTGGCAATTCTAGAGCTAATACATGC
TCCTTAGATCGTACCTTACT--ACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC
TCCTTAGATCGTACCCACATTTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGC

(and so on for 9 more pages)

Chordates from Turbeville rRNA data set
analyzed by parsimony



Chordates from Turbeville data analyzed
by Fitch–Margoliash distance matrix method





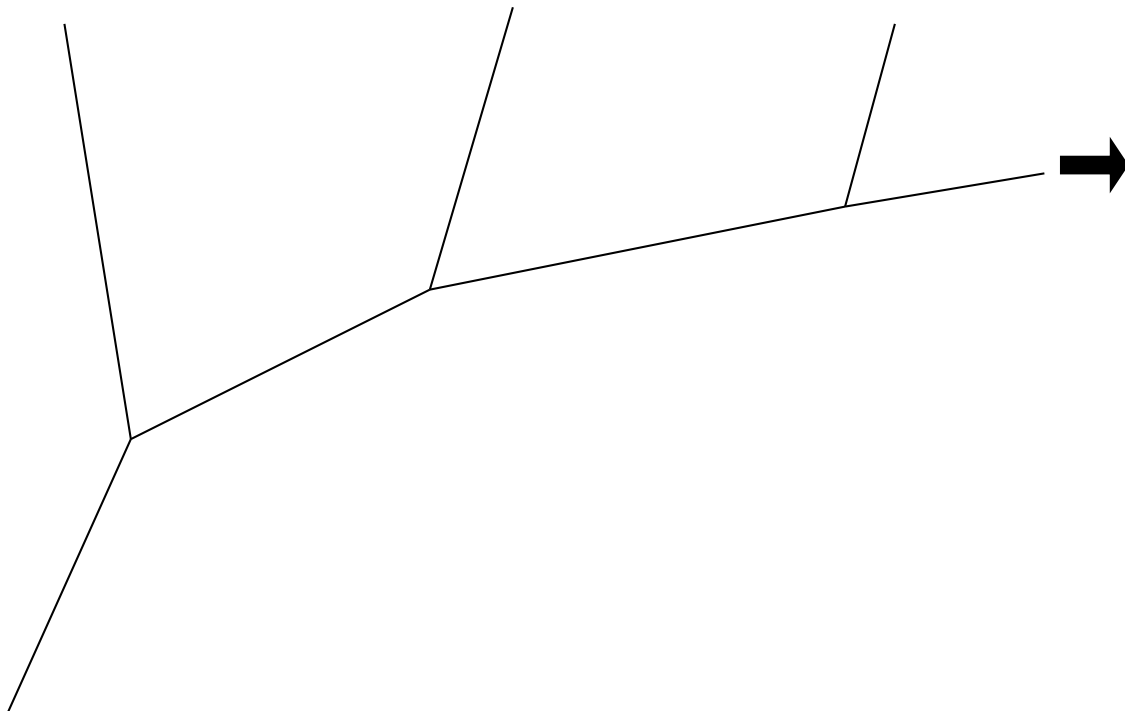
Placopecten
(scallop)



Strongylocentrotus
(sea urchin)



Saccoglossus
(acorn worm)





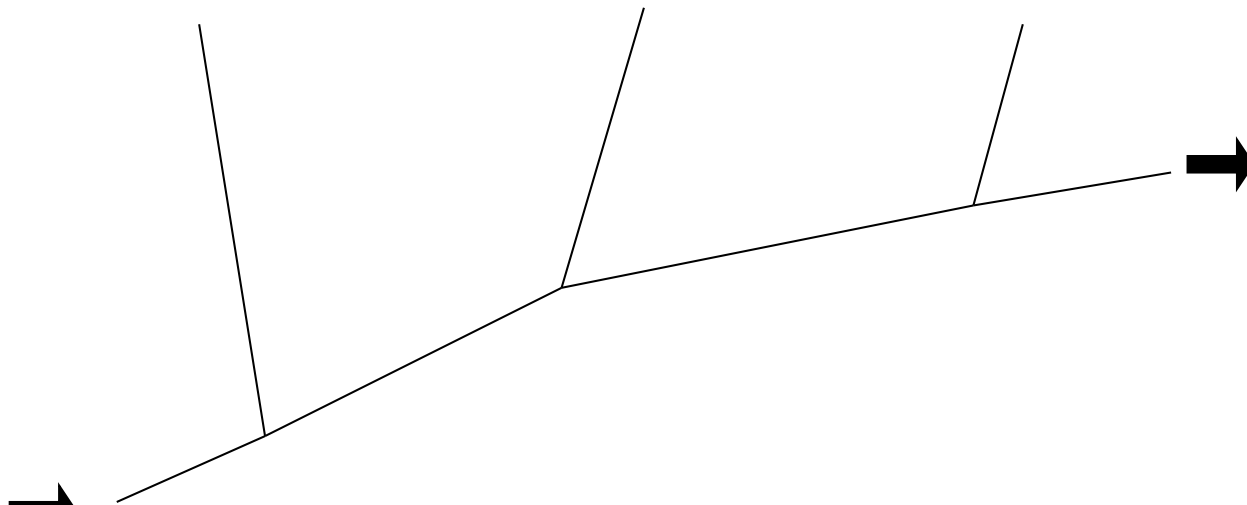
tunicate



Branchiostoma
(amphioxus)



Myxine
(hagfish)





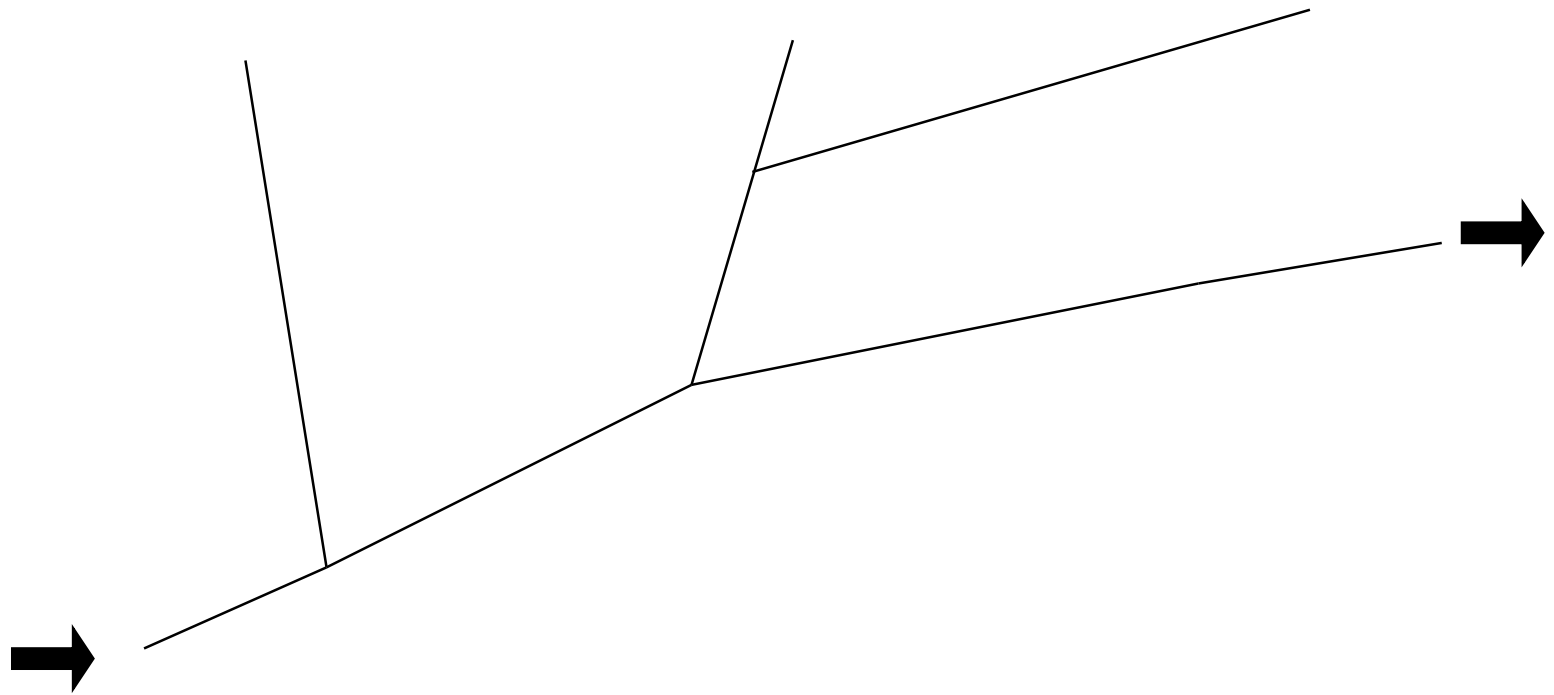
Petromyzon
(lamprey)



dogfish



dogfish





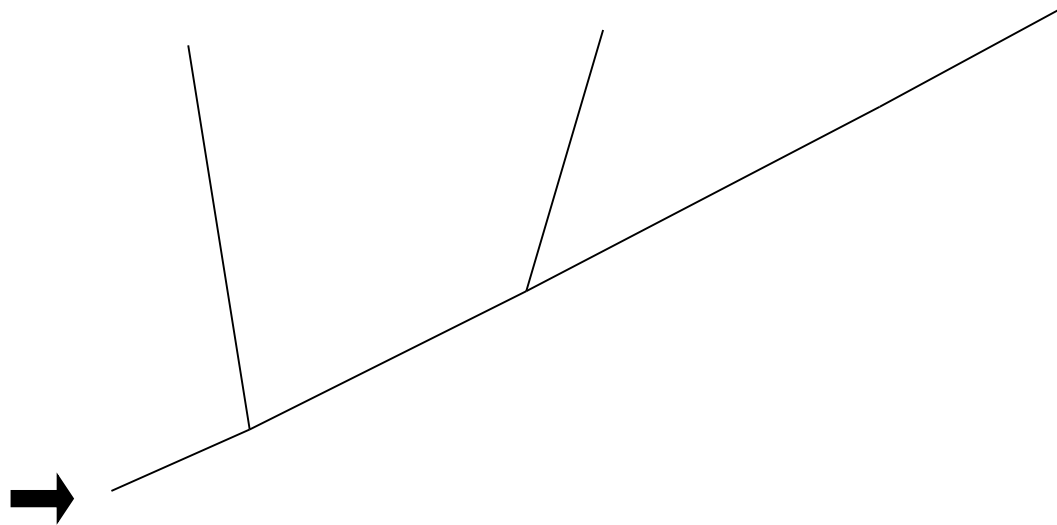
Sebastes
(rockfish)



Latimeria
(coelacanth)



Xenopus
(clawed frog)



Alignment of hemoglobin β loci of Human, Tarsier



region	bases	differences	% different
upstream	100	12	12.0
exon 1	92	9	9.8
intron 1	126	26	20.6
exon2	223	26	11.7
intron 2	820	239	29.1
exon 3	129	13	10.1
downstream	100	13	13.0

Differences in exons

position 1	10
position 2	5
position 3	33

Higher Rates of Substitution for ...

1. ... some proteins than others
2. ... some sites within proteins than others (less in active sites, interior sites)
3. ... some amino acid replacements than others (less changes to chemically more similar amino acids)
4. ... silent changes than nonsilent ones
5. ... “in-between” DNA than introns, introns than coding sequences
6. ... transitions than transversions

Morris Goodman tabulation for β hemoglobin

where	sites	change/100myr
Heme contacts	21	0.02
Nonheme contacts	10	0.02
Salt bridges β - β	4	0.00
2,3-DPG binding	4	0.10
Nonsalt bridge α , β contacts	16	0.16
Remaining interior	21	0.09
Remaining exterior	70	0.20
All	146	0.13

Rate of molecular evolution
from neutral and selective mechanisms

Neutral mutations

A fraction μ of all copies of a gene mutate. Of these $\frac{1}{2N}$ (equal to the initial frequency of the mutant) succeed in drifting to fixation for the mutant.

There are in all $2N$ copies of the gene available to mutate.

The resulting rate of substitution is

$$\mu \times \frac{1}{2N} \times 2N = \mu$$

So the rate of substitution of neutral mutations is equal to the mutation rate (the mutation rate of neutral mutants, not the total mutation rate).

Rate of molecular evolution
from neutral and selective mechanisms

Selectively advantageous mutations

A fraction μ of copies of the gene mutate. There are in all $2N$ copies available. A fraction $2s$ succeed in fixing.

The resulting rate of substitution is

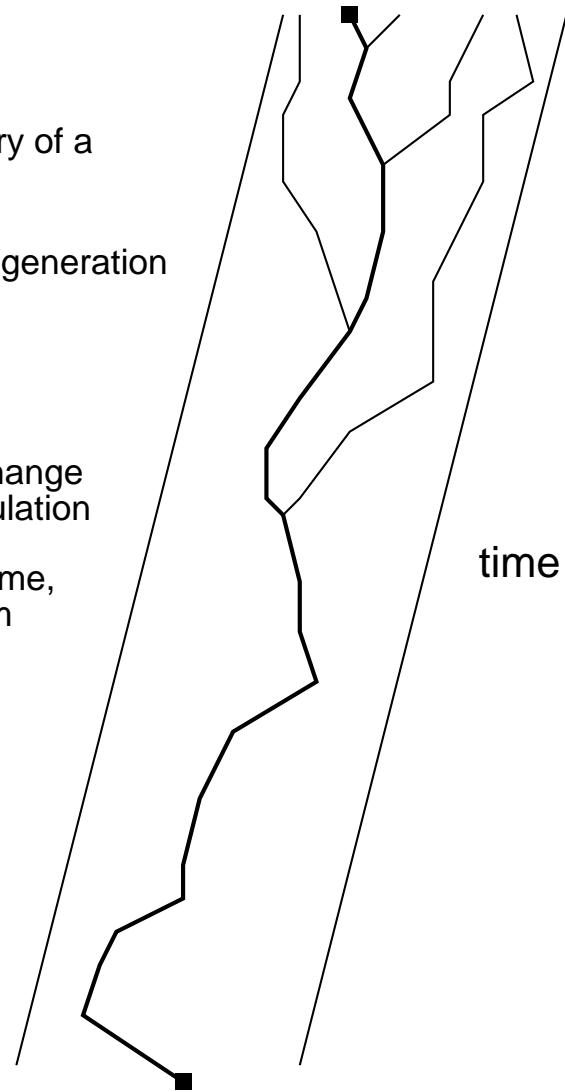
$$\mu \times 2N \times 2s = 4Ns\mu$$

Note that this is $4Ns$ times as high as for neutral mutants, if the mutation rate in both categories were equal (which it isn't).

Another way of getting Kimura's result

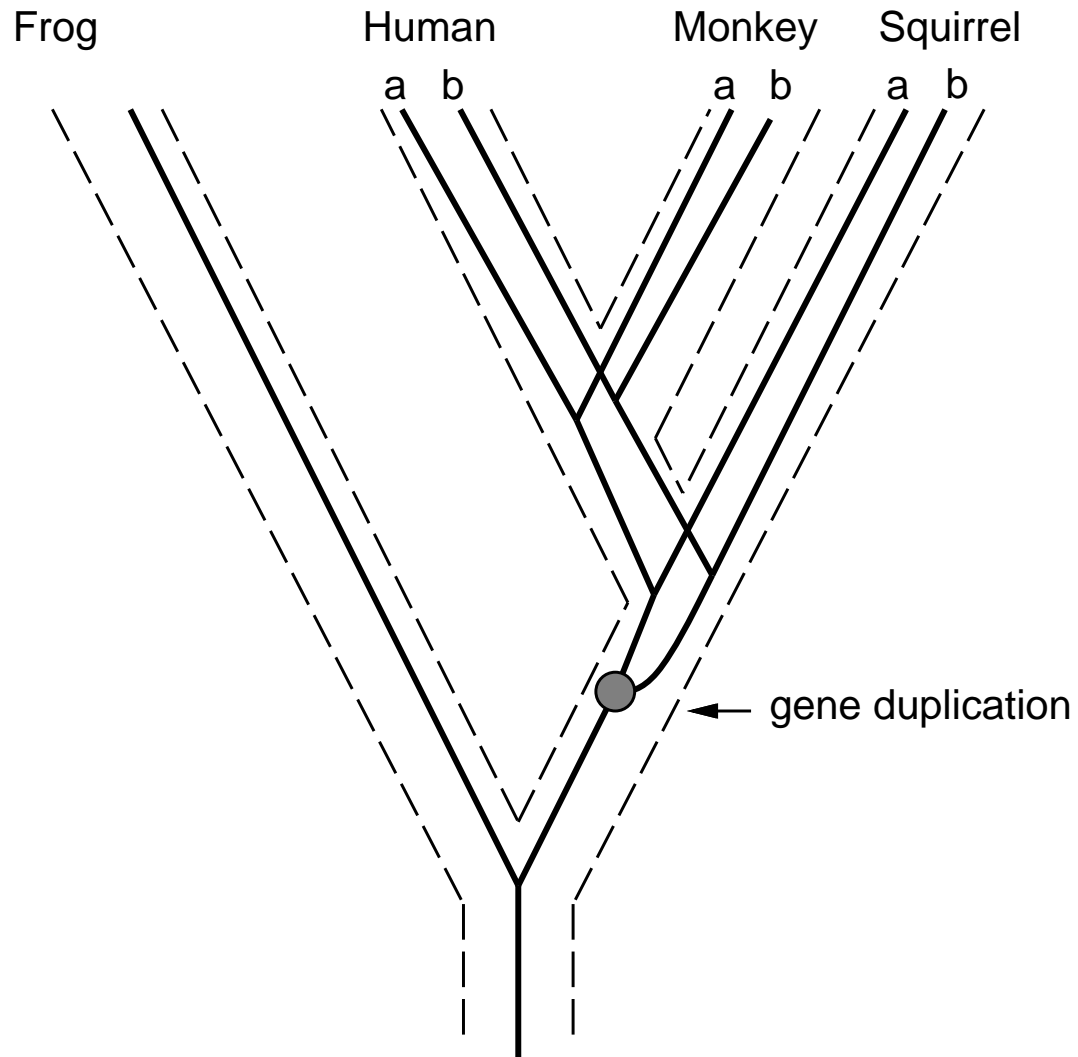
end of a lineage which is the ancestry of a
single copy of the gene:
change is expected to be μ per generation

Interestingly, the expected rate of change
is the same no matter what the population
size is (small populations make all
copies more likely to become the same,
but all are still expected to differ from
their ancestors by this amount



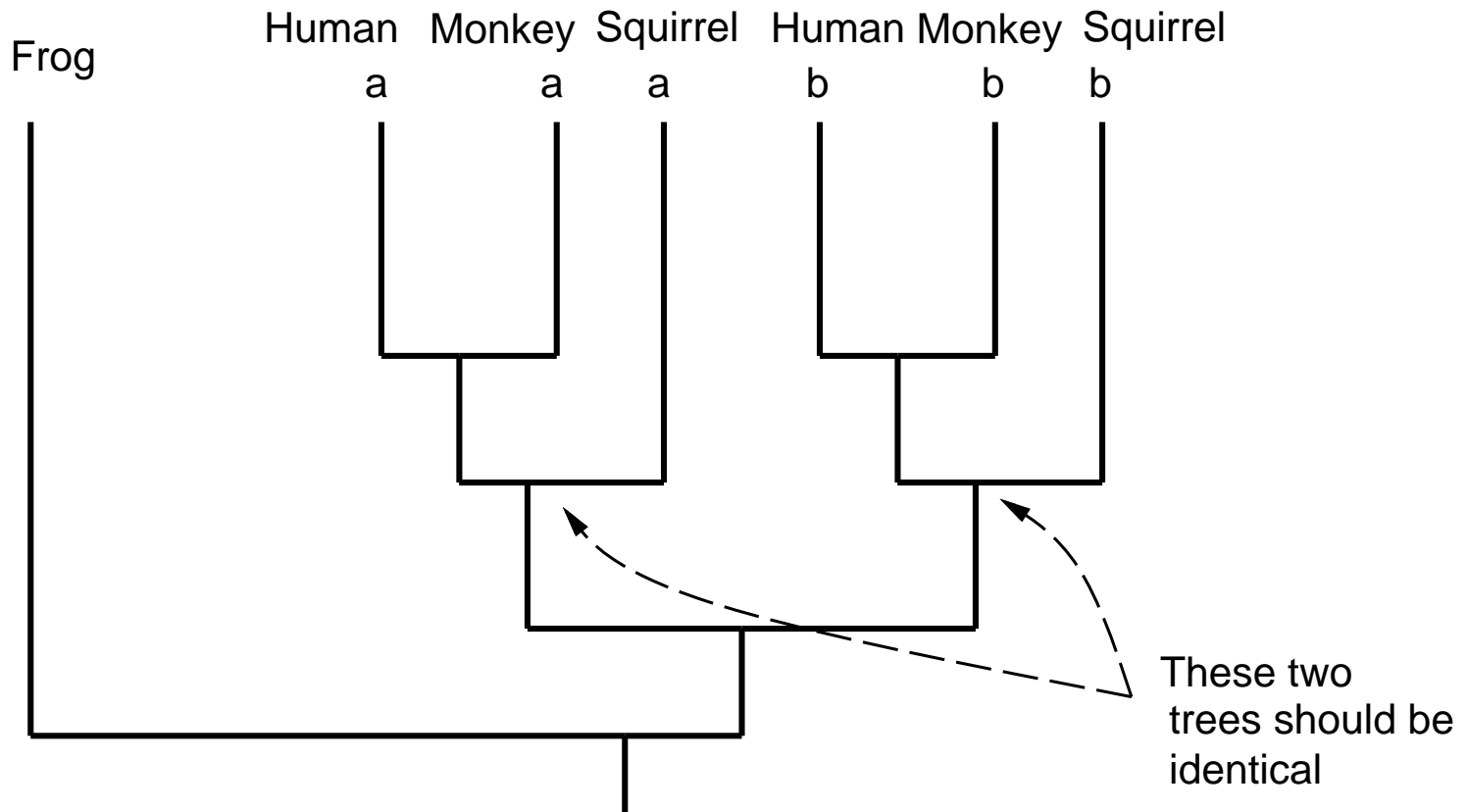
Phylogenies, gene trees, and gene duplication

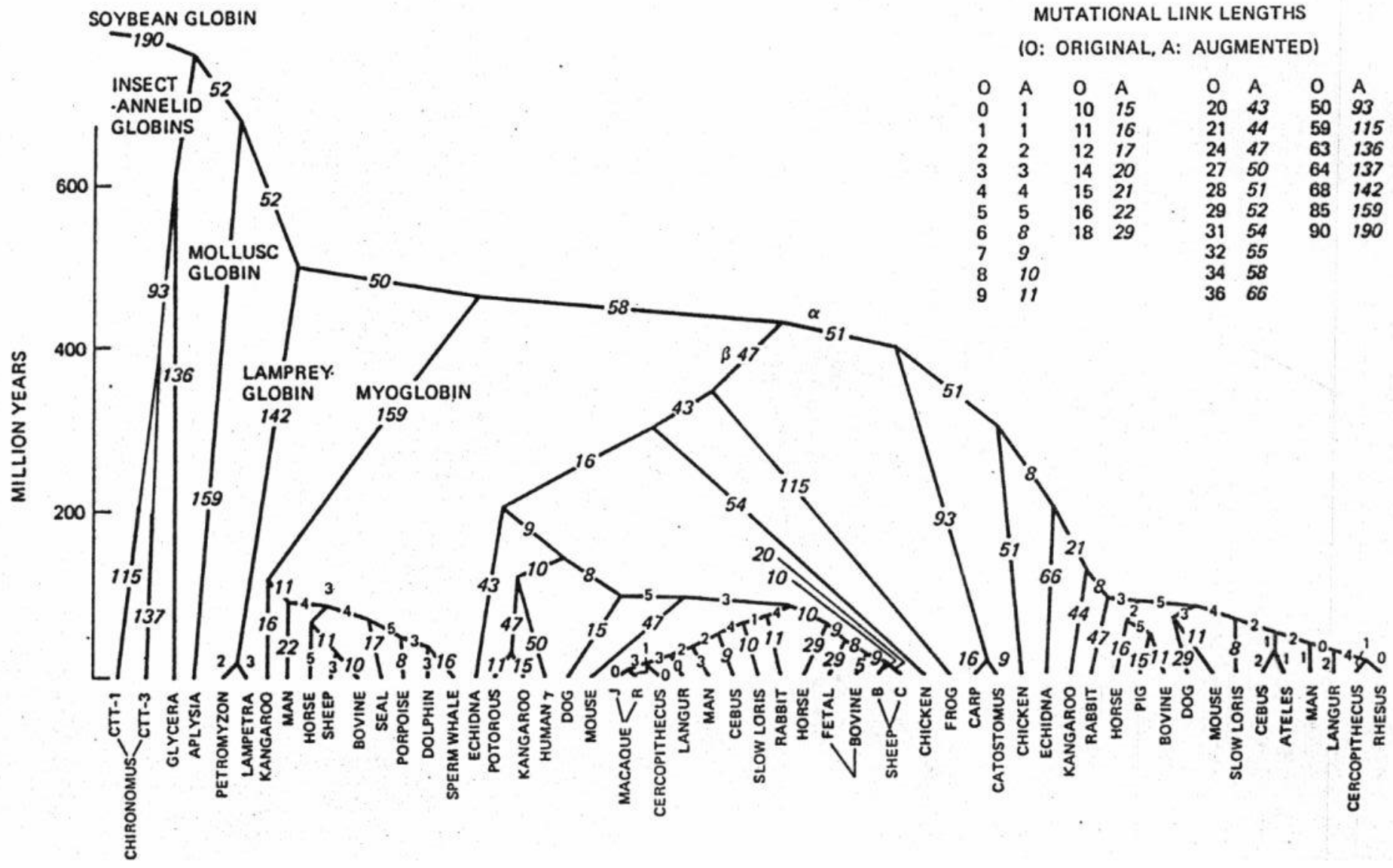
A phylogeny with a gene duplication event:



Phylogenies, gene trees, and gene duplication

So when genes are all aligned with each other, their "gene tree" is:





MUTATIONAL LINK LENGTHS
(O: ORIGINAL, A: AUGMENTED)

0	A	0	A	0	A	0	A
0	1	10	15	20	43	50	93
1	1	11	16	21	44	59	115
2	2	12	17	24	47	63	136
3	3	14	20	27	50	64	137
4	4	15	21	28	51	68	142
5	5	16	22	29	52	85	159
6	8	18	29	31	54	90	190
7	9			32	55		
8	10			34	58		
9	11			36	66		

This freeware-friendly presentation prepared with

- Linux (operating system)
- PDFLaTeX (mathematical typesetting and PDF preparation)
- Idraw (drawing program to modify plots and draw figures)
- Adobe Acrobat Reader (to display the PDF in full-screen mode)

(except that we had to use Microsoft Windows to project this as the X server I have in Linux is not too great)