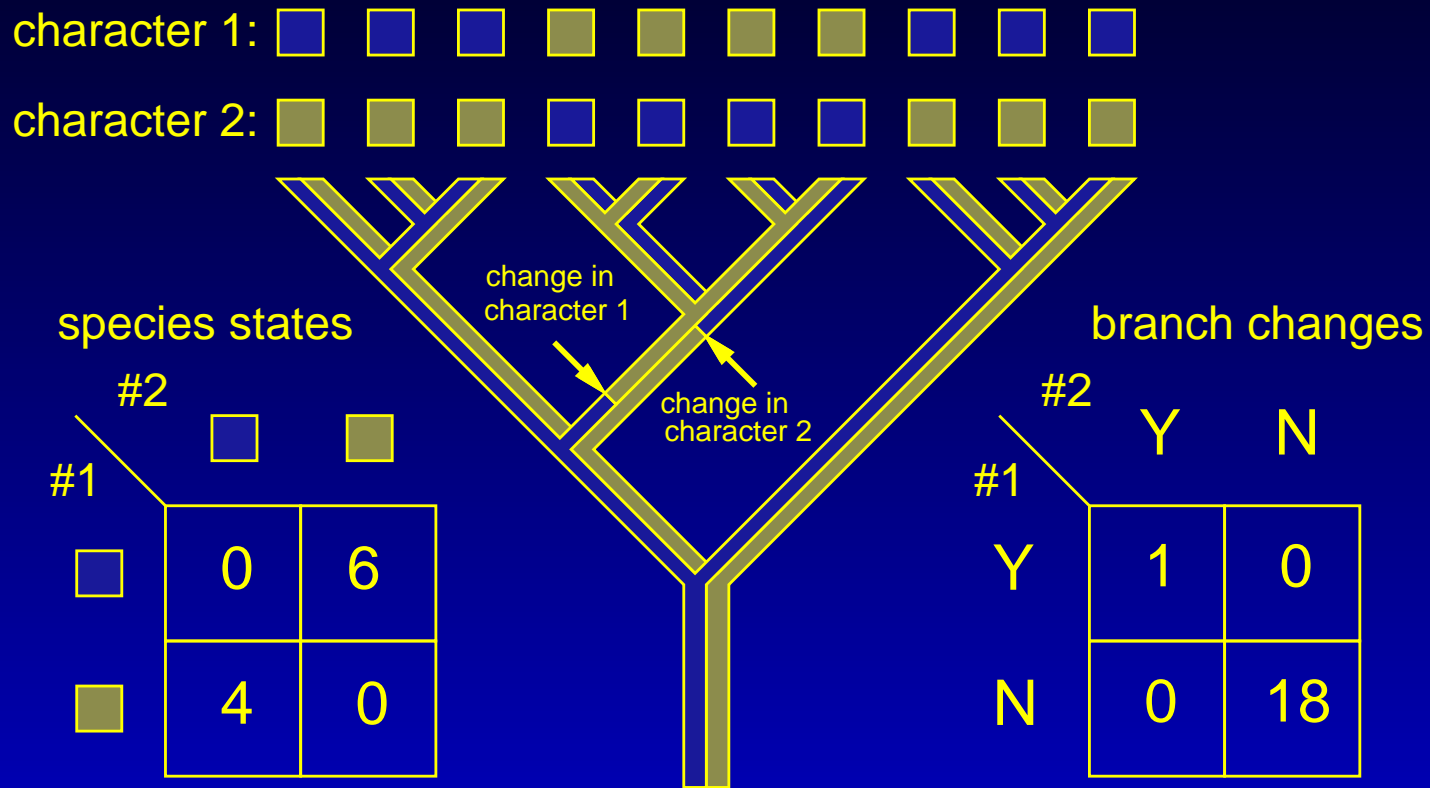Some of these slides have been borrowed from Dr. Paul Lewis, Dr. Joe Felsenstein. Thanks!
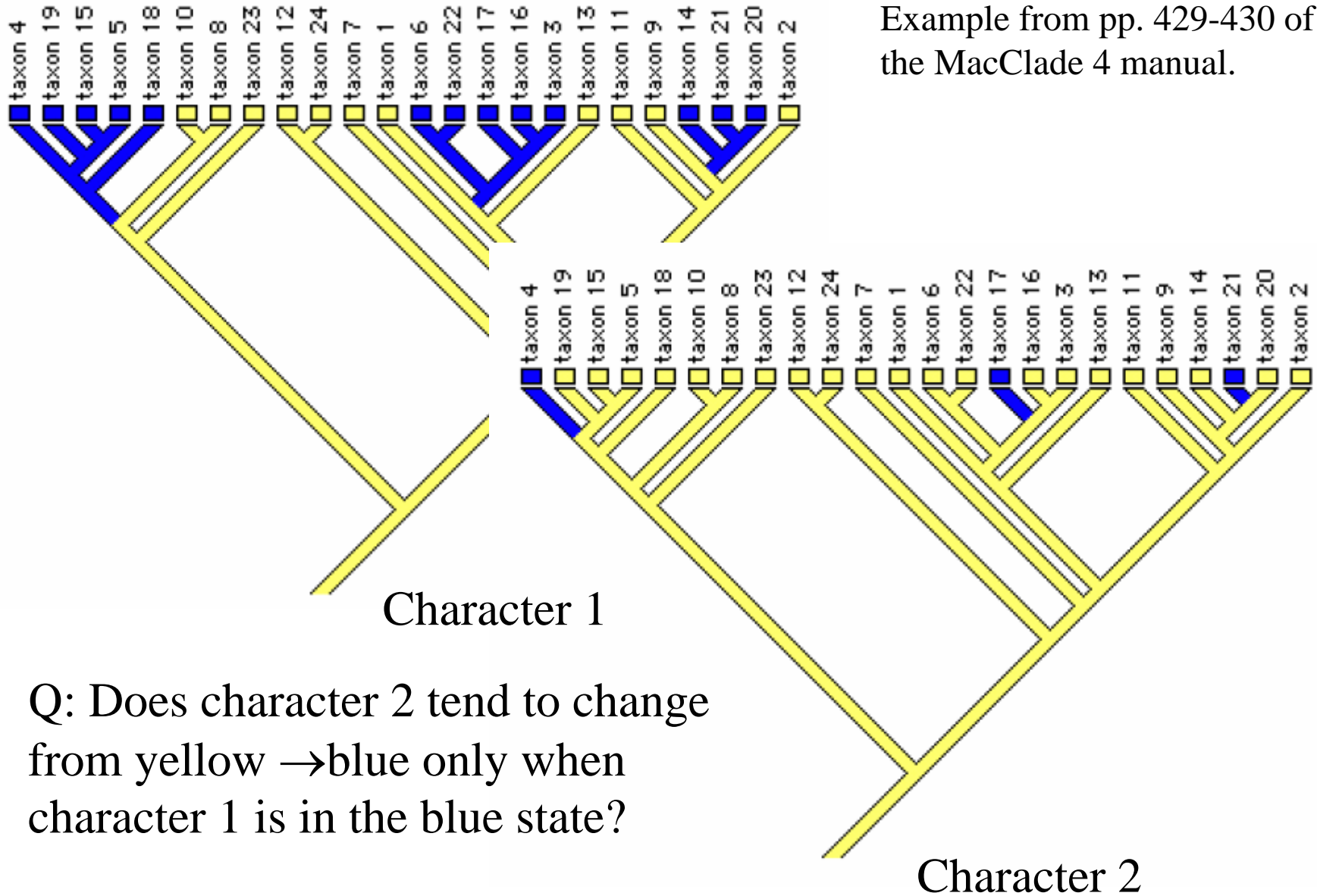
Paul has many great tools for teaching phylogenetics at his web site:

http://hydrodictyon.eeb.uconn.edu/people/plewis

Correlation of states in a discrete-state model

character 1: ▢ ▢ ▢ ▨ ▨ ▨ ▨ ▢ ▢ ▢

character 2: ▨ ▨ ▨ ▢ ▢ ▢ ▢ ▨ ▨ ▨

change in
character 1

change in
character 2

species states

branch changes

#2

	▢	▨
▢	0	6
▨	4	0

#1

#2

	Y	N
Y	1	0
N	0	18

#1

Example from pp. 429-430 of the MacClade 4 manual.

Character 1

Character 2

Q: Does character 2 tend to change from yellow →blue only when character 1 is in the blue state?

2

Concentrated Changes Test
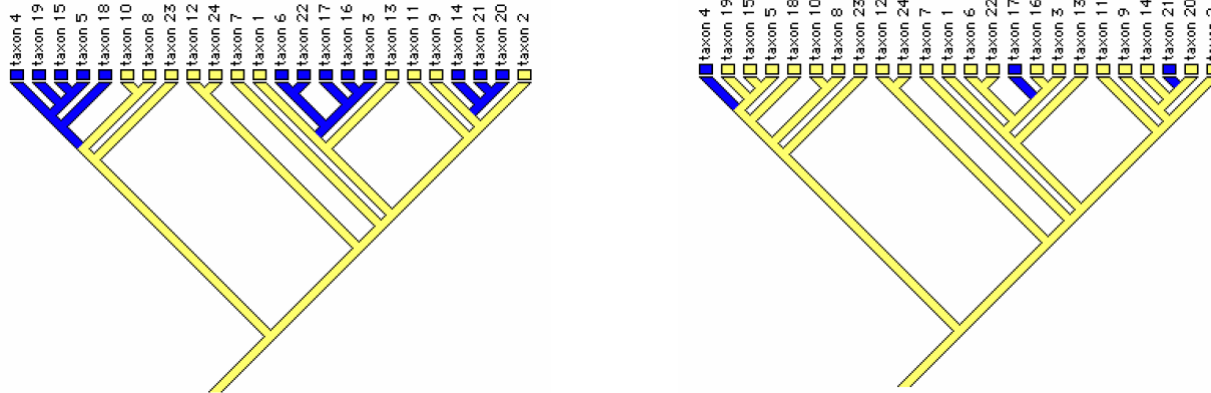
More precisely: how often would three yellow→blue changes occur in the blue areas of the cladogram on the left if these 3 changes were *thrown down at random* on branches of the tree?

Answer: 12.0671% of the time. Thus, we cannot reject the null hypothesis that the observed coincident changes in the two characters were simply the result of chance.

3

Data for Two Characters, X and Y

	X	Y
A	27	122
B	33	124
C	18	126
D	22	128

Var(X)   =   42.000
Var(Y)   =   6.667
Cov(XY)   =   -10.000
Correlation   =   -0.5976

The negative correlation is fairly strong, but would it weaken if it were recognized that there are not really 4 independent data points here...
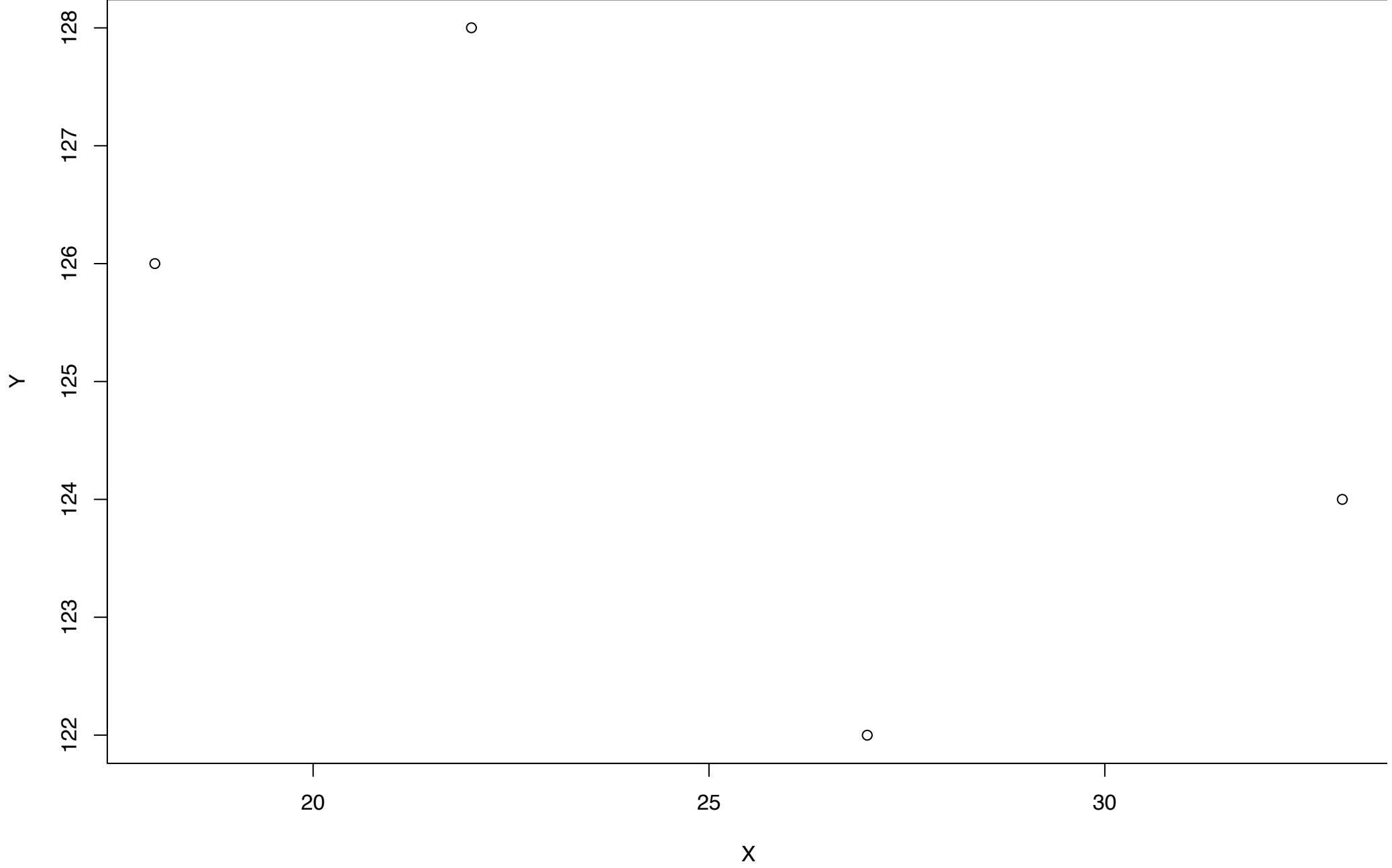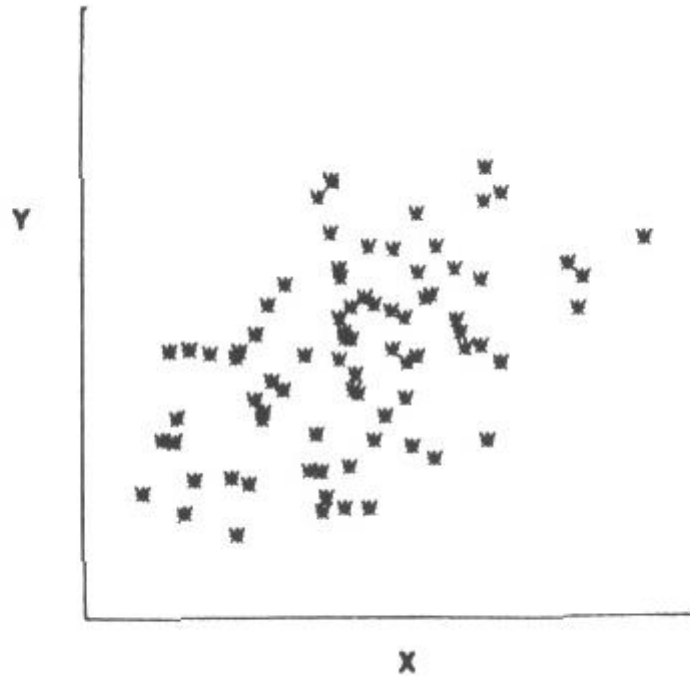
Fig. 6 from Felsenstein (1985)

FIG. 6.—A typical data set that might be generated for the phylogeny in fig. 5 using the model of independent Brownian motion (normal increments) in each character.
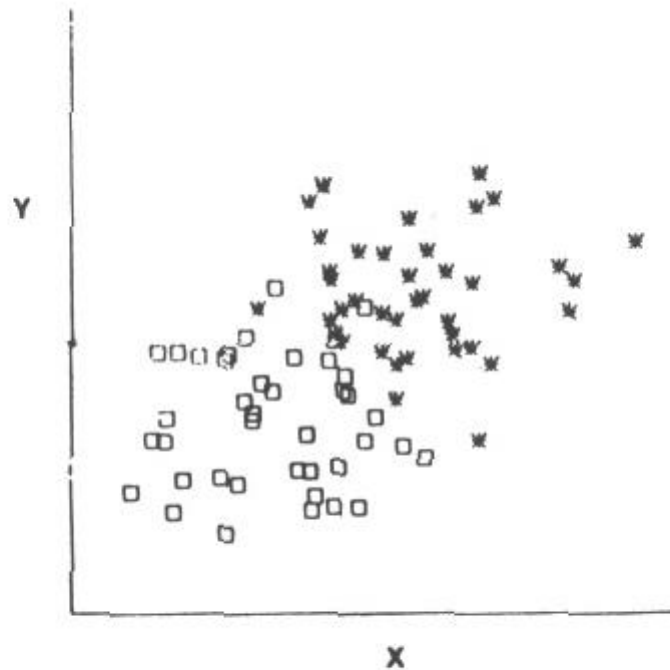
3

Fig. 7 from Felsenstein (1985)

FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.
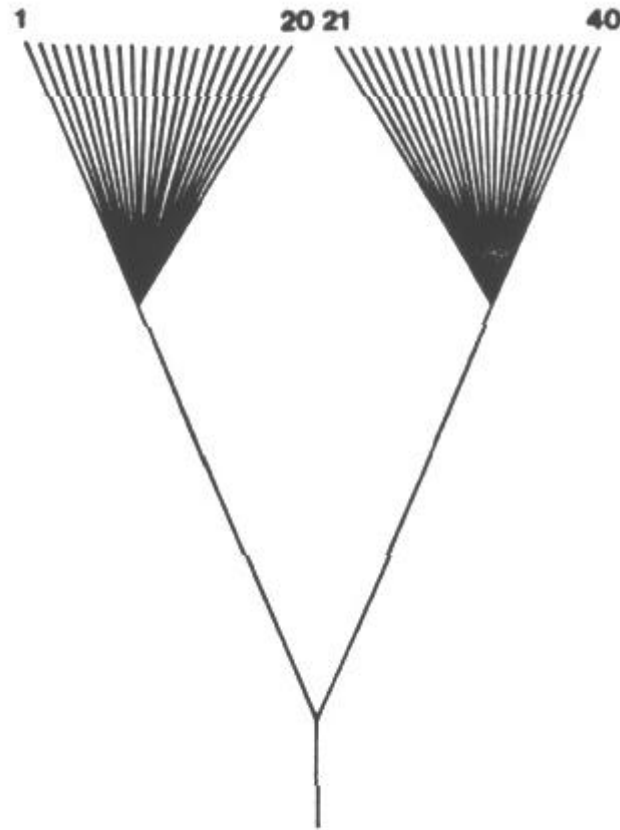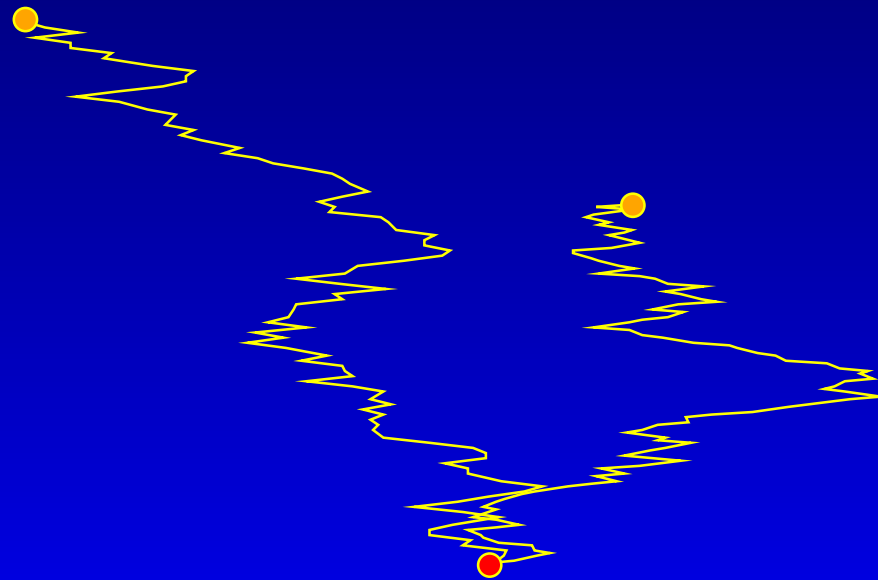
Fig. 5 from Felsenstein (1985)

FIG. 5.—A "worst case" phylogeny for 40 species, in which there prove to be 2 groups each of 20 close relatives.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1-15.
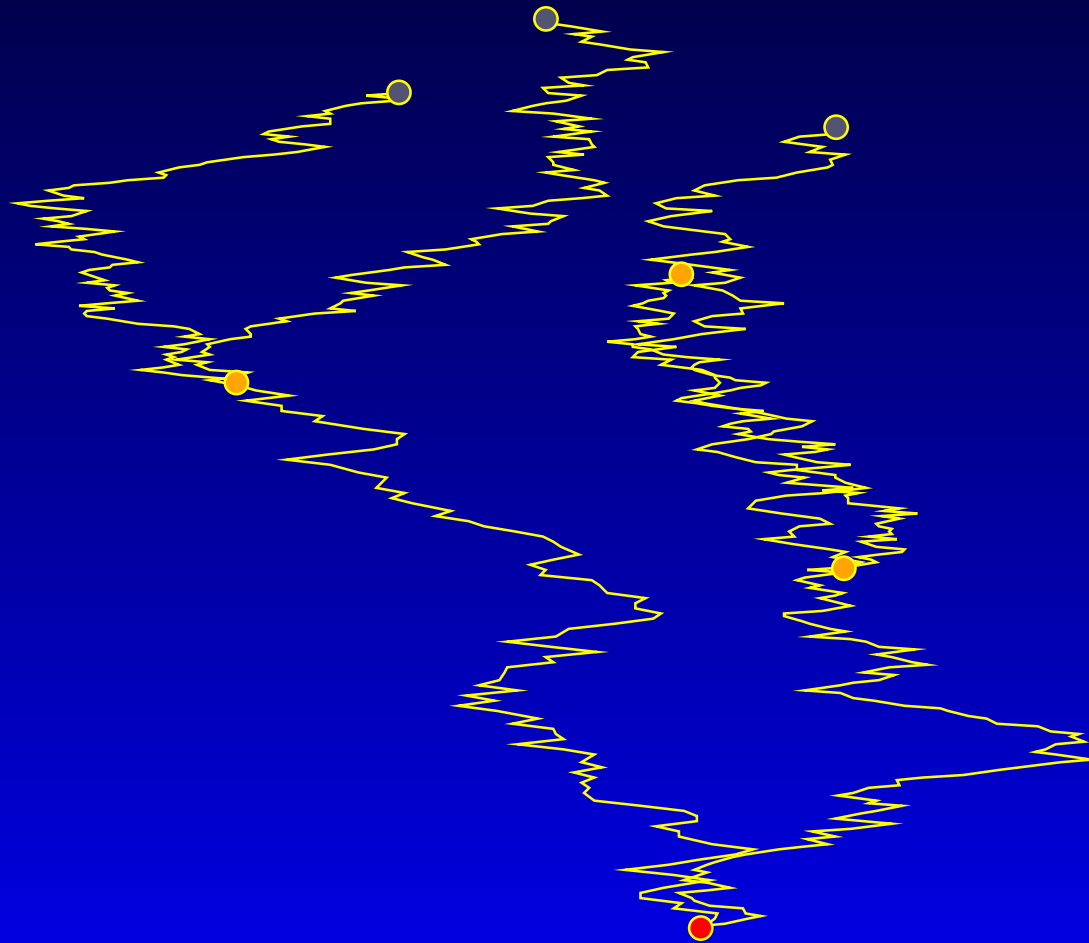
An outcome of Brownian motion on a 5-species tree

An outcome of Brownian motion on a 5-species tree

An outcome of Brownian motion on a 5-species tree

An outcome of Brownian motion on a 5-species tree
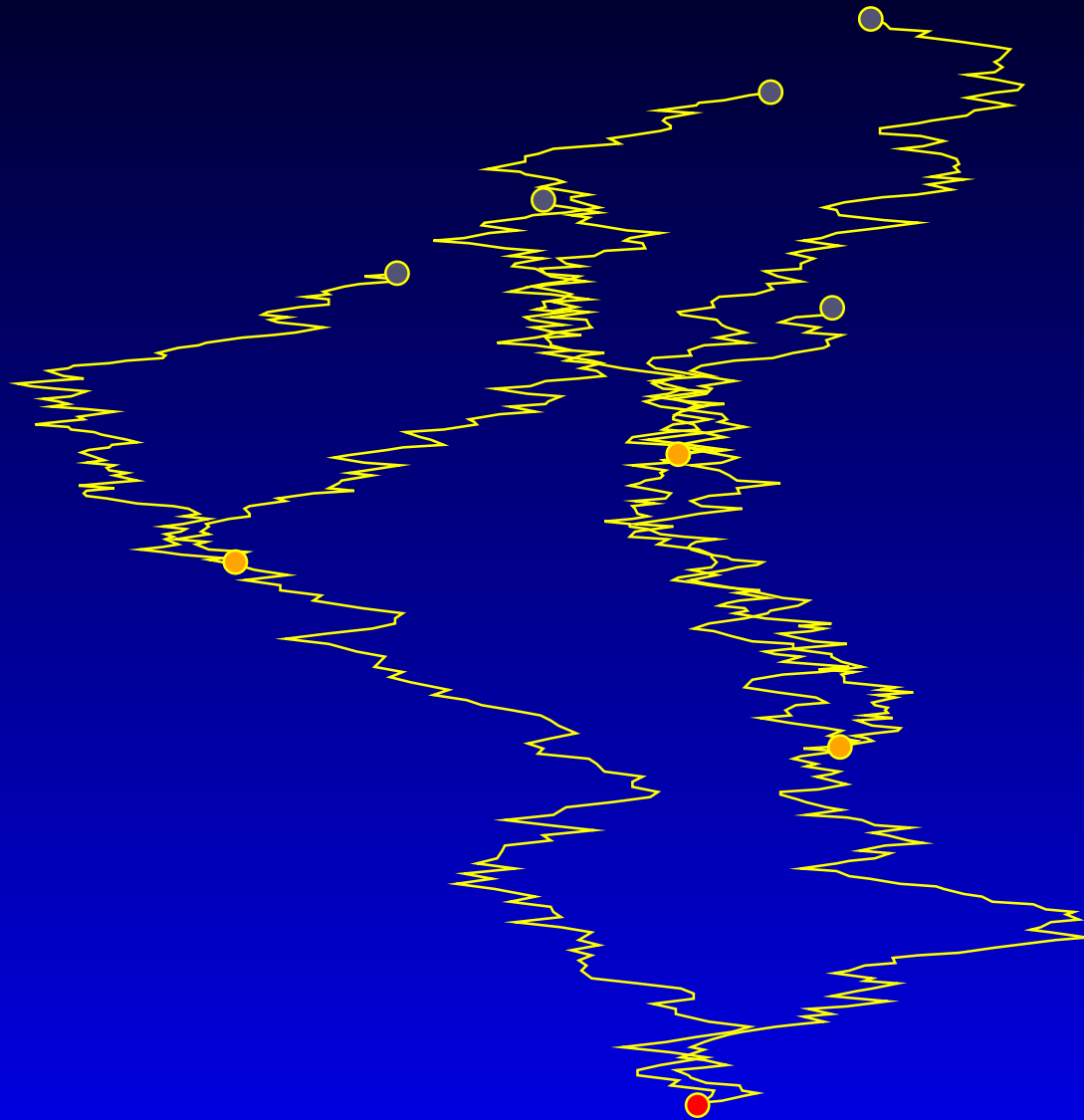
Likelihood under Brownian motion with two species

$$f\left(x; \mu, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$L = \prod_{i=1}^{p} \frac{1}{(2\pi)\sigma_i^2 \sqrt{v_1 v_2}} \exp\left(-\frac{1}{2\sigma_i^2}\left[\frac{(x_{1i} - x_{0i})^2}{v_1} + \frac{(x_{2i} - x_{0i})^2}{v_2}\right]\right)$$

Brownian motion along a tree
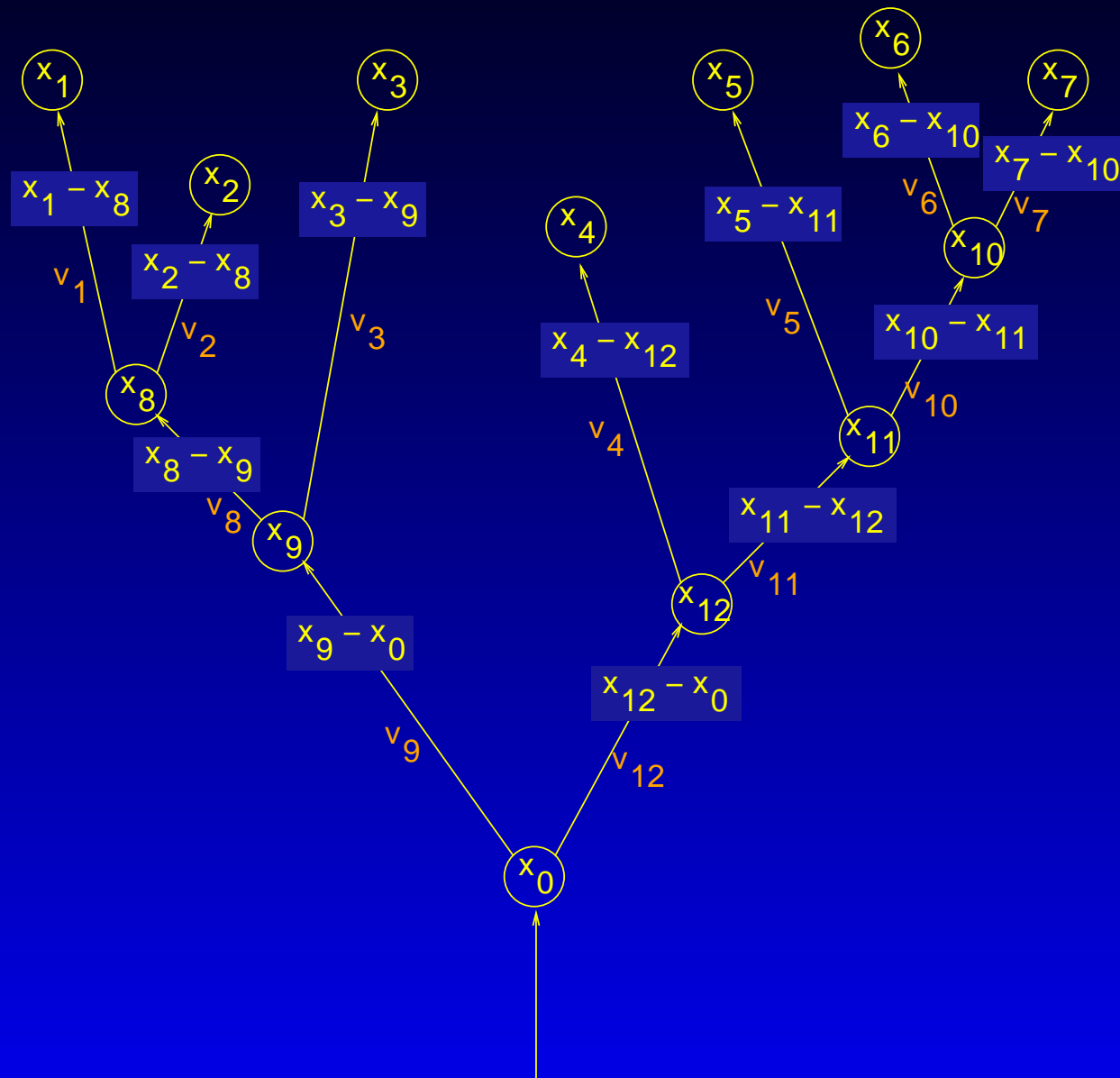
Covariances of species on the tree

$$
\begin{bmatrix}
v_1 + v_8 + v_9 & v_8 + v_9 & v_9 & 0 & 0 & 0 & 0 \\
v_8 + v_9 & v_2 + v_8 + v_9 & v_9 & 0 & 0 & 0 & 0 \\
v_9 & v_9 & v_3 + v_9 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & v_4 + v_{12} & v_{12} & v_{12} & v_{12} \\
0 & 0 & 0 & v_{12} & v_5 + v_{11} + v_{12} & v_{11} + v_{12} & v_{11} + v_{12} \\
0 & 0 & 0 & v_{12} & v_{11} + v_{12} & v_6 + v_{10} + v_{11} + v_{12} & v_{10} + v_{11} + v_{12} \\
0 & 0 & 0 & v_{12} & v_{11} + v_{12} & v_{10} + v_{11} + v_{12} & v_7 + v_{10} + v_{11} + v_{12}
\end{bmatrix}
$$

Covariances are of form

$$
\begin{bmatrix}
a & b & c & 0 & 0 & 0 & 0 \\
b & d & c & 0 & 0 & 0 & 0 \\
c & c & e & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & f & g & g & g \\
0 & 0 & 0 & g & h & i & i \\
0 & 0 & 0 & g & i & j & k \\
0 & 0 & 0 & g & i & k & l
\end{bmatrix}
$$

Brownian Motion Model

X_i

X_j

Model predicts that the *uncertainty* associated with the difference in trait values (i.e. $D_{ij} = X_i - X_j$) grows linearly with time. That is,

$$Var(D_{ij}) = (t_i + t_j) \, \sigma^2$$

So the longer you wait, the less you can predict about D_{ij}

t_i

t_j

X_k

Brownian Motion Model

X_i X_j

t_i t_j
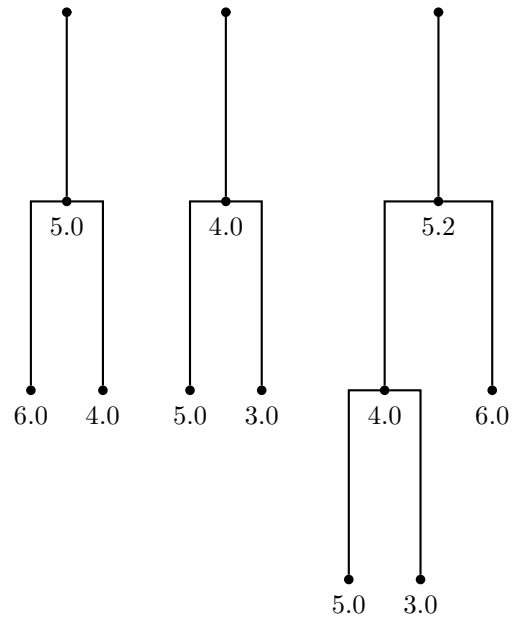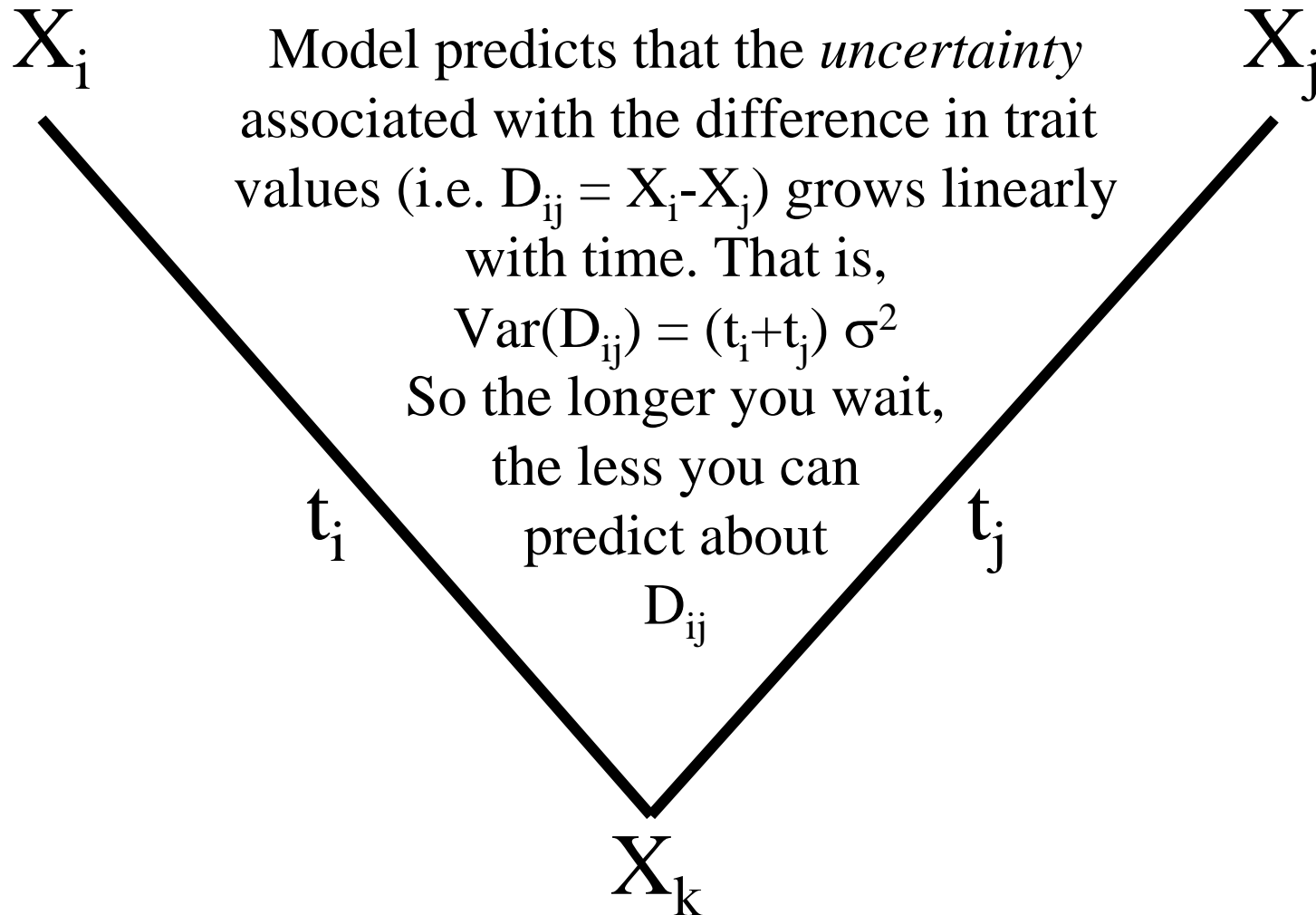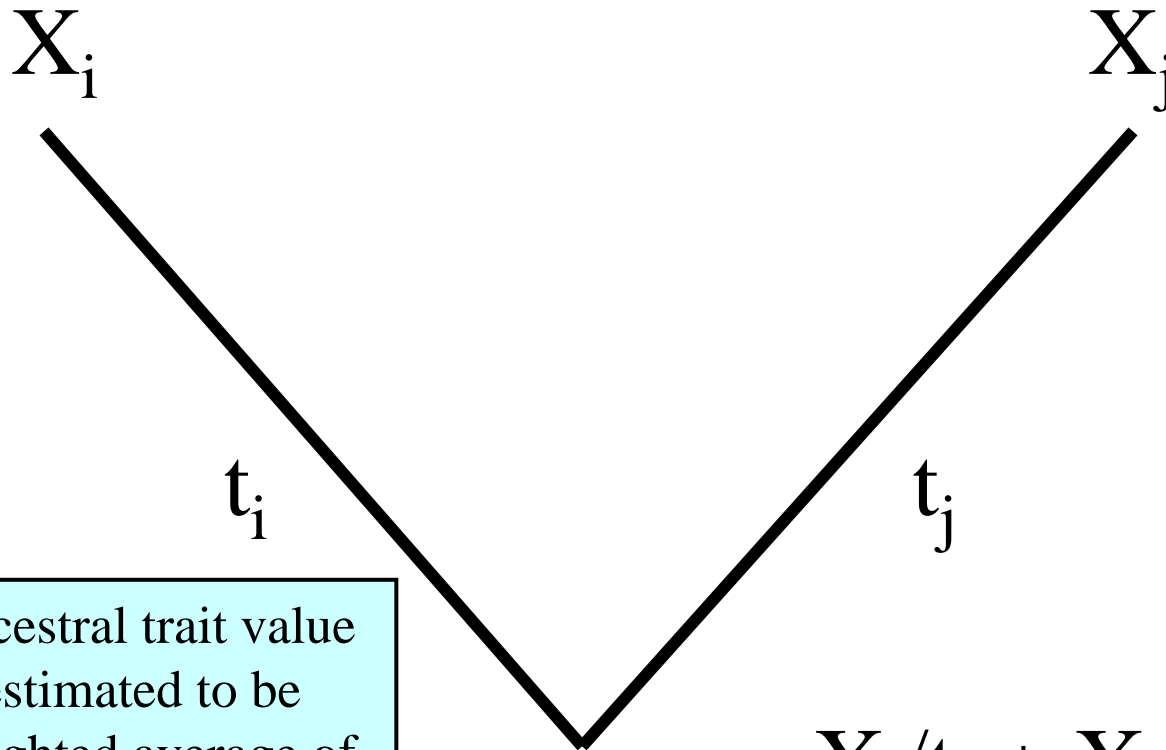
Ancestral trait value estimated to be weighted average of descendants' values. Weights are inverses of branch lengths (i.e. times)

$$X_k = \frac{X_i/t_i + X_j/t_j}{1/t_i + 1/t_j}$$

Brownian Motion Model

X_i X_j
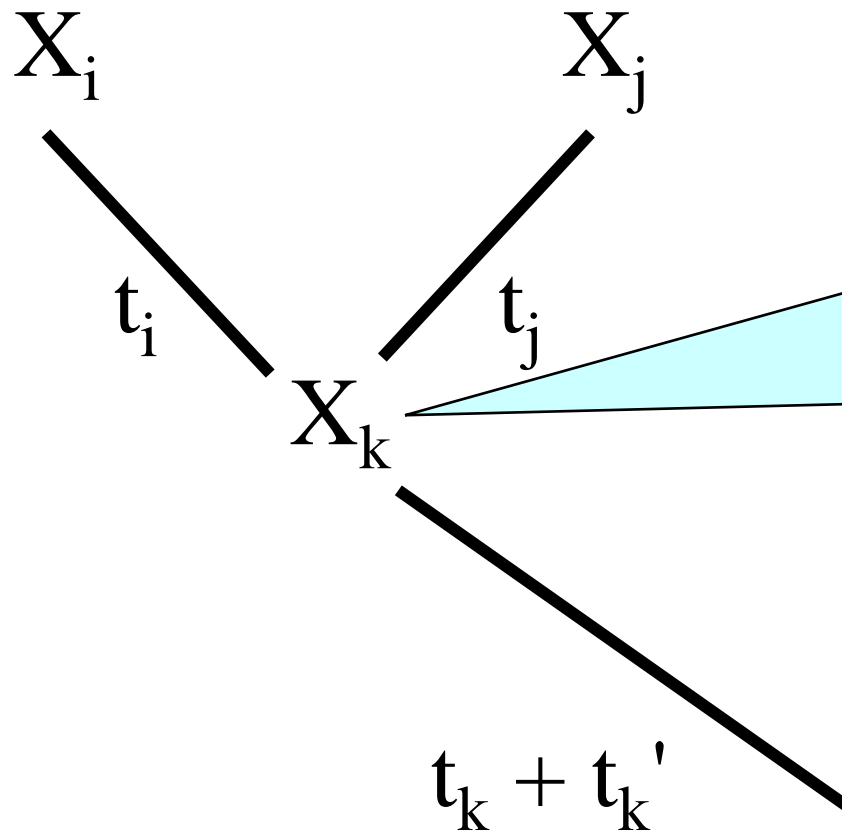
t_i t_j

X_k

$t_k + t_k'$

X_k is not observed, and thus has some extra uncertainty associated with its value. This extra uncertainty can be modeled by adding an extra bit of length (t_k') to the branch subtending X_k

The amount of extra uncertainty that should be added is:
$$t_k' = t_i\, t_j\, / \,(t_i + t_j)$$

Data for Two Characters on Tree

X	27	33	18	22
Y	122	124	126	128

A   B   C   D

2   2   2   2

E   F

3.5   3.5
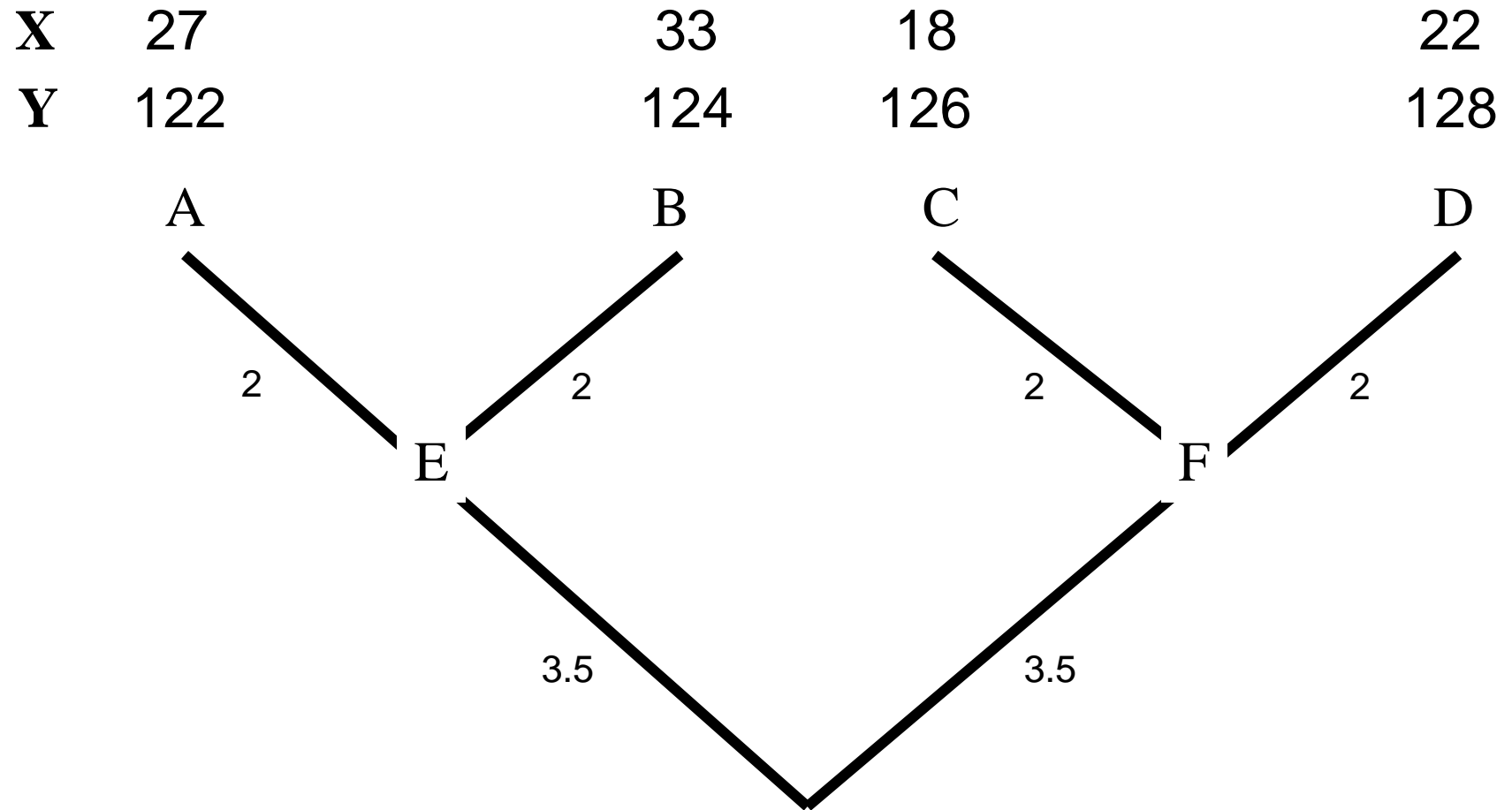
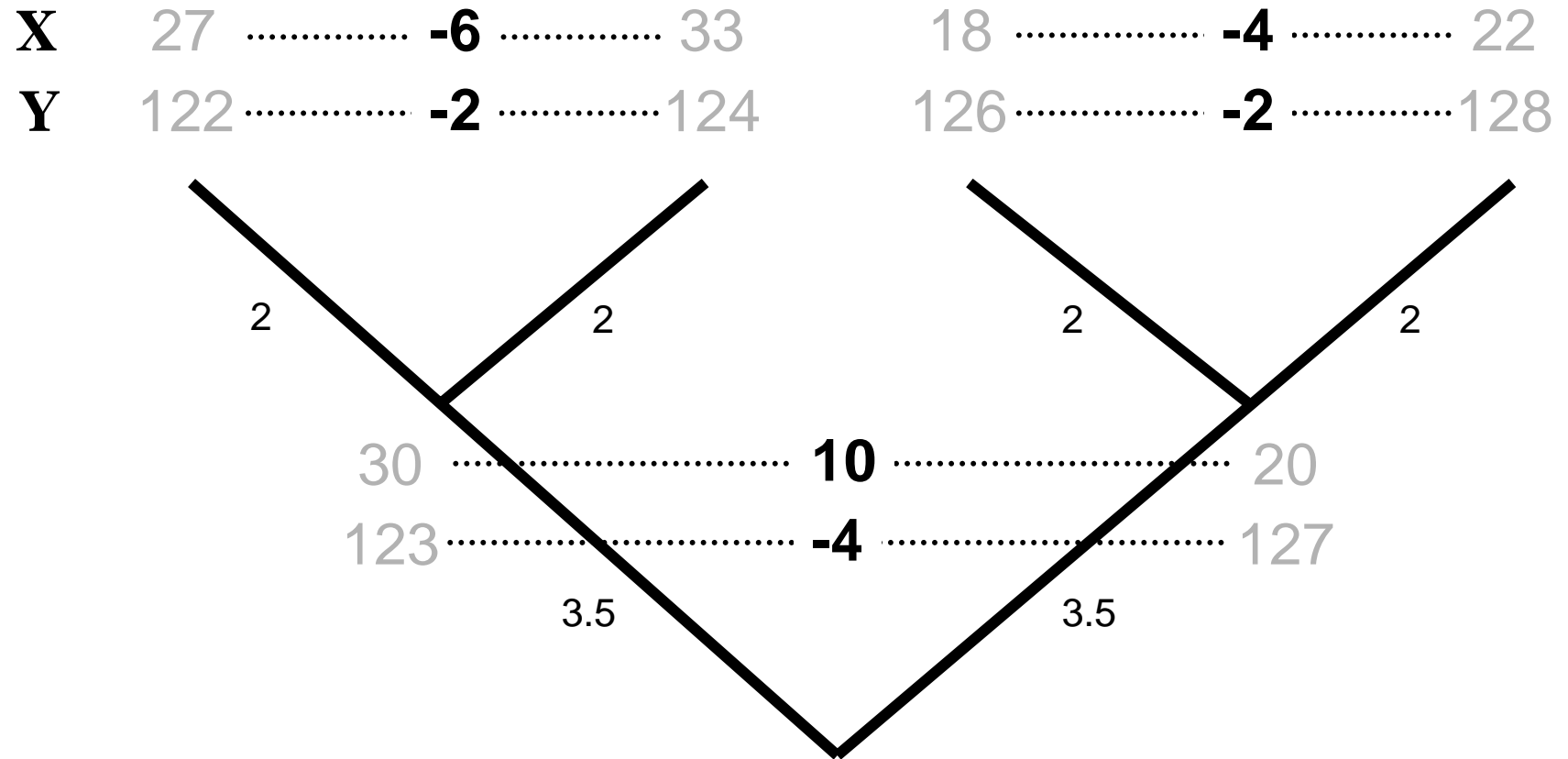# Contrasts (left minus right)

X   27 ·········· **-6** ·········· 33   18 ··········· **-4** ·········· 22

Y 122 ·········· **-2** ·········· 124 126 ·········· **-2** ·········· 128

2   2   2   2

30 ·········· **10** ·········· 20

123 ·········· **-4** ·········· 127

3.5   3.5

Scaling Contrasts

- A-B and C-D contrasts are expected to be on the same scale because the path length associated with both of these is 4

- E-F has a path length of 7, which means this contrast is expected to be larger

- Adding in the extra uncertainty associated with estimating E and F, this path length expands from 7 to 9

- Can put all 3 contrasts on same scale by dividing by standard deviation (square root of variance)

Rescaled Contrasts

Asterisks indicate that these original trait values have been scaled by dividing by the square root of the path length.

	X	Y	Variance proportional to	X*	Y*
A-B	-6	-2	4	-6/2 = -3	-2/2 = -1
C-D	-4	-2	4	-4/2 = -2	-2/2 = -1
E-F	10	-4	9	10/3 = 3.33	-4/3 = -1.33

These path lengths are proportional to the variance, and thus they are all that are needed to place the 3 contrasts for a given trait on the same scale.
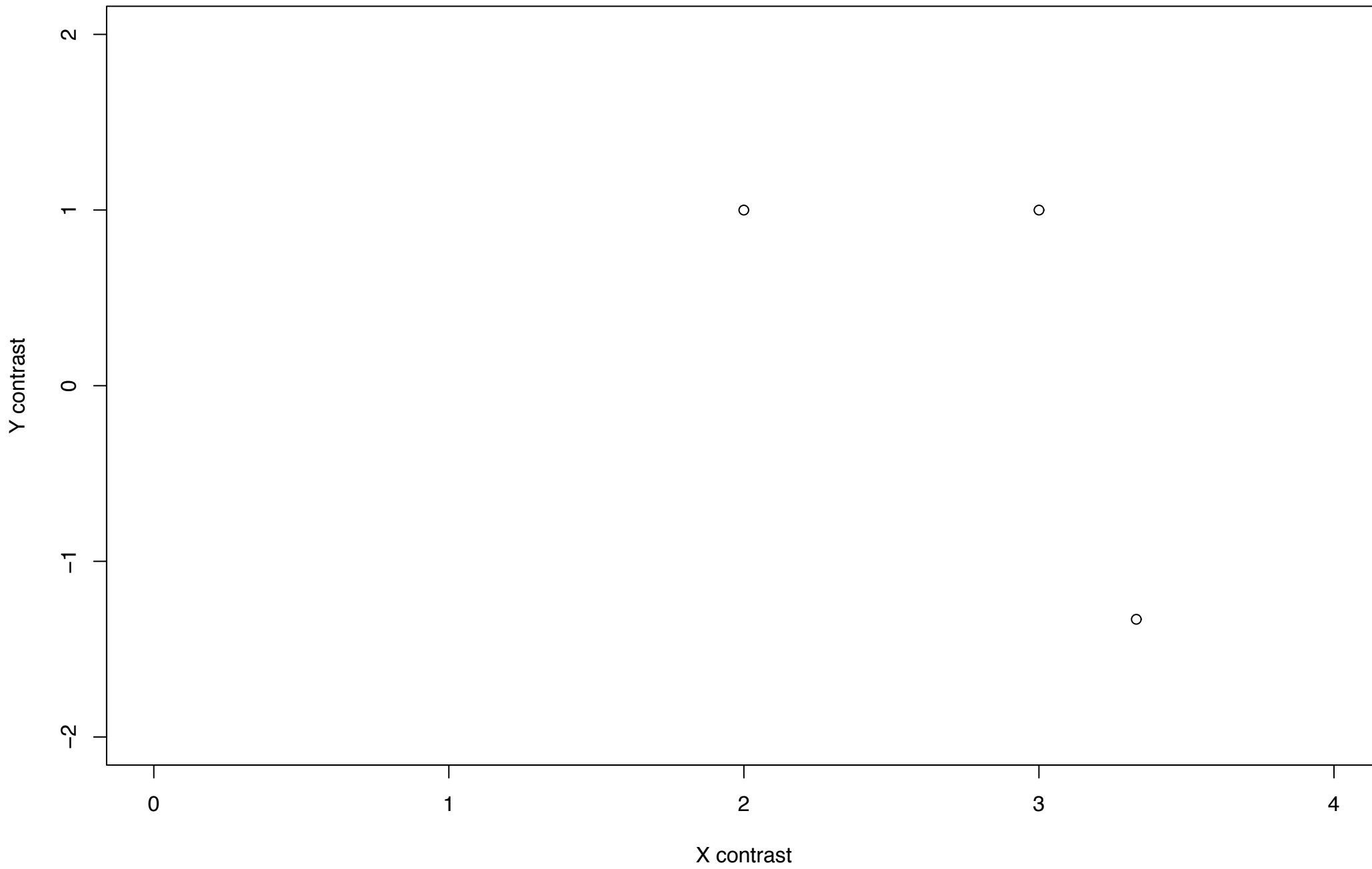
Correlation of Contrasts

	X	Y
A-B	-3	-1
C-D	-2	-1
E-F	3.33	-1.33

$$Var(X) = 8.0370$$
$$Var(Y) = 1.2593$$
$$Cov(XY) = 0.1852$$
$$Correlation = 0.05821$$

Correlation of the raw X and Y trait values was -0.5976, which is both stronger and of opposite sign. Note that the sample size is now 3 rather than 4.

The CONTRAST program in Joe Felsenstein's PHYLIP package performs independent contrasts