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:

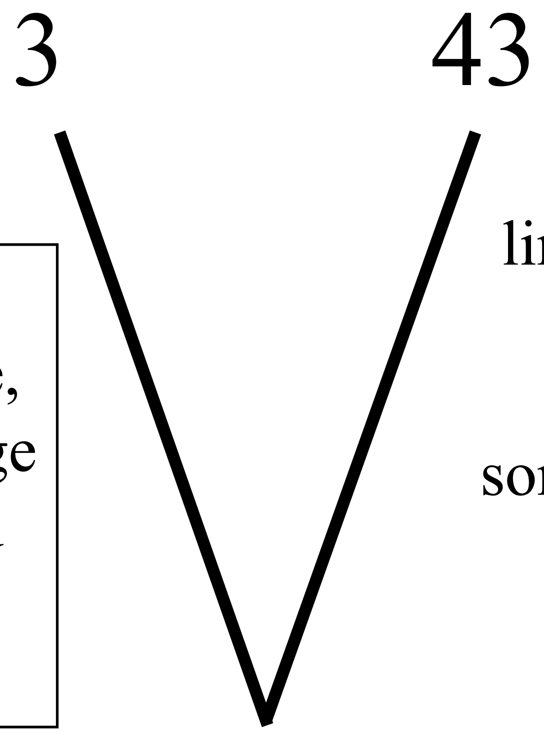
`http://hydrodictyon.eeb.uconn.edu/people/plewis`

Key Innovations

Clade contrast approach

Left-hand lineage
has few species

There is obviously a difference in diversity here, but is this difference so large that it cannot be explained using a model of random speciation?

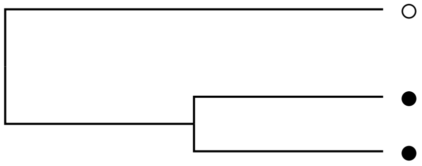


Right-hand lineage has many species, and is associated with some trait thought to favor species diversity

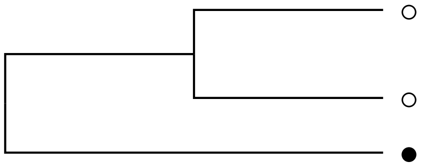
Slowinski, J. B., and C. Guyer. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. The American Naturalist 134:907-921.

Slowinski, J. B., and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification: an improved method based on a model of random speciation and extinction. The American Naturalist 142:1019-1024.

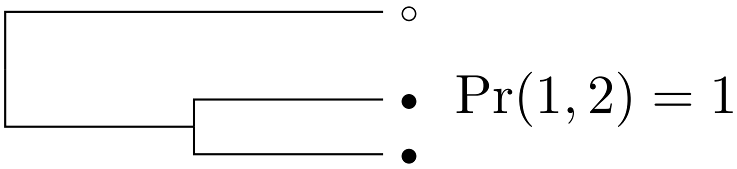
$\Pr(1, 1) = 1.0$

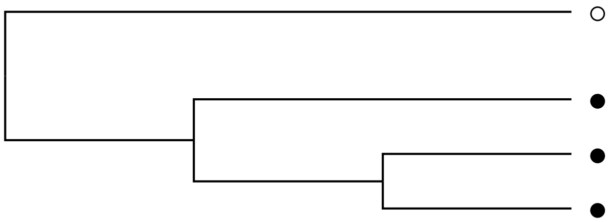


$\Pr(1,2) = 0.5$

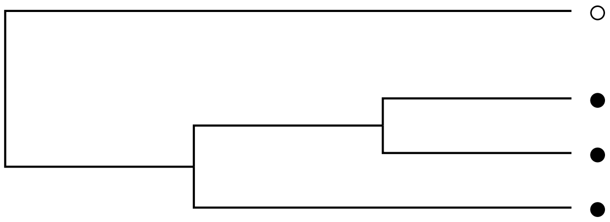


$\Pr(2,1) = 0.5$

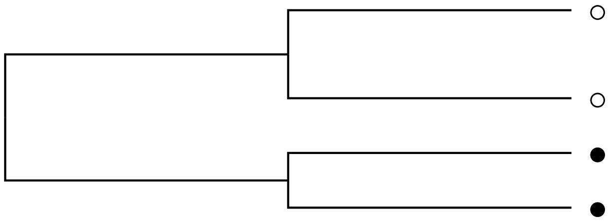

$\Pr(1, 2) = 1$



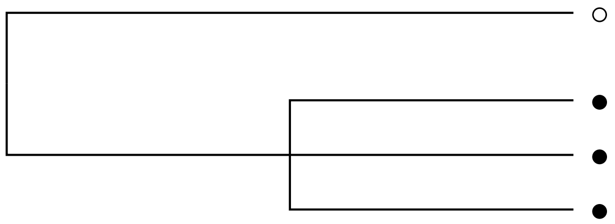
$$\Pr(1,3) = \frac{1}{3}$$



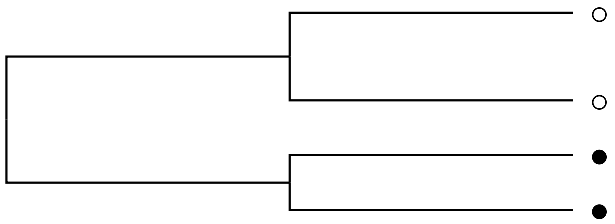
$$\Pr(1,3) = \frac{1}{3}$$



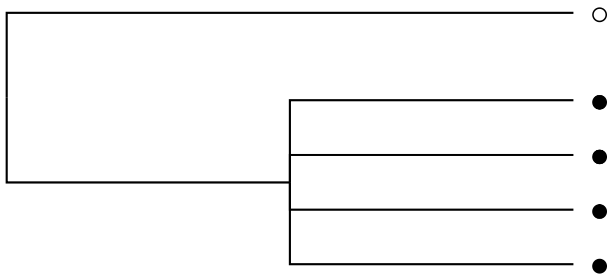
$$\Pr(2,2) = \frac{1}{3}$$



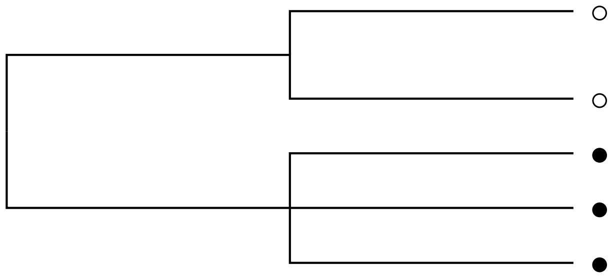
$$\Pr(1,3) + \Pr(3,1) = \frac{2}{3}$$



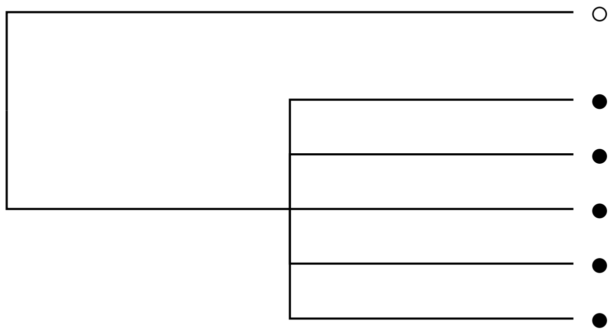
$$\Pr(2,2) = \frac{1}{3}$$



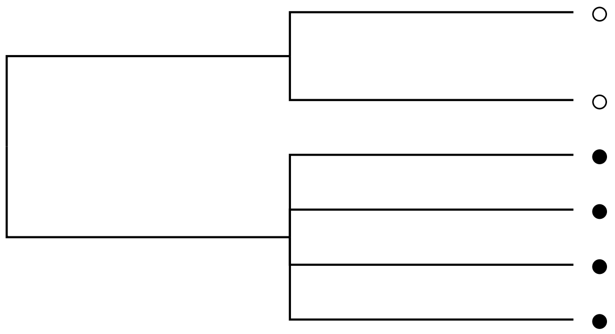
$$\Pr(1,4) + \Pr(4,1) = \tfrac{1}{2}$$



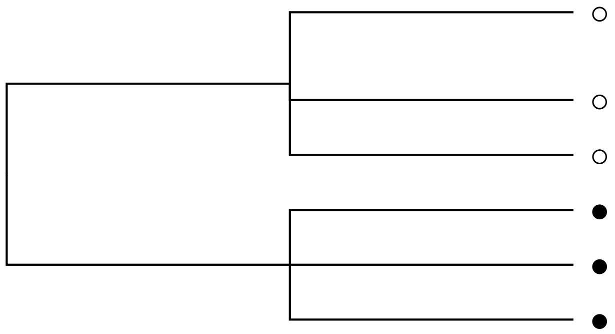
$$\Pr(2,3) + \Pr(3,2) = \tfrac{1}{2}$$



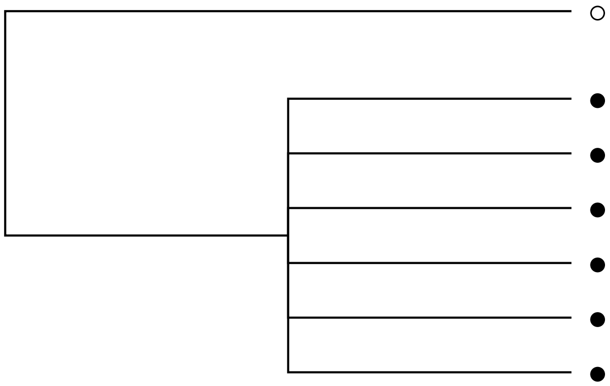
$$\Pr(1,5) + \Pr(5,1) = \frac{2}{5}$$



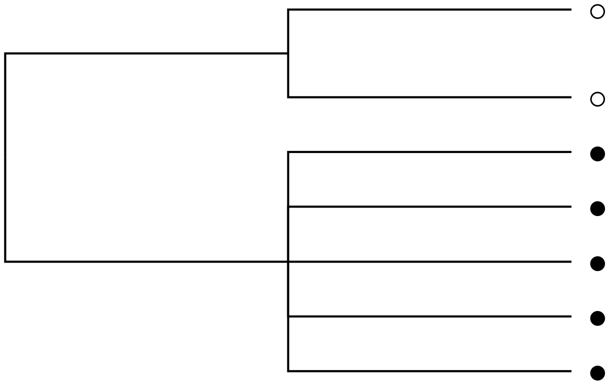
$$\Pr(2,4) + \Pr(4,2) = \frac{2}{5}$$



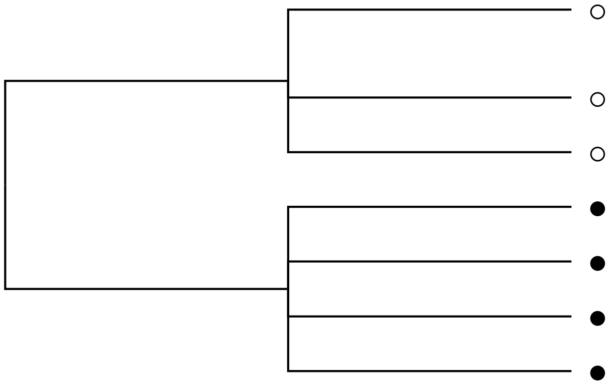
$$\Pr(3,3) = \frac{1}{5}$$



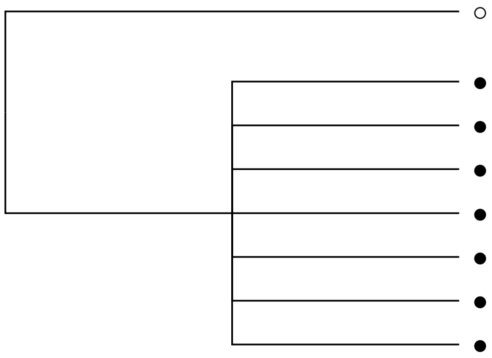
$$\Pr(1,6)+\Pr(6,1)=\tfrac{1}{3}$$



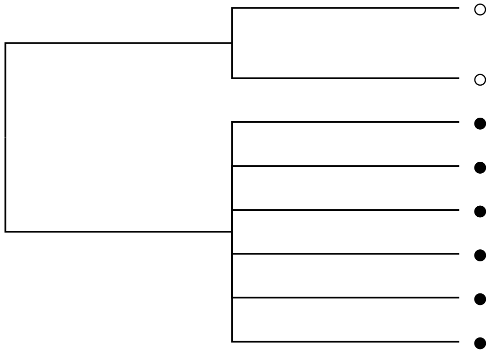
$$\Pr(2,5)+\Pr(5,2)=\tfrac{1}{3}$$



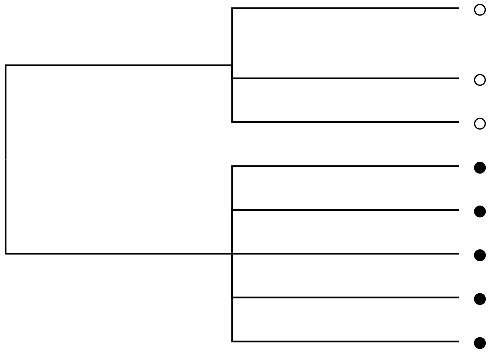
$$\Pr(3,4)+\Pr(4,3)=\tfrac{1}{3}$$



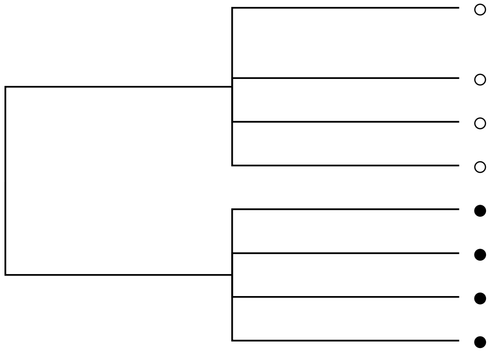
$$\Pr(1,7)+\Pr(7,1)=\tfrac{2}{7}$$



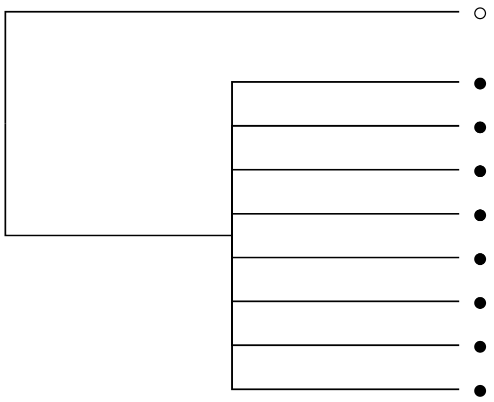
$$\Pr(2,6)+\Pr(6,2)=\tfrac{2}{7}$$



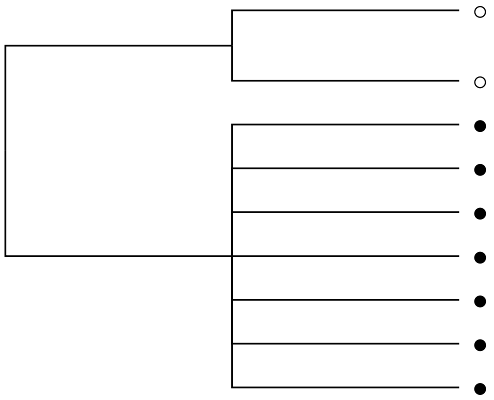
$$\Pr(3,5)+\Pr(5,3)=\tfrac{2}{7}$$



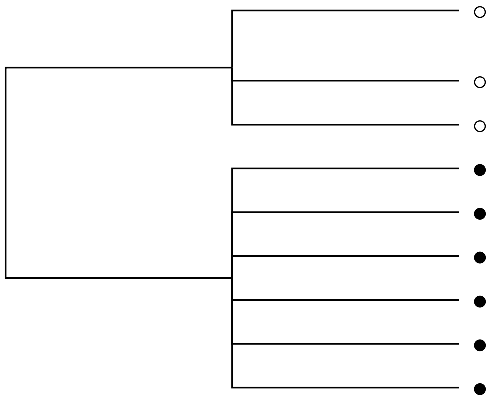
$$\Pr(4,4)=\tfrac{1}{7}$$



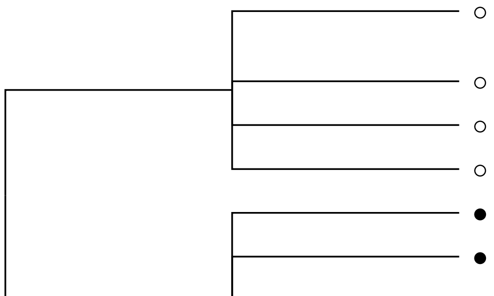
$$\Pr(1,8)+\Pr(8,1)=\frac{1}{4}$$



$$\Pr(2,7)+\Pr(7,2)=\frac{1}{4}$$



$$\Pr(3,6)+\Pr(6,3)=\frac{1}{4}$$



$$\Pr(4,5)+\Pr(5,4)=\frac{1}{4}$$

Key Innovations – clade size comparison

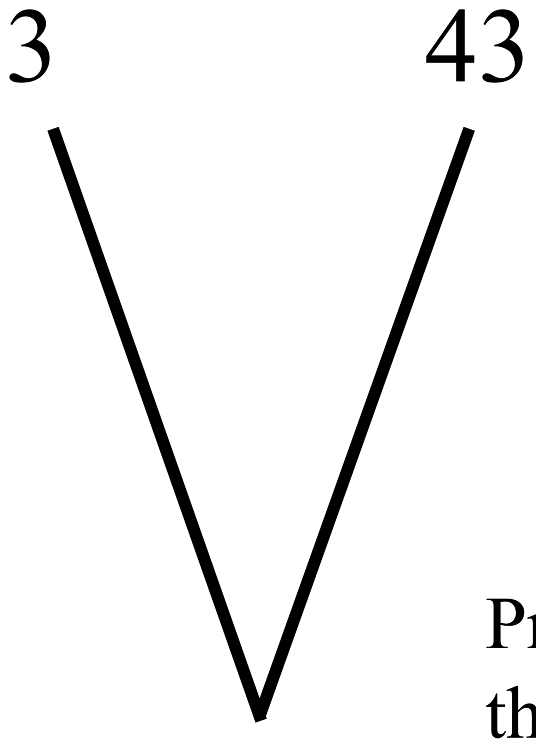
If you have an *a priori* reason to expect one state to lead to more species, then you can conduct the test as a one-tailed test.

This (roughly) divides the probabilities by one half.

$$\Pr(x, y) = \frac{1}{x + y - 1}$$

You have the pair of sister clades have a total of 46 species; 43 are in one clade and three are in another. What is the probability of seeing this much imbalance in clade size even if the character does not affect clade size?

Clade contrast approach



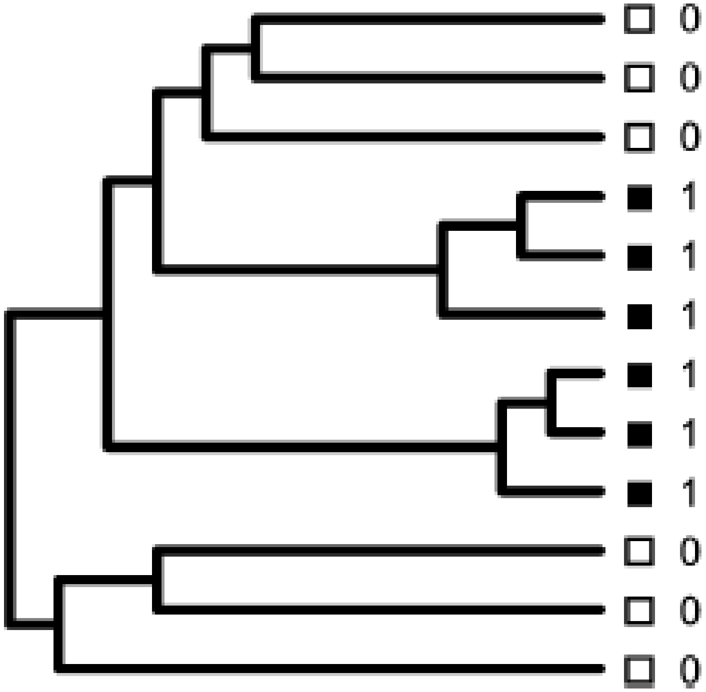
Observed:
Pr(3,43) = 1/(3+43-1) = 1/45

More extreme:
Pr(2,44) = 1/45
Pr(1,45) = 1/45

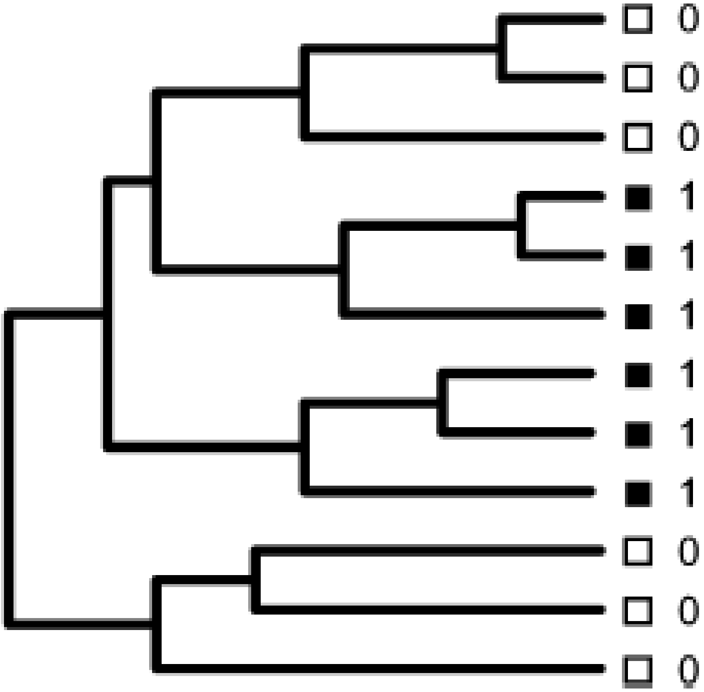
Probability of a contrast at least this extreme is 3/45 = 0.067 (n.s.)

Problem: test has low power (i.e. contrast needs to be quite extreme before this test can detect it). Also, assumes no reversals.

Ree's approach



Shorter average times between speciation events are associated with state 1 → **evidence** that state 1 is a key innovation



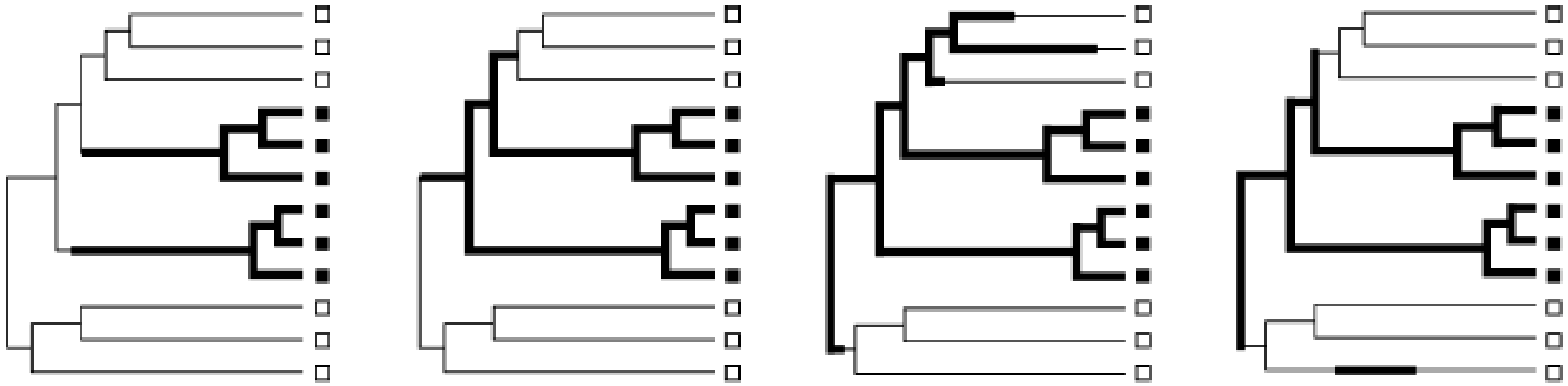
No association of internode times with state 1 → **no evidence** that state 1 is a key innovation

Fig. 1 from: Ree, R. H. 2005. Evolution 59(2): 257-265.

Ree's approach

- Question: is diversification rate elevated when state 1 is present in a lineage
- Answer involves three parts:
  - Decide when state 1 is present and when it is absent
  - Measure average diversification rate for both cases and form a contrast
  - Decide whether the contrast is significantly large based on null-model expectation

Part 1: When is state 1 present and when is it absent?

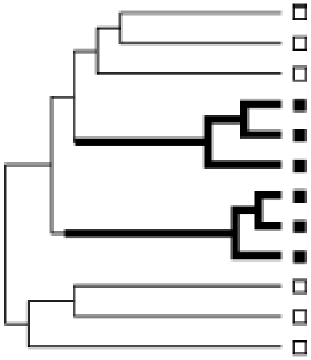


Use stochastic character mapping to determine this: the figure shows four mappings in which bold lines represent portions of lineages in which character state 1 was in effect

If state 1 is indeed a key innovation, then expect a higher rate of diversification in the bold segments of the tree.

Fig. 2 from: Ree, R. H. 2005. Evolution 59(2): 257-265.

Part 2: Measuring diversification rates



White state: 7 "speciations" over 70% of tree
Black state: 4 "speciations" over 30% of tree

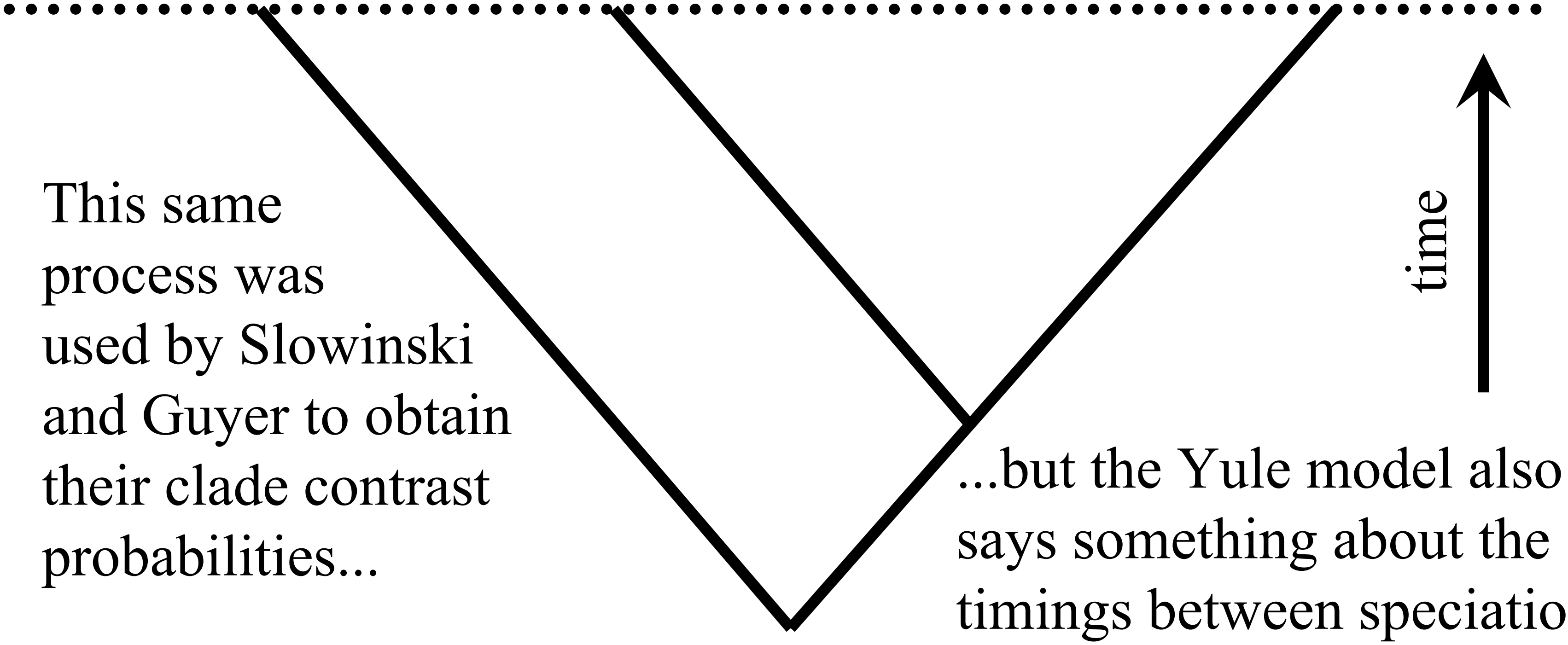
$d_0 = 7/0.7 = 10.00$

$d_1 = 4/0.3 = 13.33$

$d_1 - d_0 = 3.33$ ← is this a large number?

Part 3: Null model is the Yule process

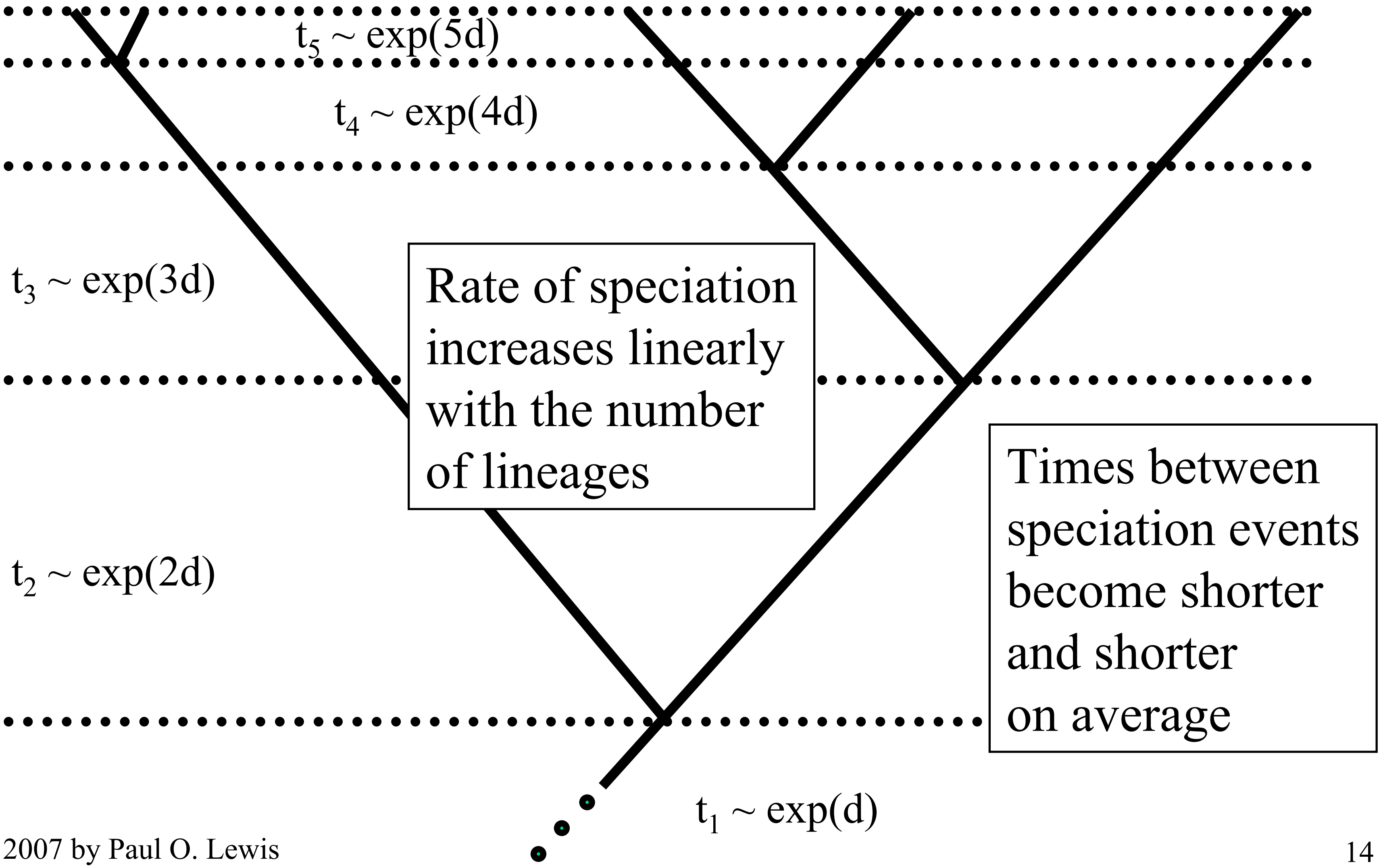
At every point in time, all lineages in existance have the same probability of speciating (no extinction allowed)



This same process was used by Slowinski and Guyer to obtain their clade contrast probabilities...

...but the Yule model also says something about the timings between speciations

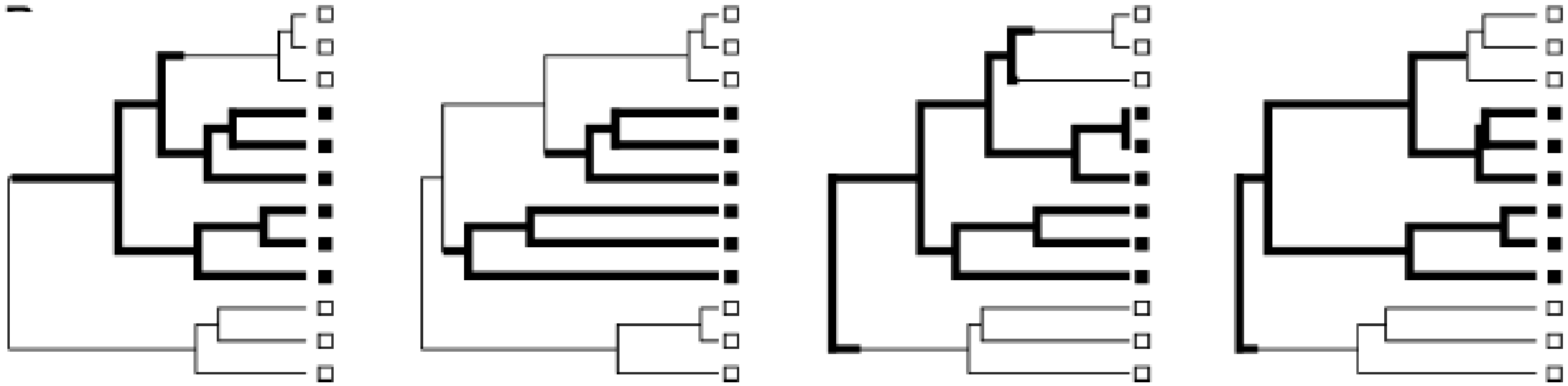
Part 3: Yule process



Rate of speciation increases linearly with the number of lineages

Times between speciation events become shorter and shorter on average

Part 3: Yule model null



- These stochastic mappings are on Yule trees
- Calculate contrasts $d_1 - d_0$ for these mappings just as you did for the stochastic mappings on the estimated trees
- Build up distribution of $d_1 - d_0$ contrasts from Yule trees to determine cutoff for statistical significance
- If original diversification rate contrast is in the upper 5% tail of this null distribution, then random speciation model cannot explain why diversification rate is higher when state 1 is present

Are columbine spurs a key adaptation?

Nectar-filled spurs are an adaptation to hummingbird pollination. Non-spurred species are hawkmoth pollinated

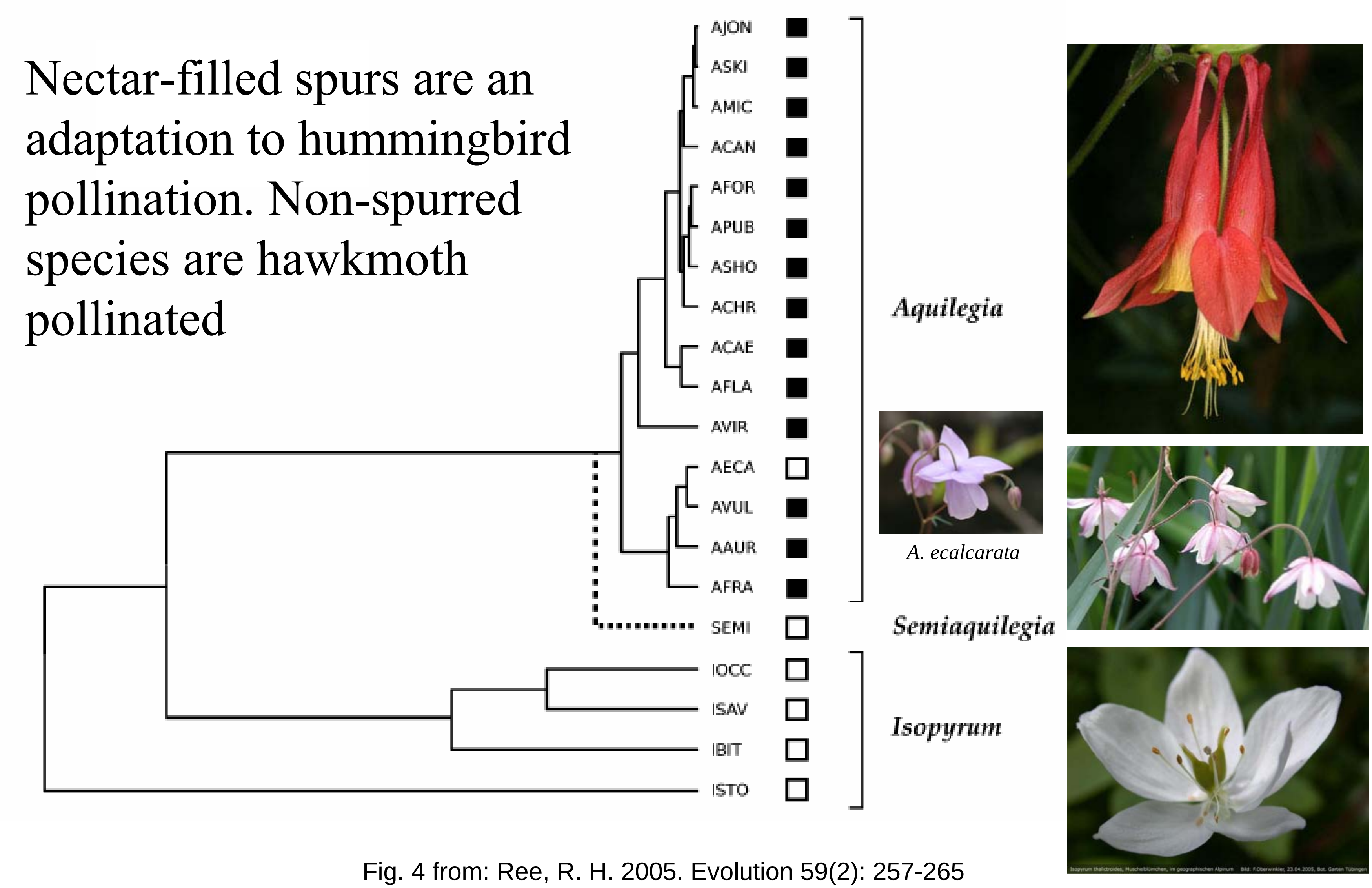


Fig. 4 from: Ree, R. H. 2005. Evolution 59(2): 257-265
Semiaquilegia image from http://www.flemings.com.au/australiangarden/plants.htm
Isopyrum thalictroides from http://www.botgarden.uni-tuebingen.de/tiki-index.php?page=Isopyrum
Aquilegia ecalcarata http://www.botanic.jp/plants-ha/fuodam.htm

Columbine example

- Used relaxed clock method in Beast program to sample trees with branch lengths proportional to divergence times
- Slowinski & Guyer clade contrast method yields Pr(1,15) = 1/15 = 0.067 (note that homoplasy is a complication)
- Ree's approach yielded P-value = 0.0006 (appears to be much more powerful and homoplasy is not a complicating factor)

Potential weakness of Ree's approach

- When testing for character correlations, stochastic character mapping can be weaker than Pagel's method because the model used to infer the mapping assumes independence.
- In Ree's approach the stochastic character mapping is done using a model that assumes that state changes are independent of the probability of cladogenesis.
- A potentially more powerful approach is to use a model that allows speciation and extinction rates to vary depending on a character state.
- As is often the case: being less powerful may make Ree's approach more robust!

BiSSE model of Maddison et al. (2007)

Calculate the probability of tree shape and character distribution:

$$\Pr(X, T, \boldsymbol{\nu} | \theta)$$

rather than:

$$\Pr(X, | T, \boldsymbol{\nu}, \theta)$$

which is done by assuming that the evolution of a character is independent of tree shape.

BiSSE model of Maddison et al. (2007)

μ_0 the extinction rate of a species that displays character state 0

μ_1 the extinction rate of a species that displays character state 1

λ_0 the speciation rate of a species that displays character state 0

λ_1 the speciation rate of a species that displays character state 1

q_{01} the rate of $0 \longrightarrow 1$ transitions.

q_{10} the rate of $1 \longrightarrow 0$ transitions.

BiSSE model of Maddison et al. (2007)

Sweep tip-to-root. $D_{N0}(t+\Delta t)$ is the probability of an species with character state 0 at time $t+\Delta t$ being the ancestor of a particular clade of N taxa at time 0.

$D_{N0}(t+\Delta t)$ is $(1-\mu_0\Delta t)$ times the sum of:

$\Pr(\text{No changes in } \Delta t)$	$(1-q_{01}\Delta t)(1-\lambda_0\Delta t)D_{N0}(t)$
$\Pr(\text{state change in } \Delta t)$	$(q_{01}\Delta t)(1-\lambda_0\Delta t)D_{N1}(t)$
$\Pr(\text{Spec. + extinct. in } \Delta t)$	$(1-q_{01}\Delta t)(\lambda_0\Delta t)E_0(t)D_{N0}(t)$
$\Pr(\text{Spec. + extinct. in } \Delta t)$	$(1-q_{01}\Delta t)(\lambda_0\Delta t)E_0(t)D_{N0}(t)$

BiSSE model of Maddison et al. (2007)

$E_0(t + \Delta t)$ is the probability of an species with character state 0 at time $t + \Delta t$ giving rise to no descendants at time 0.
$E_0(t + \Delta t)$ is the sum of:

$\mathrm{Pr}($Extinction in $\Delta t)$	$\mu_0 \Delta t$
$\mathrm{Pr}($No changes in $\Delta t)$	$(1 - \mu_0 \Delta t)(1 - q_{01} \Delta t)(1 - \lambda_0 \Delta t) E_0(t)$
$\mathrm{Pr}($State change in $\Delta t)$	$(1 - \mu_0 \Delta t)(q_{01} \Delta t)(1 - \lambda_0 \Delta t) E_1(t)$
$\mathrm{Pr}($Spec. in $\Delta t)$	$(1 - \mu_0 \Delta t)(1 - q_{01} \Delta t)(\lambda_0 \Delta t) E_0(t)^2$

BiSSE model of Maddison et al. (2007)

Initial conditions:

A $D_{10}(0) = 1$ term for every tip that has state 0
A $D_{11}(0) = 1$ term for every tip that has state 1
$E_0(0) = E_1(0) = 0$

You can use likelihood ratios to test if $\lambda_0 = \lambda_1$ and $\mu_0 = \mu_1$.

Implemented (by Peter Midford) in Mesquite.

Continuous characters

Fig. 6 from Felsenstein (1985)

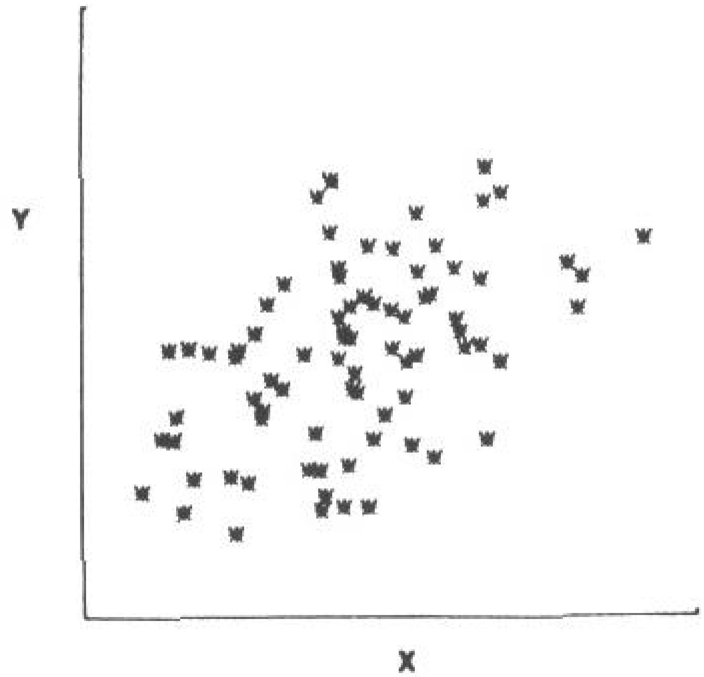


FIG. 6.—A typical data set that might be generated for the phylogeny in fig. 5 using the model of independent Brownian motion (normal increments) in each character.

Fig. 7 from Felsenstein (1985)

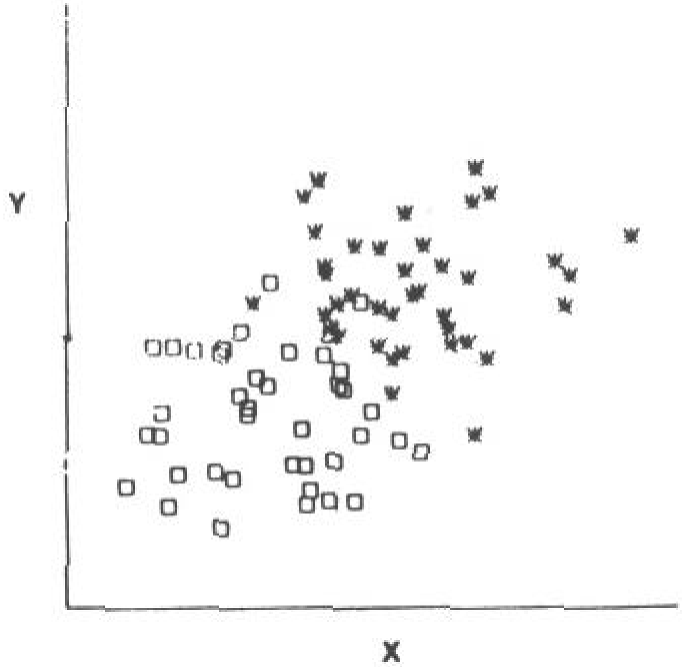


FIG. 7.—The same data set, with the points distinguished to show the members of the 2 monophyletic taxa. It can immediately be seen that the apparently significant relationship of fig. 6 is illusory.

Fig. 5 from Felsenstein (1985)

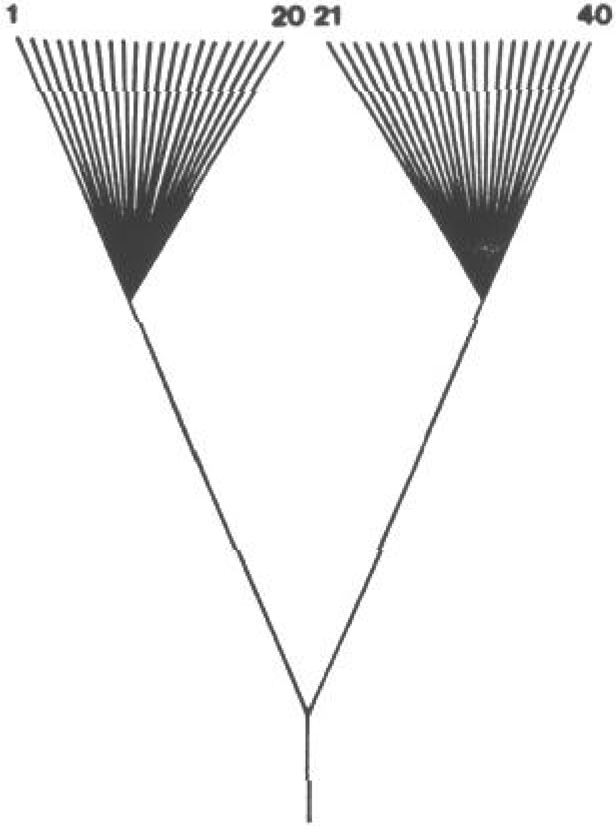
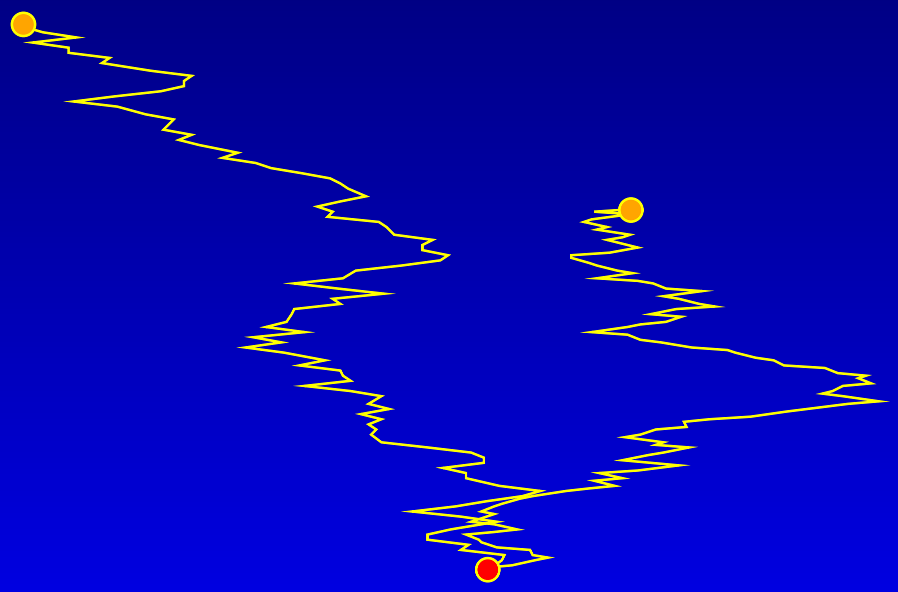


FIG. 5.—A "worst case" phylogeny for 40 species, in which there prove to be 2 groups each of 20 close relatives.

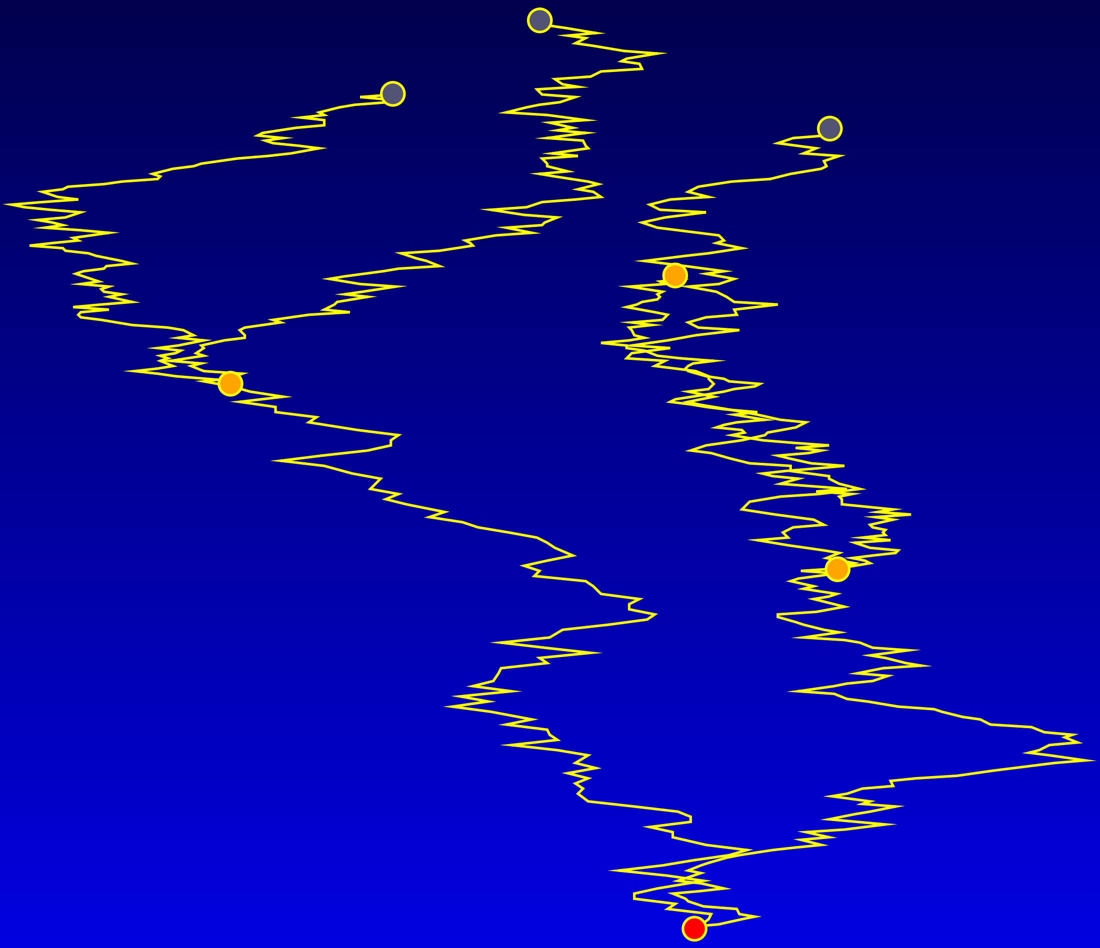
Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist 125:1-15.

An outcome of Brownian motion on a 5-species tree

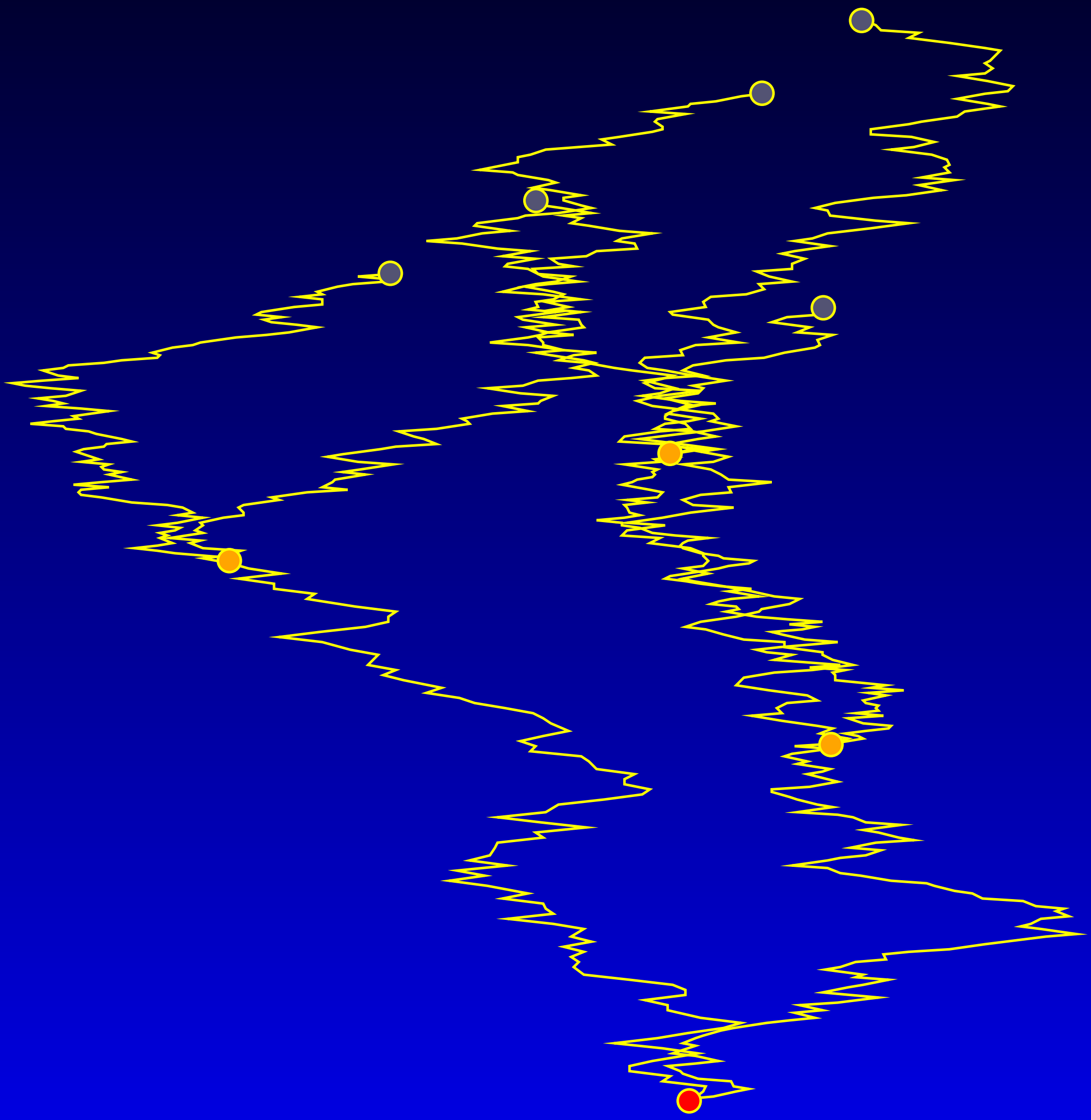
An outcome of Brownian motion on a 5-species tree



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

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

$$L \;=\; \prod_{i=1}^{p}\frac{1}{(2\pi)\sigma_i^2\sqrt{v_1v_2}}\exp\left(-\frac{1}{2\sigma_i^2}\left[\frac{\left(x_{1i}\;-\;x_{0i}\right)^2}{v_1}+\frac{\left(x_{2i}\;-\;x_{0i}\right)^2}{v_2}\right]\right)$$

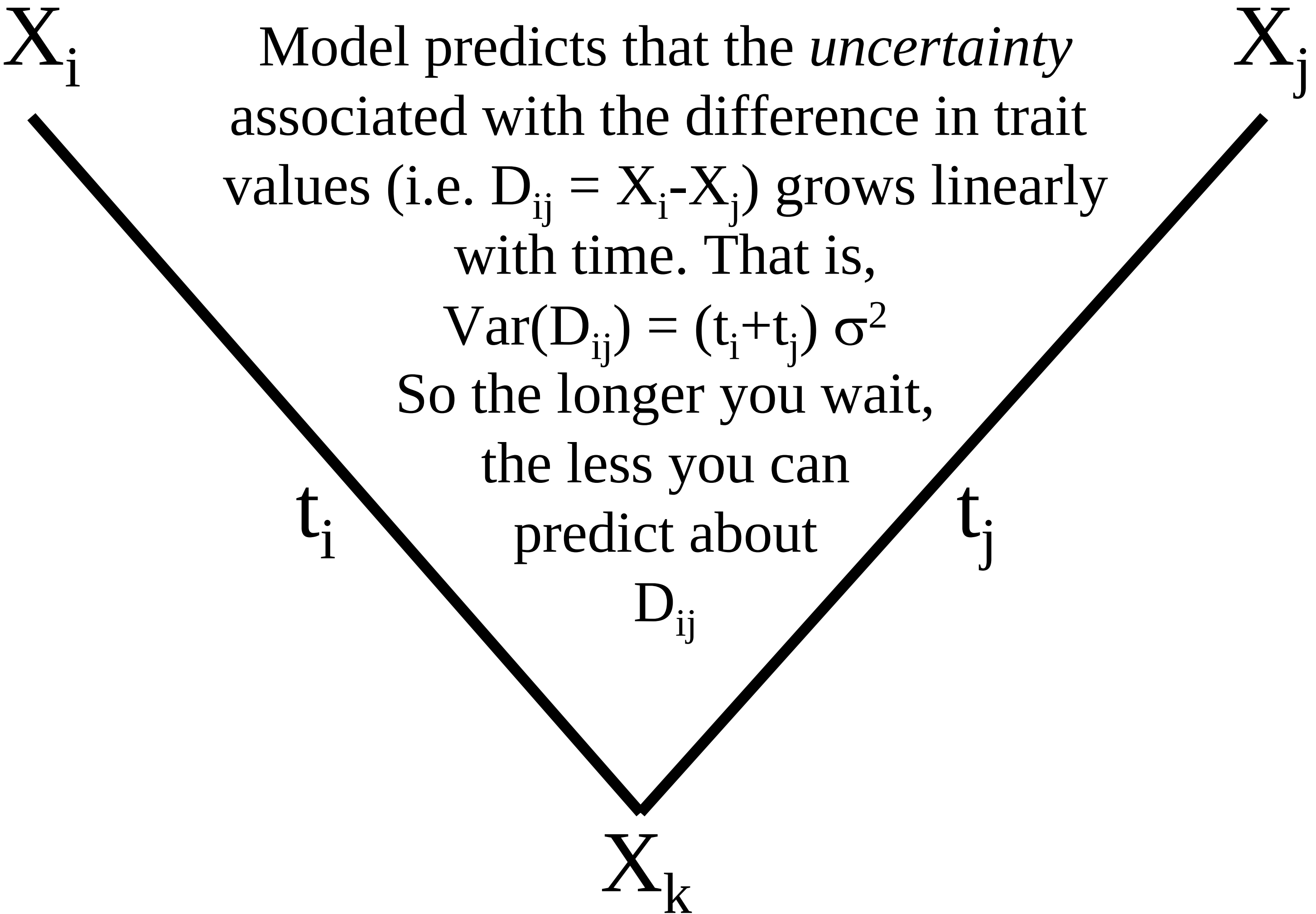
Data for Two Characters, X and Y

	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...

Brownian Motion Model

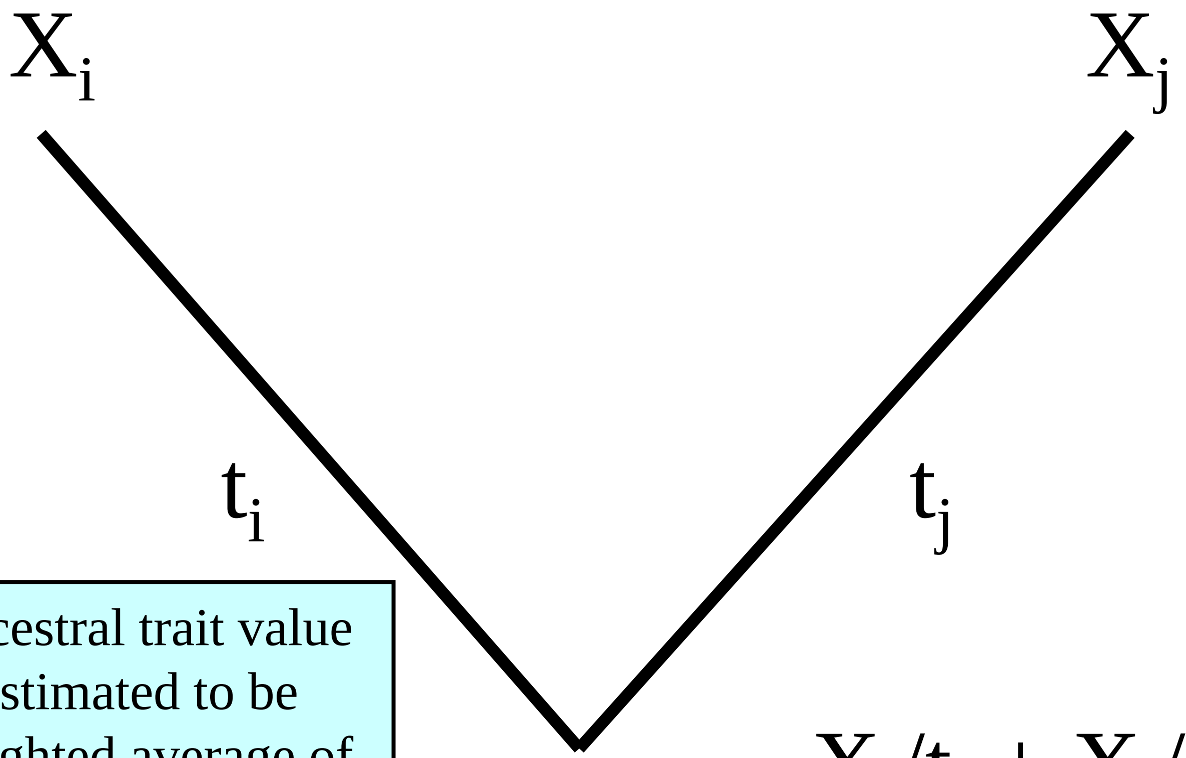


Model predicts that the *uncertainty* associated with the difference in trait values (i.e. $D_{ij} = X_i - X_j$) grows linearly with time. That is,

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

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

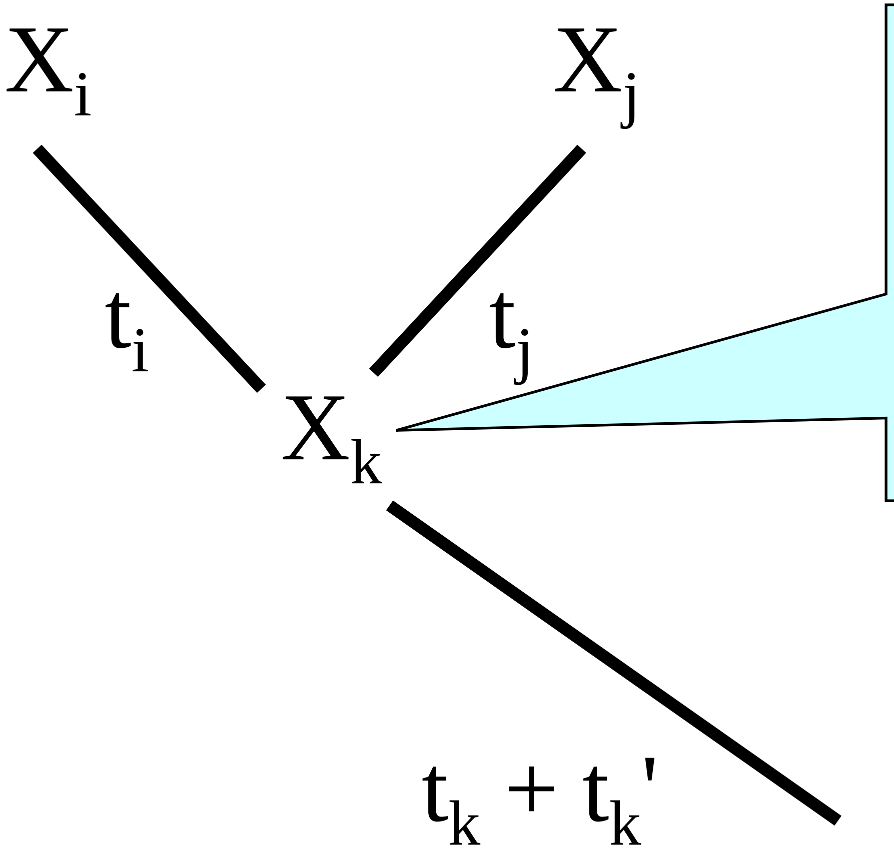
Brownian Motion Model



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

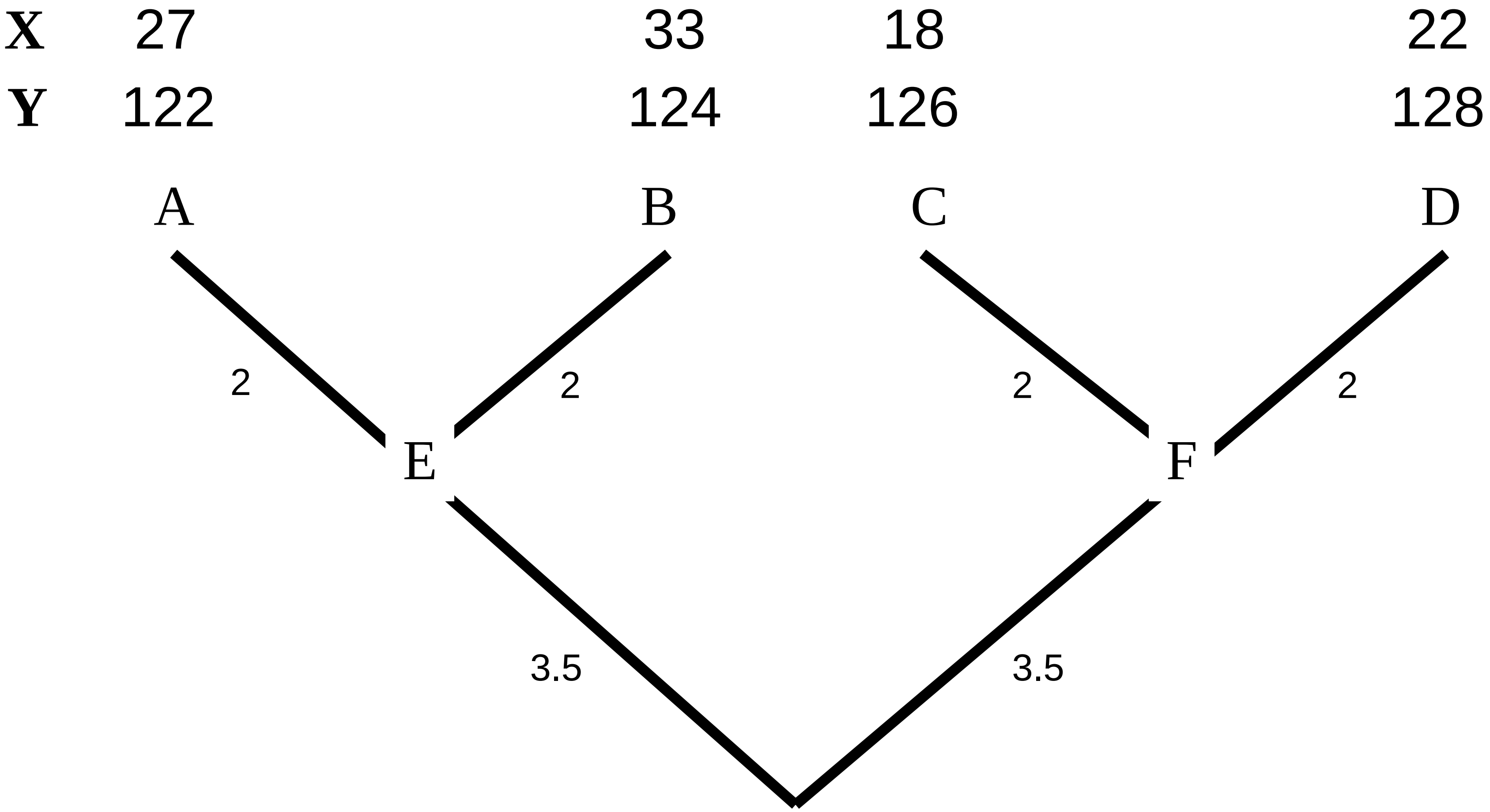


X_k is not observed, and thus has some extra uncertainty associated with its value. This extra uncertainty can be modeled by adding an extra bit of length (t_k') to the branch subtending X_k

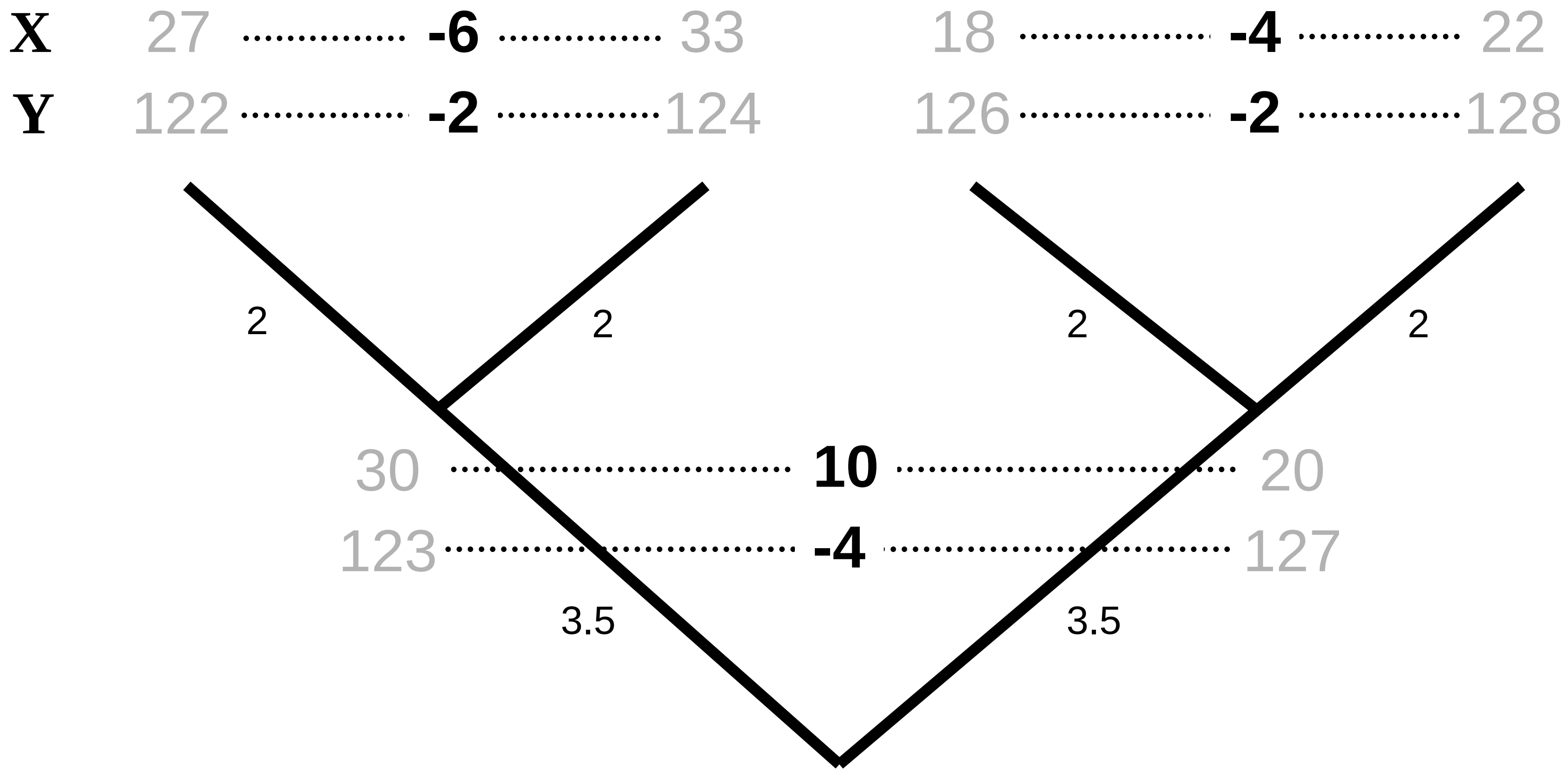
The amount of extra uncertainty that should be added is:

$$t_k' = t_i \, t_j / (t_i + t_j)$$

Data for Two Characters on Tree



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

Estimating phylogenies from continous traits

- Felsenstein, J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. *American Journal of Human Genetics* 25: 471-492.
- Same brownian motion model used for independent contrasts method
- CONTML program in the PHYLIP package: `http://evolution.genetics.washington.edu/phylip.html`

References

Maddison, W., Midford, P. E., and Otto, S. E. (2007). Estimating a binary character's effect on speciation and extinction. systematic biology. *Systematic Biology*, 56(5):701–710.