Statistical nonmolecular phylogenetics: can molecular phylogenies illuminate morphological evolution?

28 July 2012.

Joe Felsenstein

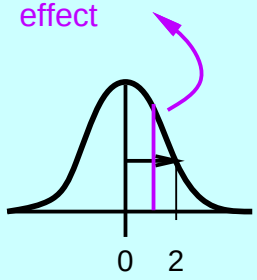
Workshop on Molecular Evolution, MBL, Woods Hole

Where this lecture fits in

We are heading (rapidly) into more complete integration of work on molecular evolution with work on between-species differences of measurable characters, and into more integration of within-species work on these characters with both of these.

A standard quantitative genetics model

$$P = \begin{Bmatrix} AA & 2 \\ Aa & 4 \\ aa & 7 \end{Bmatrix} + \begin{Bmatrix} BB & 0.6 \\ Bb & 0.1 \\ bb & -0.2 \end{Bmatrix} + \begin{Bmatrix} CC & -1 \\ Cc & 6 \\ cc & 6 \end{Bmatrix} + \begin{Bmatrix} DD & 0.3 \\ Dd & 0.3 \\ dd & 0.7 \end{Bmatrix} + \begin{Bmatrix} EE & -0.4 \\ Ee & 0.3 \\ ee & -0.3 \end{Bmatrix} + \text{environmental effect}$$



AA Bb Cc dd Ee 10

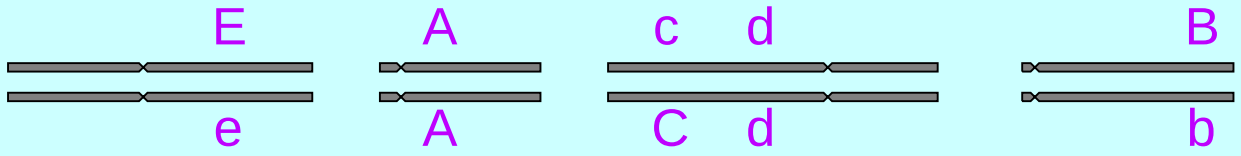
Aa bb cc DD ee

aa bb CC DD Ee

aa bb Cc DD EE

Aa Bb Cc DD Ee

0.3 + 2 + 6 + 0.7 + 0.1 + 0.9



A model of quantitative characters on a phylogeny

- We can model character change on a phylogeny as Brownian motion with multiple characters with different variances and with covariation as well.

A model of quantitative characters on a phylogeny

- We can model character change on a phylogeny as Brownian motion with multiple characters with different variances and with covariation as well.
- This started with approximating gene frequencies in the 1960s by Anthony Edwards and Luca Cavalli-Sforza.

A model of quantitative characters on a phylogeny

- We can model character change on a phylogeny as Brownian motion with multiple characters with different variances and with covariation as well.
- This started with approximating gene frequencies in the 1960s by Anthony Edwards and Luca Cavalli-Sforza.
- I expanded it to model quantitative characters determined by these genes.

Models for long-term evolution

The use of quantative genetics approximations to model long-term evolution in lineages was largely introduced by Russ Lande in the 1980s.

Russell Lande, from his website at Imperial College, U.K., where he has been in recent years.

Where do the covariances come from?

- Genetic covariances (the same loci affect two or more traits). Genetic drift or natural selection can change the gene frequencies at these loci, and thus make correlated changes in the two traits.

Where do the covariances come from?

- Genetic covariances (the same loci affect two or more traits). Genetic drift or natural selection can change the gene frequencies at these loci, and thus make correlated changes in the two traits.
- Selective covariances (Olof Tedin, 1926; G. Ledyard Stebbins 1950) The same environmental conditions can select changes in two or more traits – even though they may have no genetic covariance. This source of evolutionary covariance is widely ignored.

Part 1

Morphometrics and phylogenies

Fred Bookstein

... is co-author on this part of the talk

How to use morphometric coordinates on phylogenies?

Is it possible to simply use the coordinates of landmarks $(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)$ as continuous phenotypes $x_1, y_1, \ldots, x_p, y_p$ using Brownian motion along a phylogeny?

Yes, but ...

First we must make sure that the forms (or if we are scaling to eliminate size differences, the shapes), are represented in a proper morphometric space.

Otherwise meaningless translations (shifts) or rotations of the specimens will affect the coordinates.

In effect we are superimposing the specimens properly, although a complete superposition isn't necessary.

Dealing with translation

The specimens can be reduced to differences among the `x` x coordinates of different points, and differences among the `y` coordinates too, thus losing the grand mean of each specimen.

This amounts to taking contrasts between the different points of one specimen *(a different matter from phylogenetic contrasts, which are for the same coordinate, but between different specimens)*.

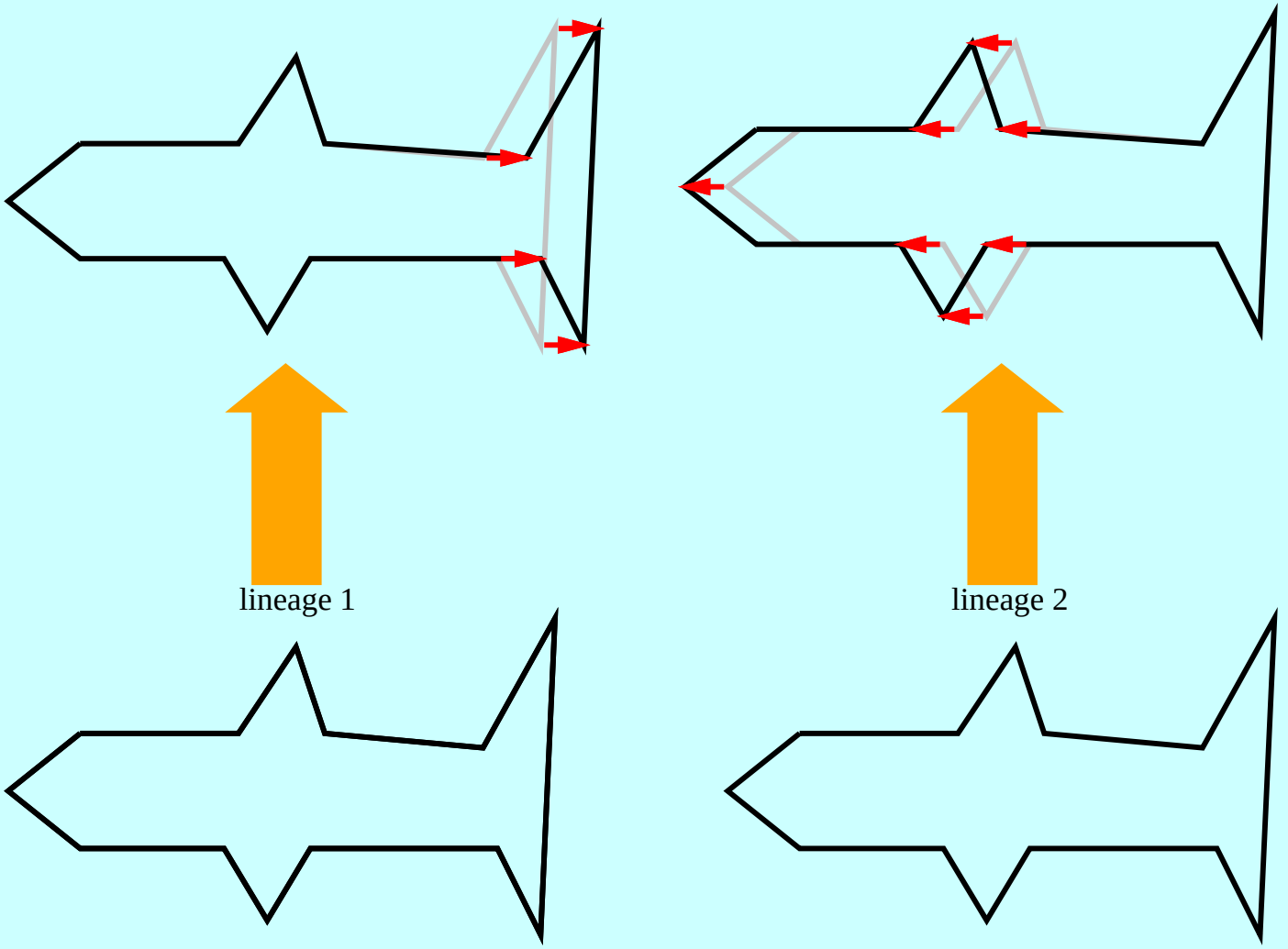
In effect one is centering each specimen so that the mean of its points is at $(0, 0)$. (The assumption is that the horizontal and vertical placement of the specimen on the digitizer is not useful information).

This has the effect of dropping two degrees of freedom so that each specimen now has $2p - 2$ coordinates. It now "lives" in a $(2p - 2)$-dimensional space.

For example, we could drop the last point (x_p, y_p) as it is then always predictable from the sum of the other points. Or we could replace the coordinates by any set of $2p - 2$ contrasts such as the Helmert transform.

Can we superpose specimens?

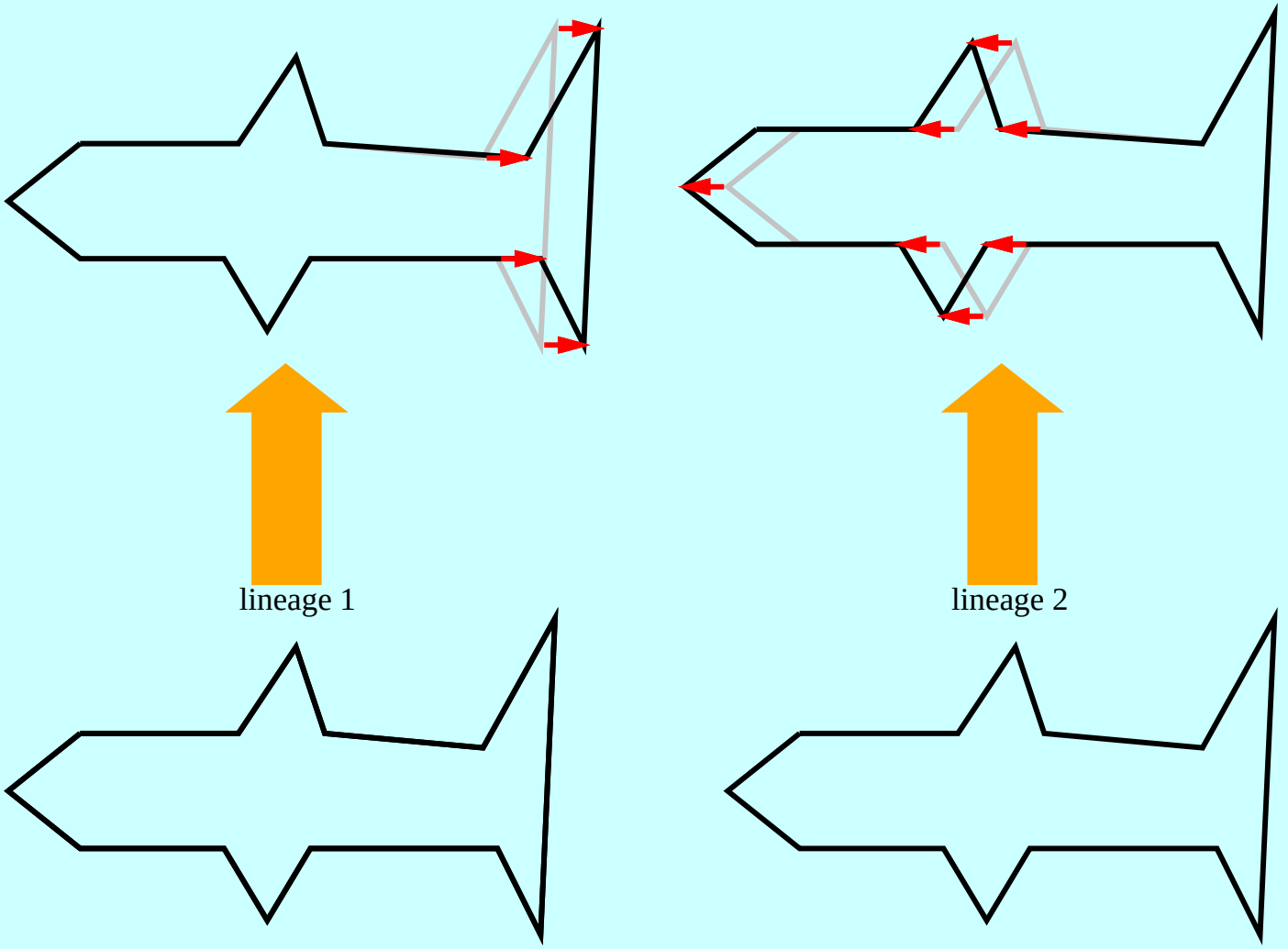
Consider two cases:



Are these different?

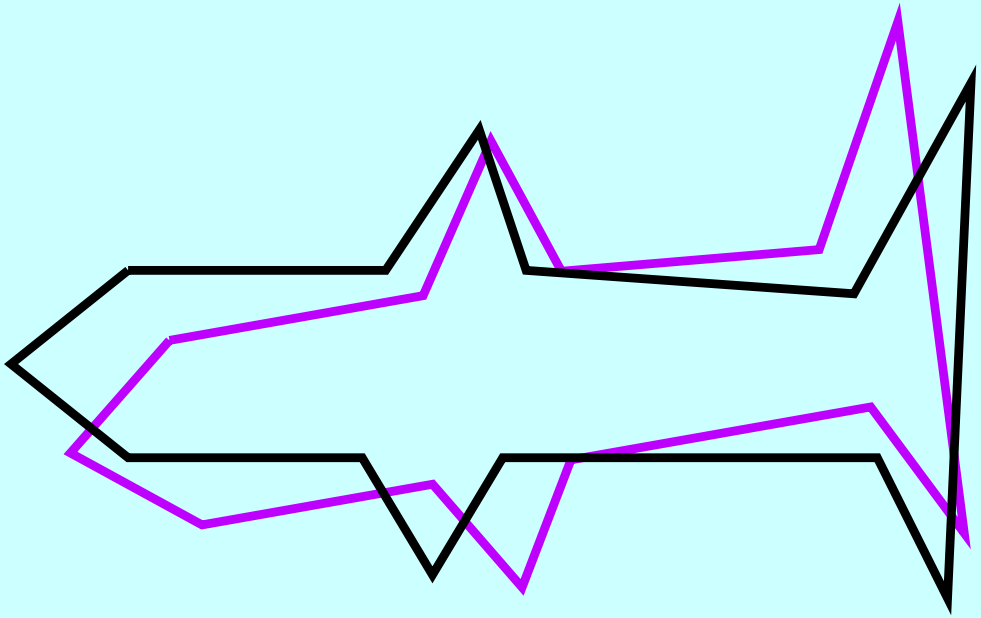
Why superposition is in principle impossible

Consider two cases:



Are these different? No!

The annoying issue of rotation



Sadly, there is no corresponding transform that tosses out rotation, as there is for translation.

Degrees of freedom and other transforms

There are other possible rotation transforms that are all approximately equivalent, including:

- Determining rotations by making a joint Procrustes (least squares) superposition.
- The “Bookstein transformation”, an approximate Procrustes.
- Choosing the angles of rotation of all but the first specimen ($\theta_2, \theta_3, \ldots, \theta_p$) to maximize the resulting likelihood.

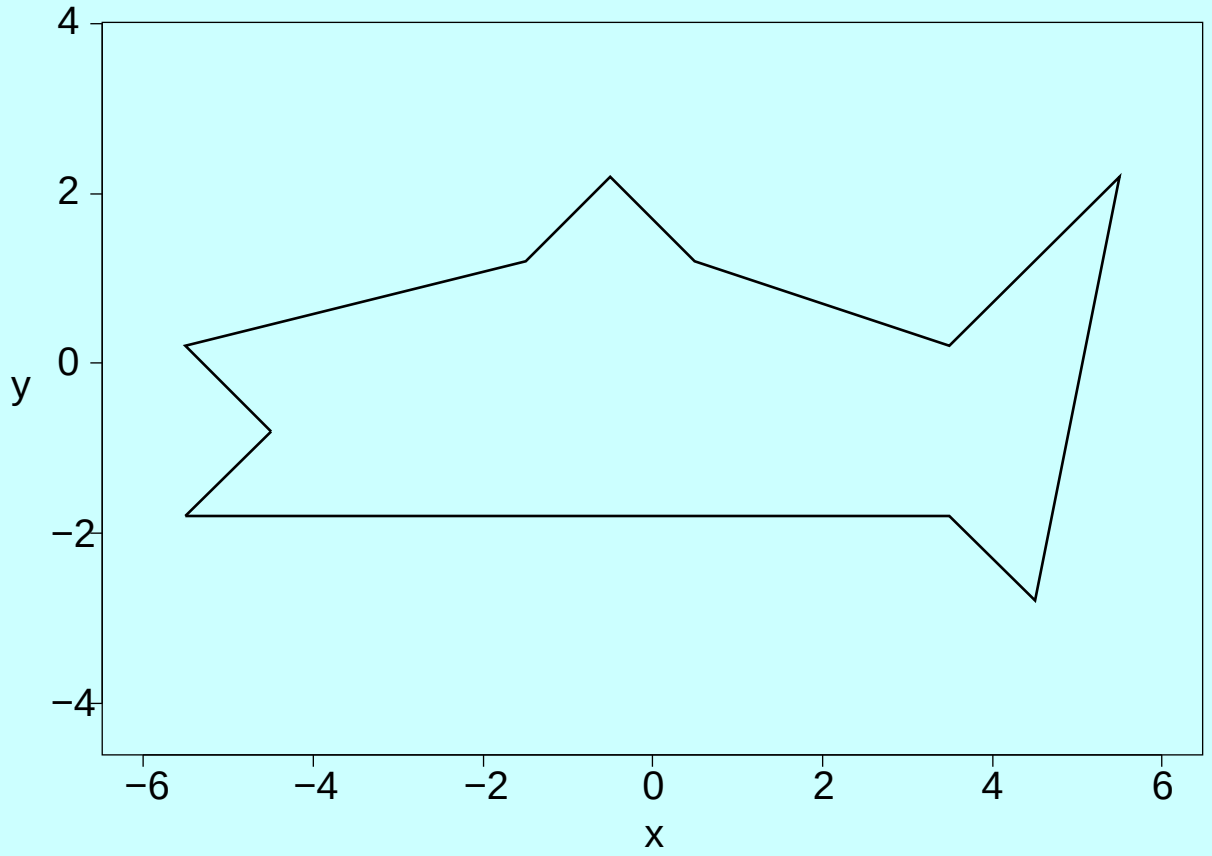
All of these reduce the degrees of freedom of each specimen by 3, to $2p - 3$.

But does this mean that the multivariate density function does not exist? No, it does exist, just in a $(2p - 3)$-dimensional subspace.

In that space, all the usual machinery of the phylogenetic comparative method is available: contrasts to evaluate covariation of characters, reconstruction of ancestors, etc.

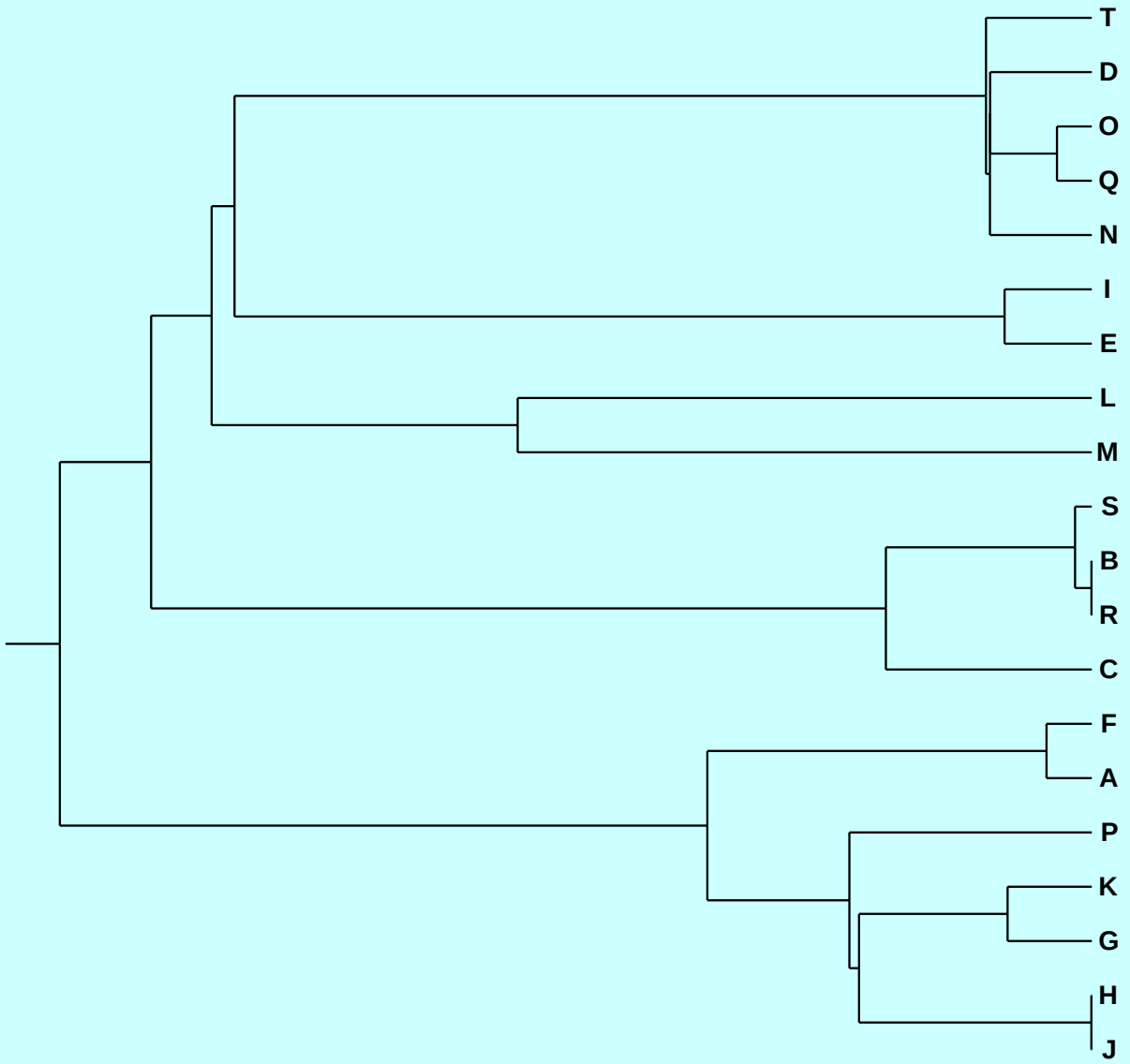
A simulation: the true ancestor

Here is a simulation showing the operation of the Bookstein Transformation in an imaginary species:



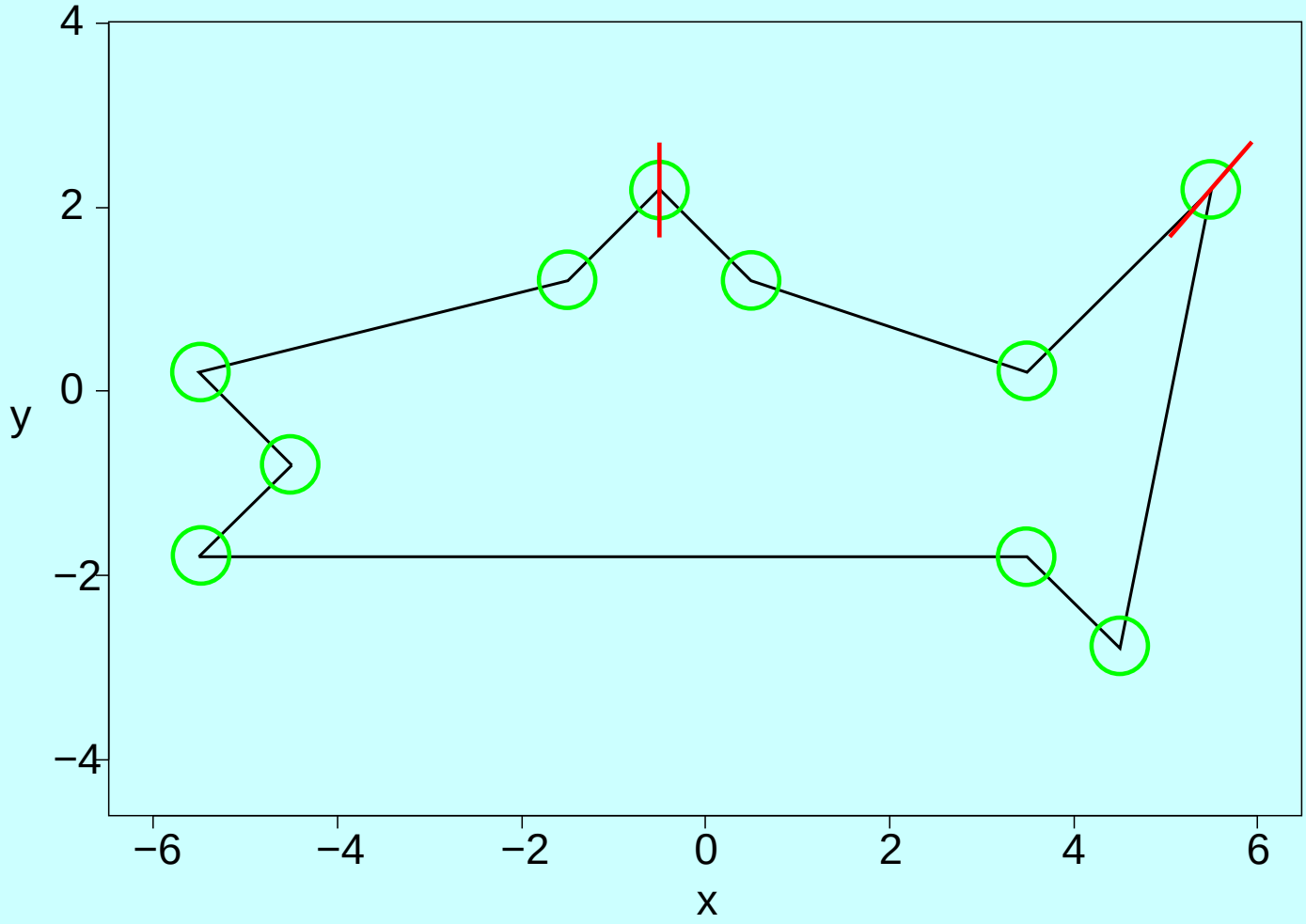
The dreaded Thresher Salmon-Shark (*Palionchorhynchus*)

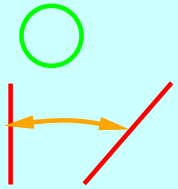
The true tree of 20 forms



(We used a 100-species tree similar to this).

The true directions of change by Brownian motion

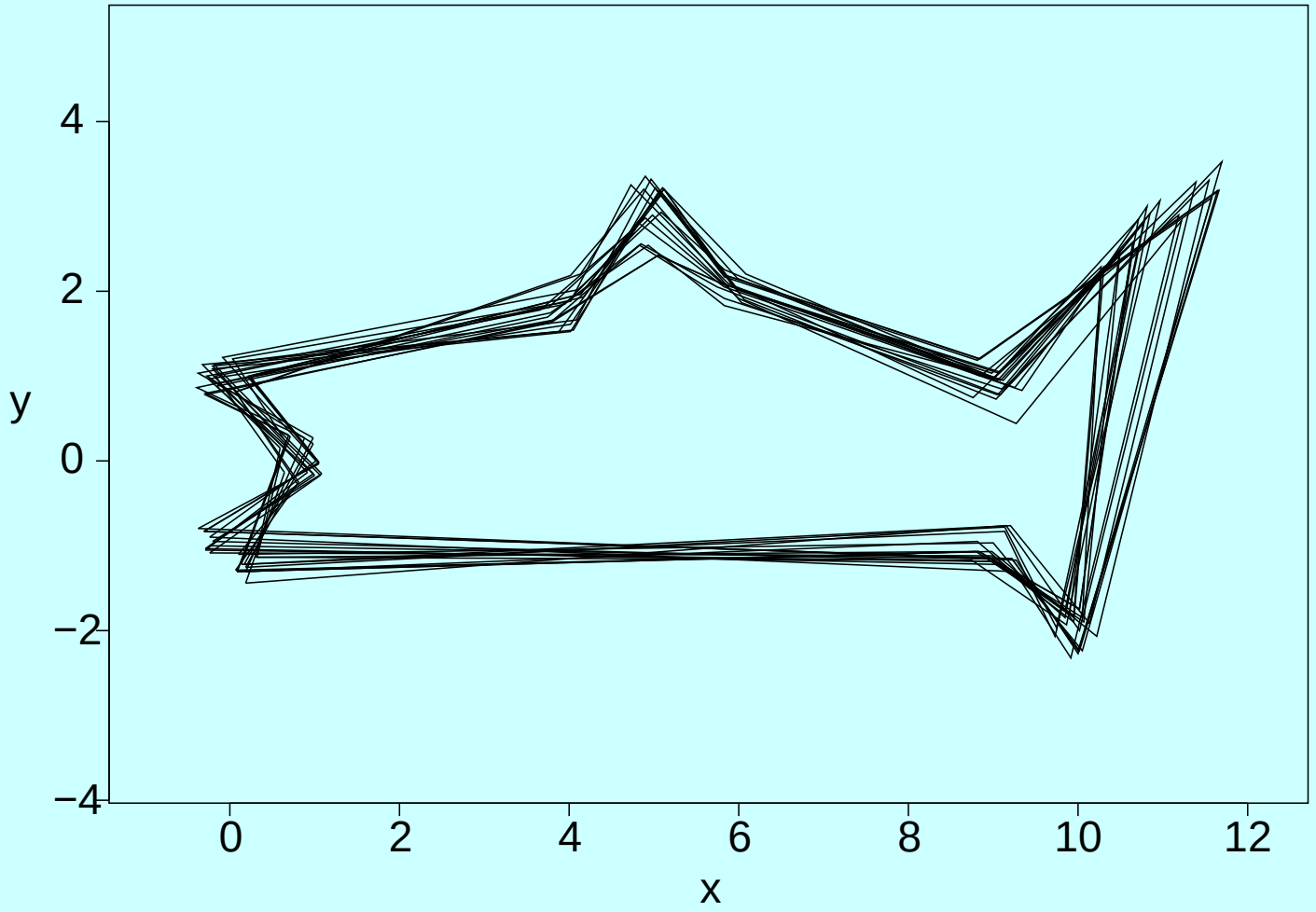




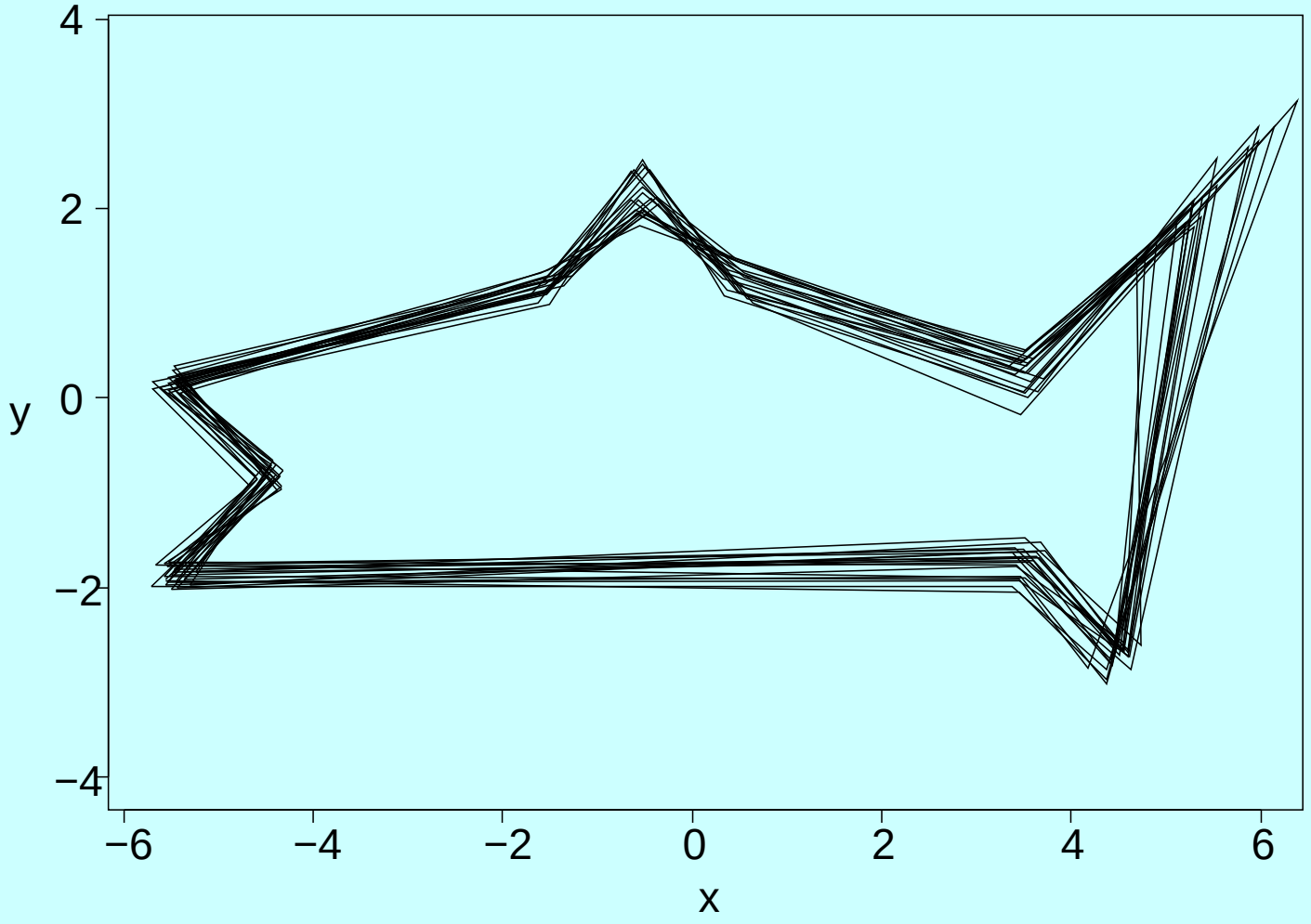
Green circles are independent circular normal change

Red lines show perfectly correlated normal change

The (unknown) true superposition



The Procrustes (least squares) superposition



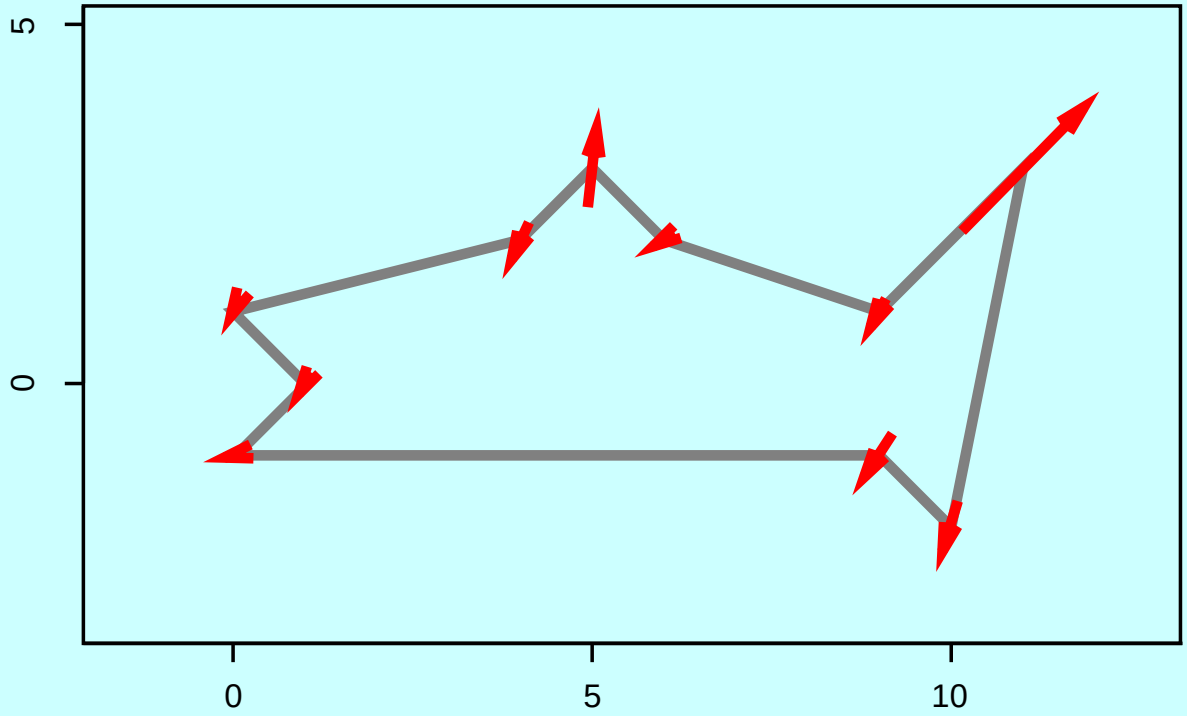
Inferring the covariances of change

Basically, you just take the standardized contrasts as independent samples, as usual. The variation will “live” in the 17-dimensional space but the best estimate of the covariance is still the empirical covariance.

We can then find eigenvalues and eigenvectors of the empirical covariances of contrasts, as usual.

The first principal component of variation

This was computed from phylogenetic contrasts on the true tree with the Procrustes superposition:

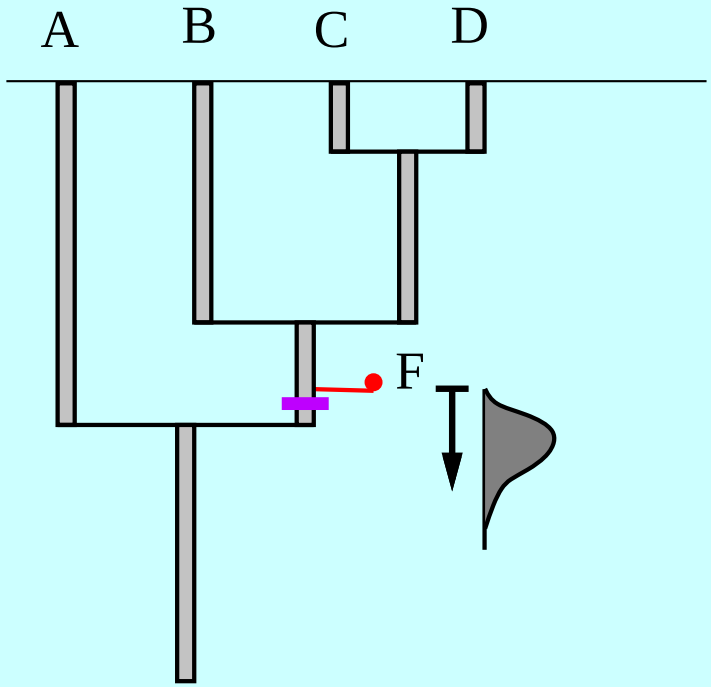


(Not too bad, though there are some signs that the program still has problems).

Part 2

Fossils and phylogenies

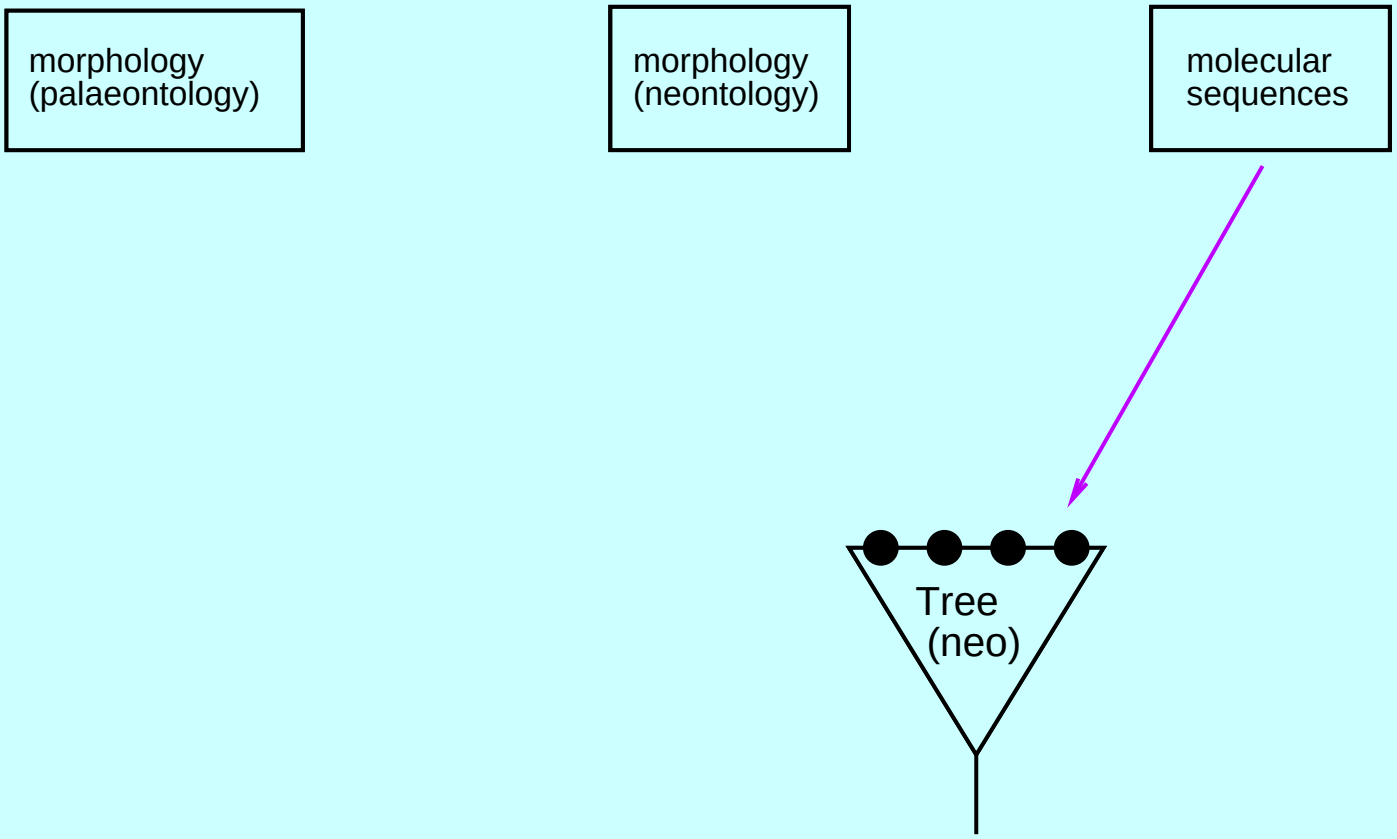
Present methods for calibration



Can take a fossil to indicate a bound on how recently a common ancestor was present. Use various priors on how much earlier or how much more recently.

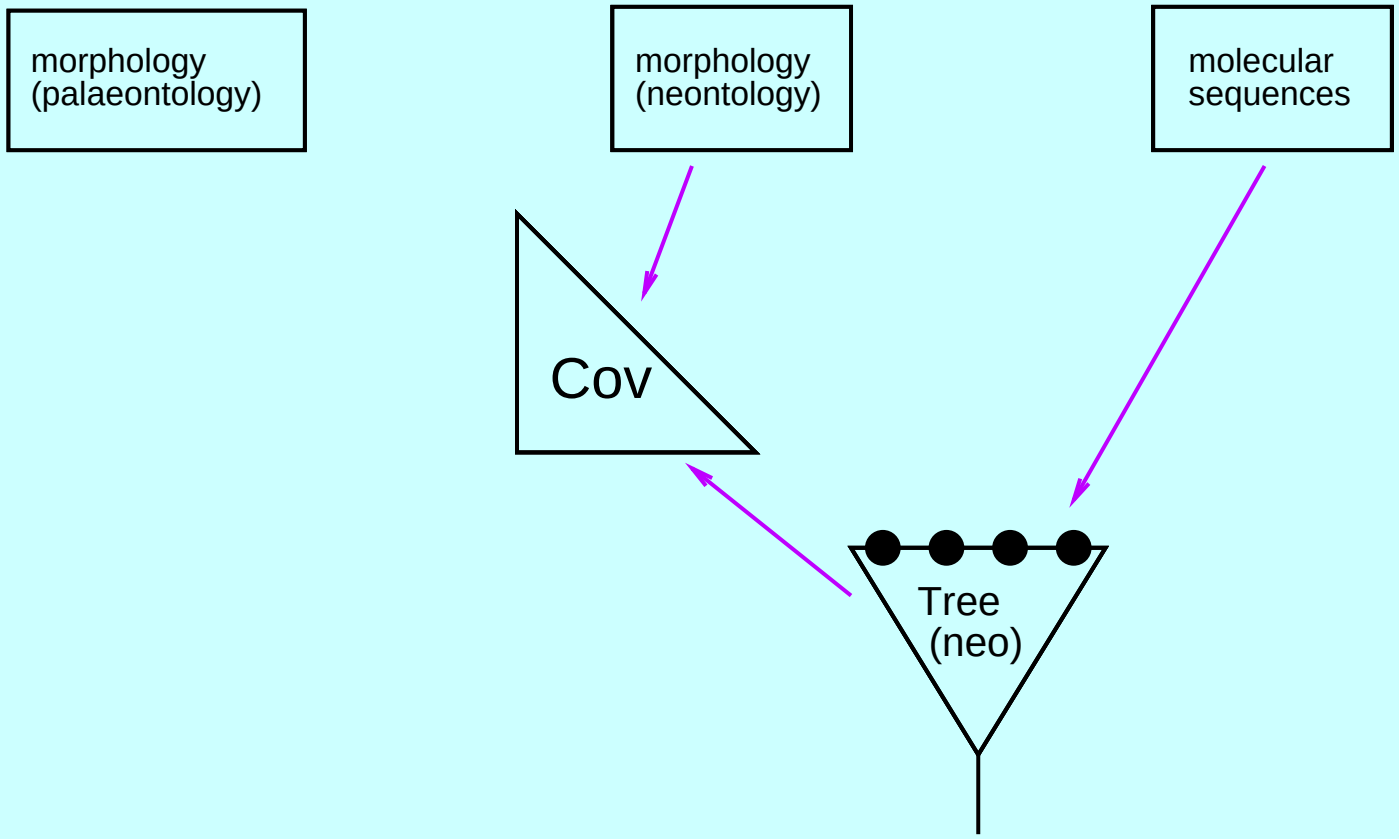
But there is another way, which is being explored by me and (independently) by Alexander Pyron (2011) and by Fredrik Ronquist et al. (2012)

Another way of using fossils



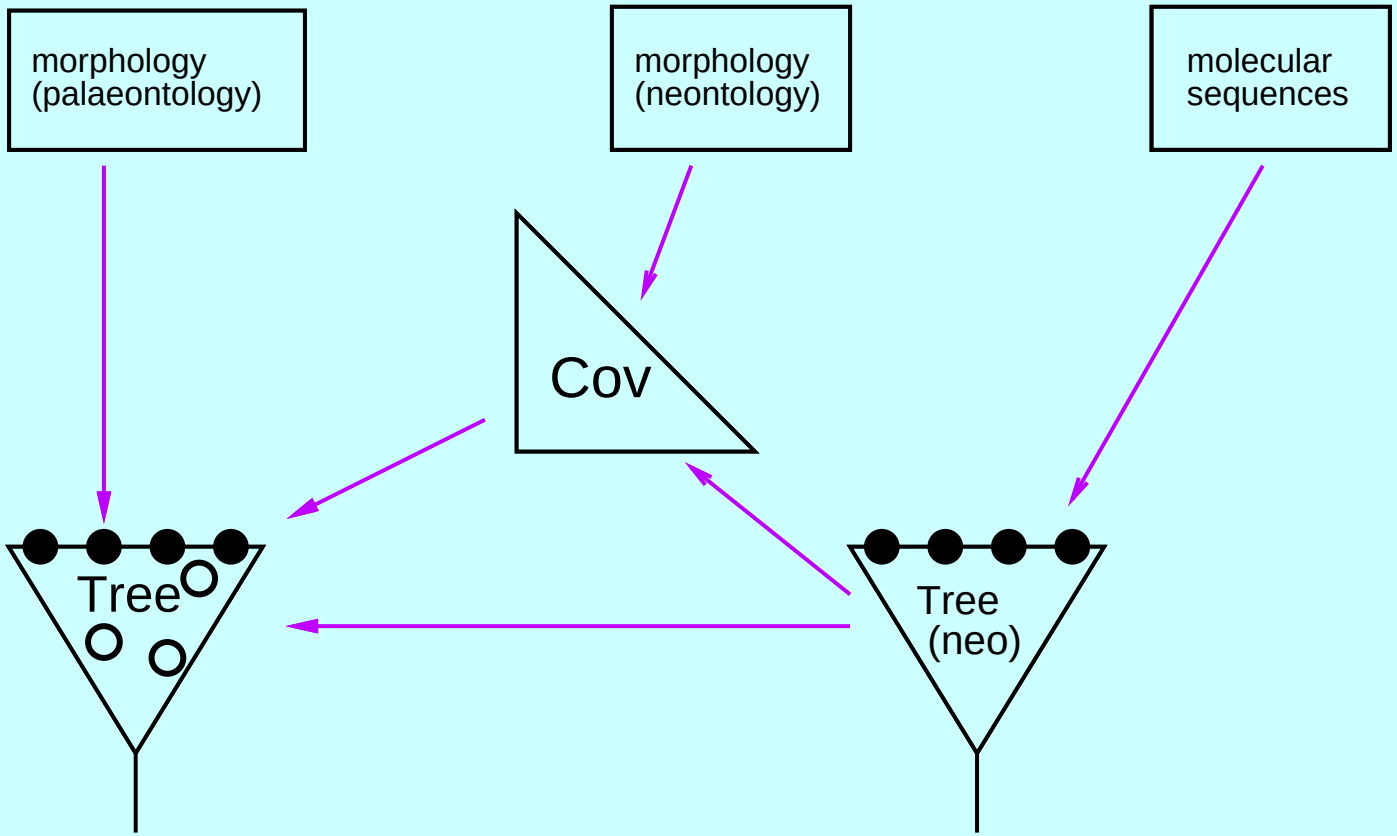
Infer tree of present-day species from molecular sequences

Using fossils



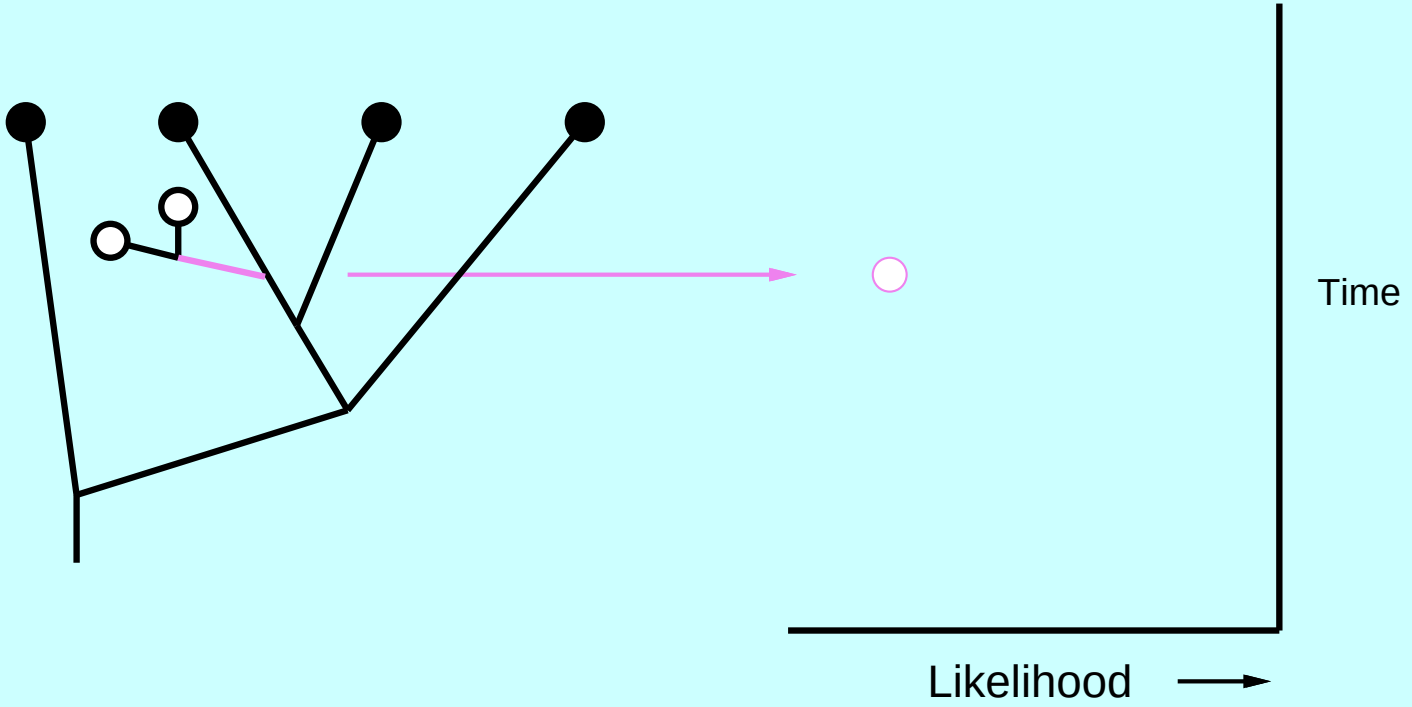
Infer covariances of morphology using it, present-day species

Using fossils



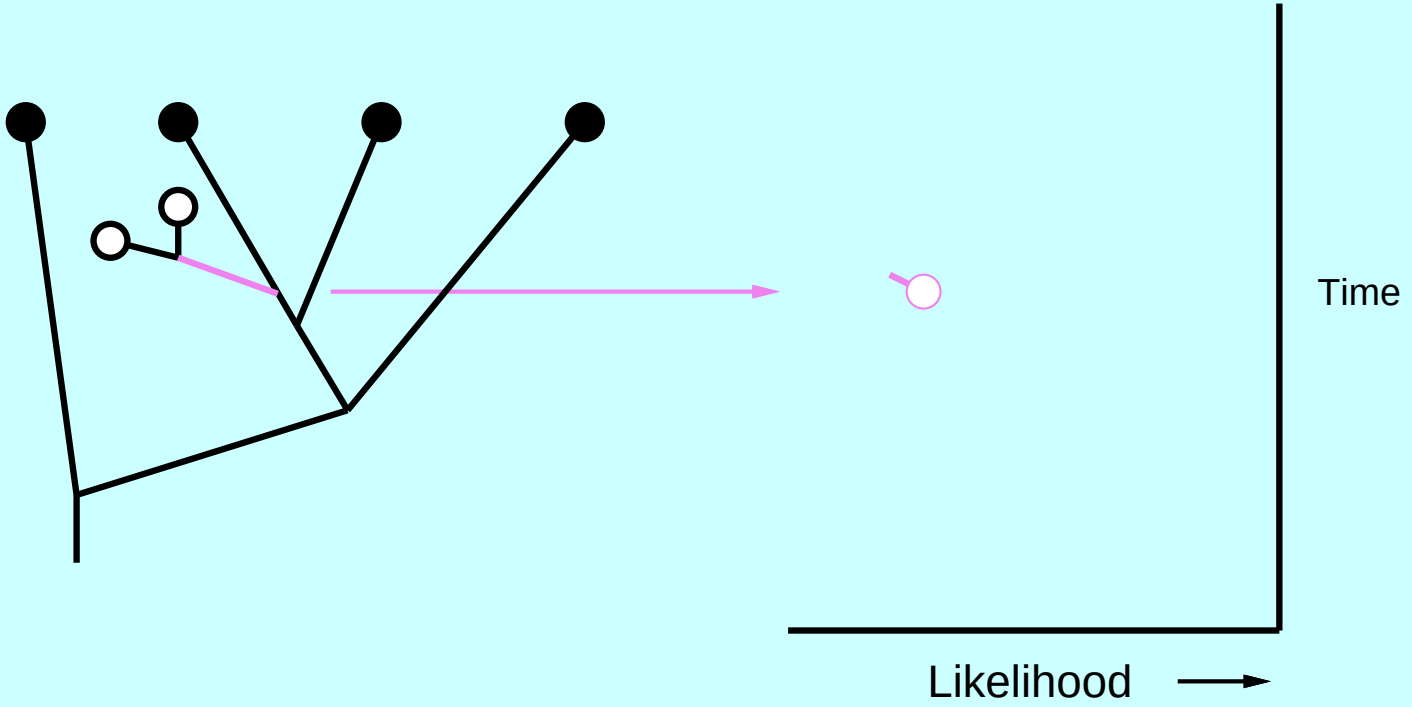
Infer placement of fossil species using their data

Using fossils



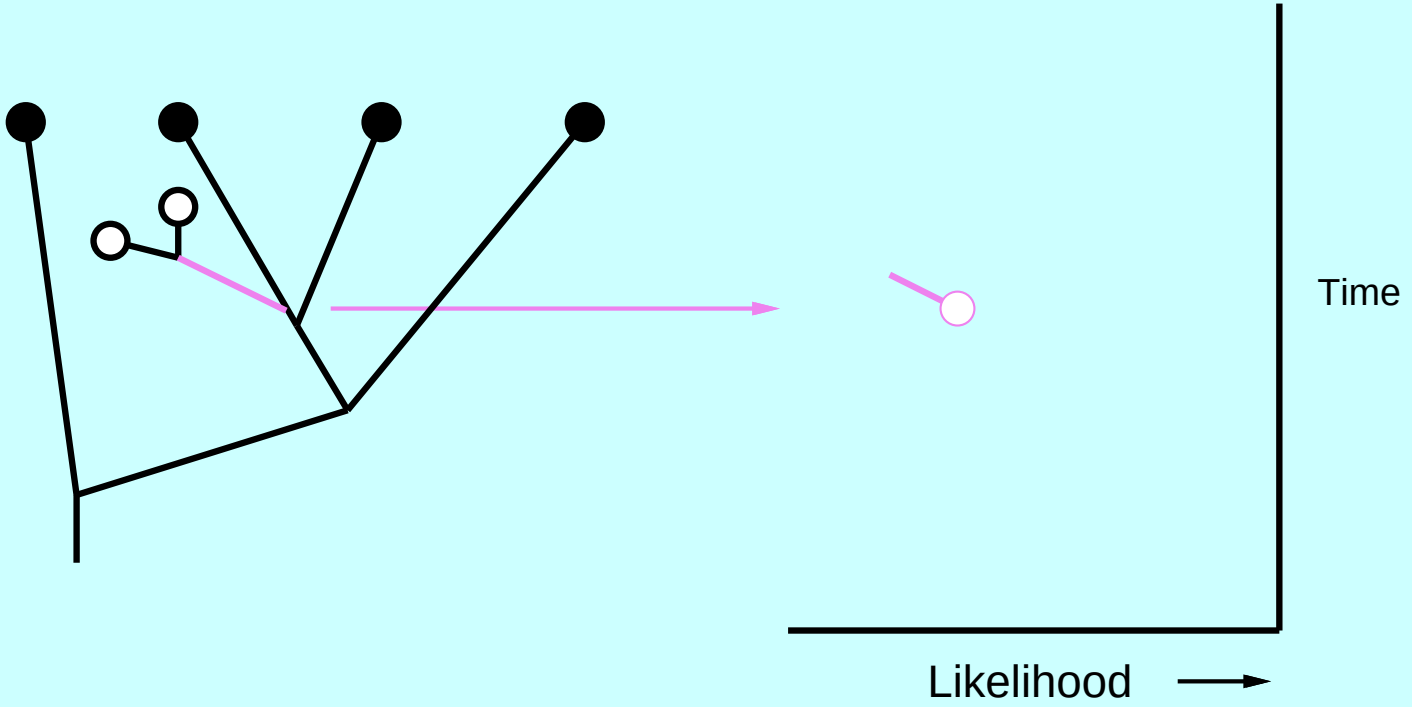
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



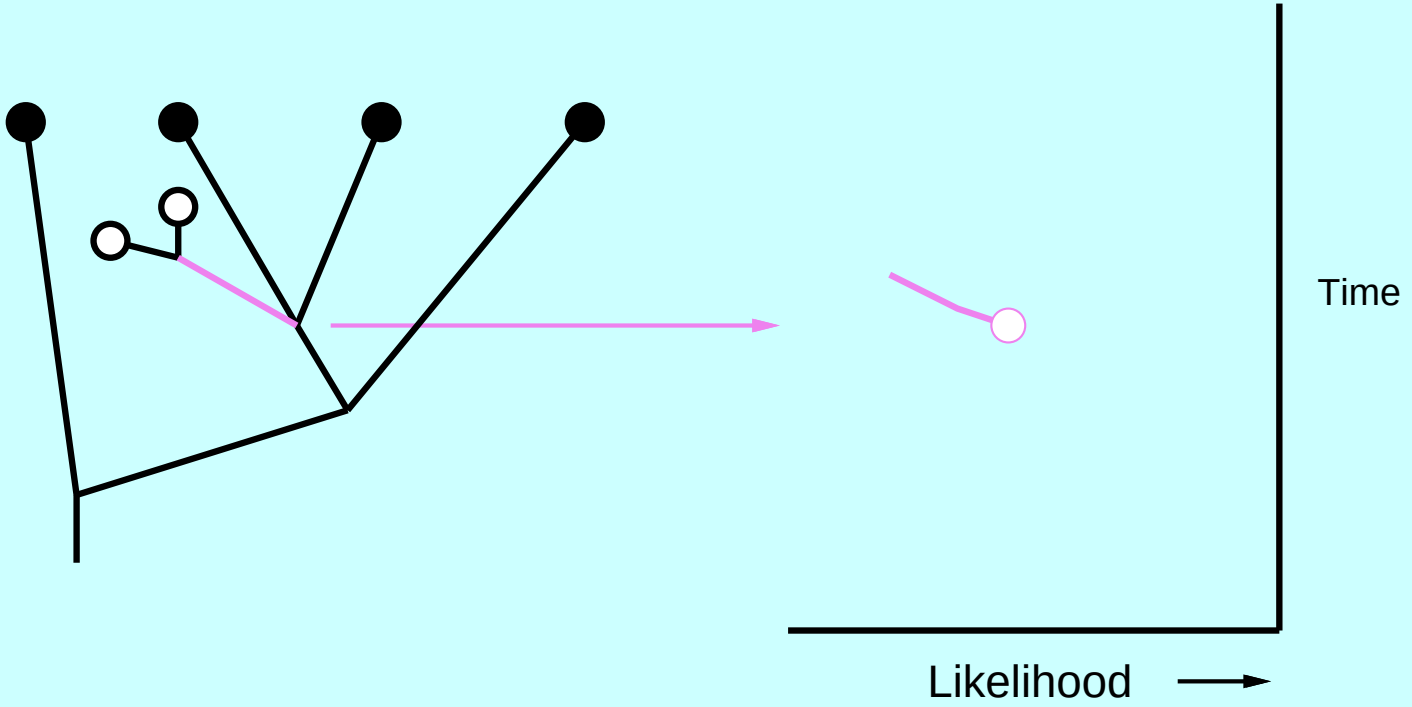
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



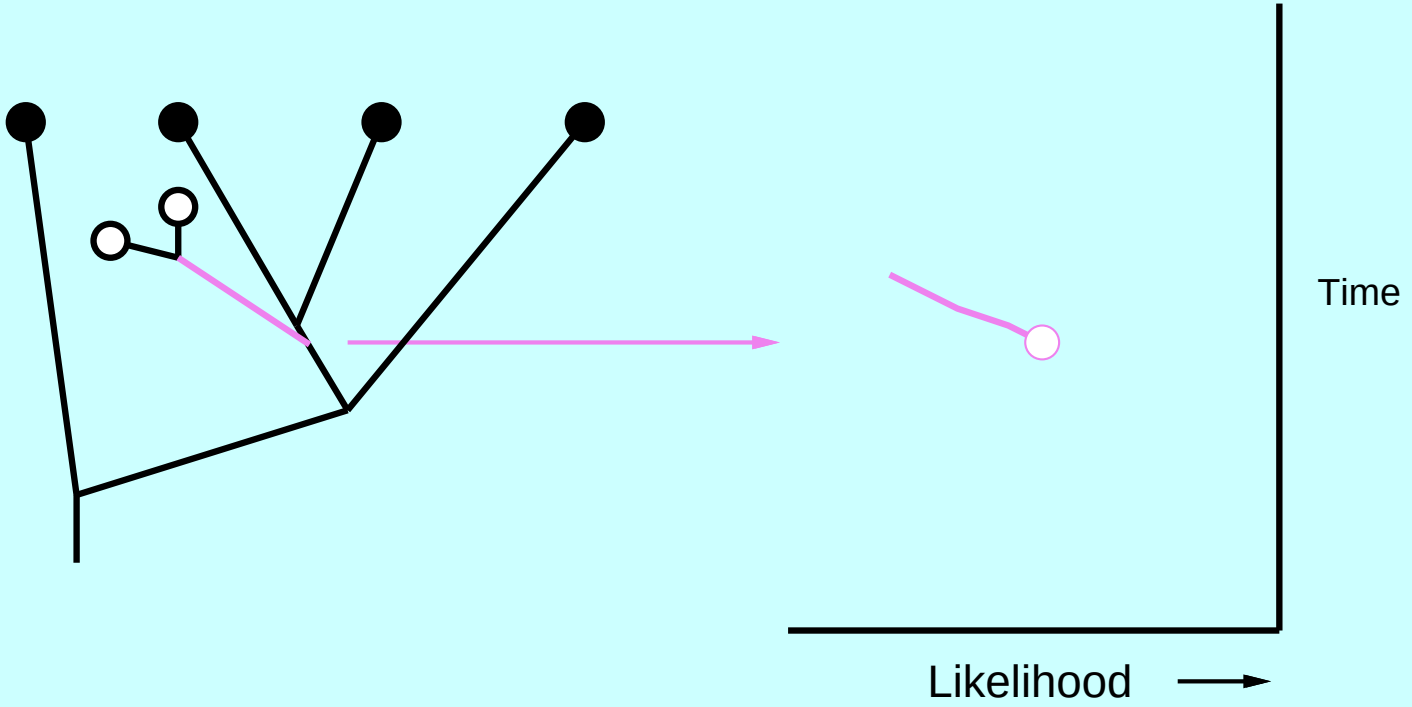
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



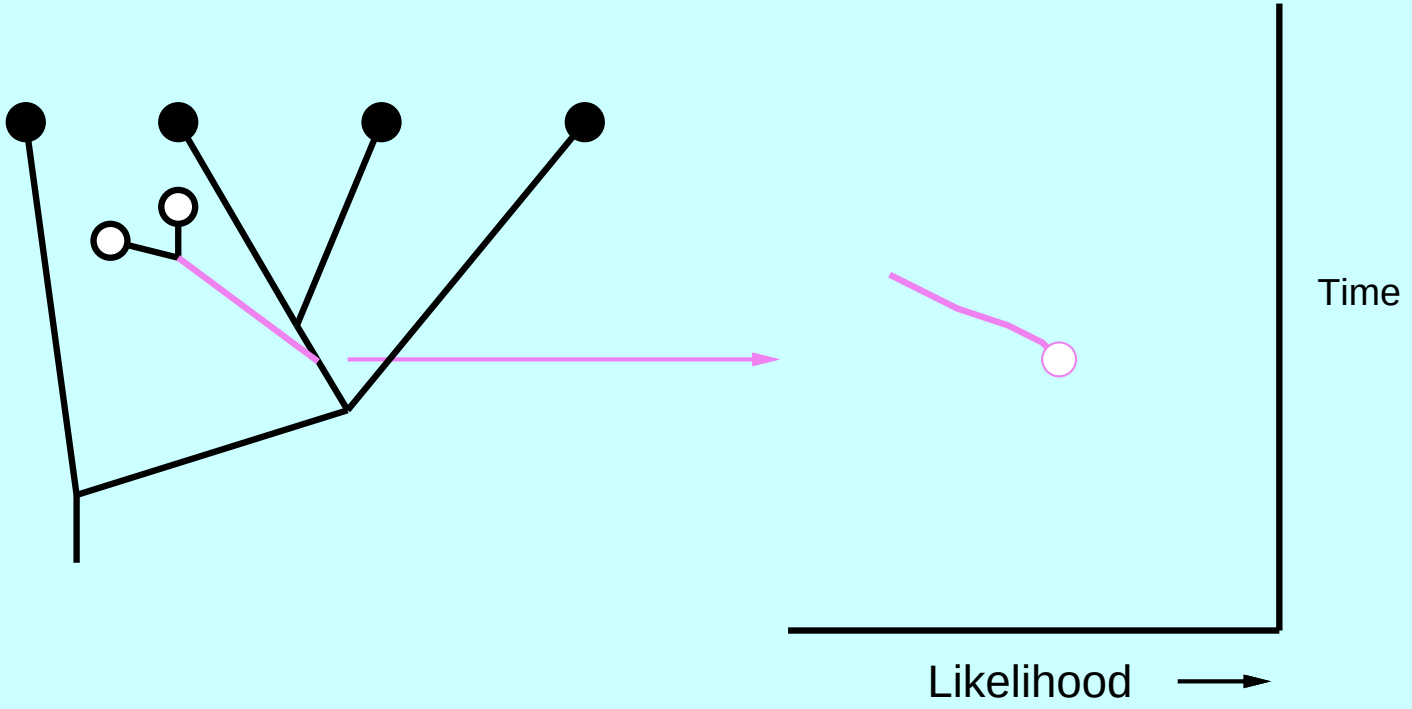
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



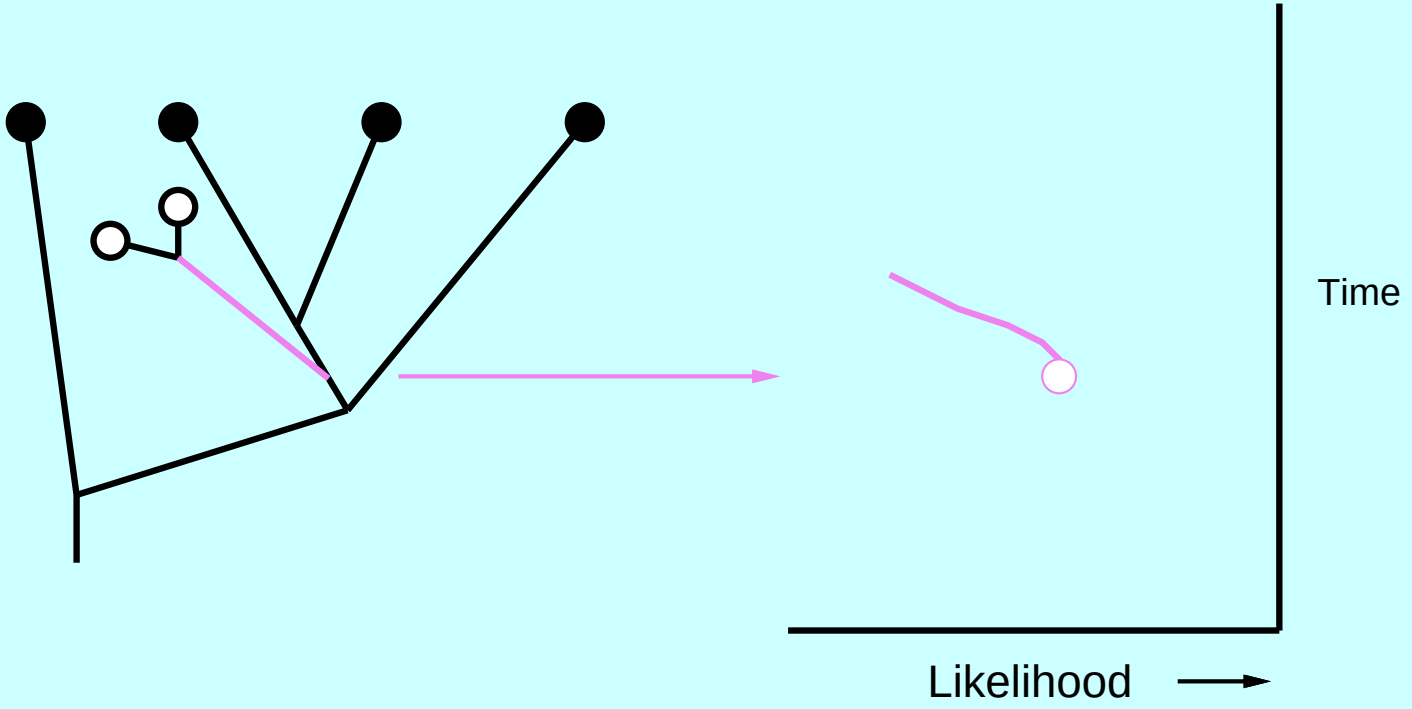
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



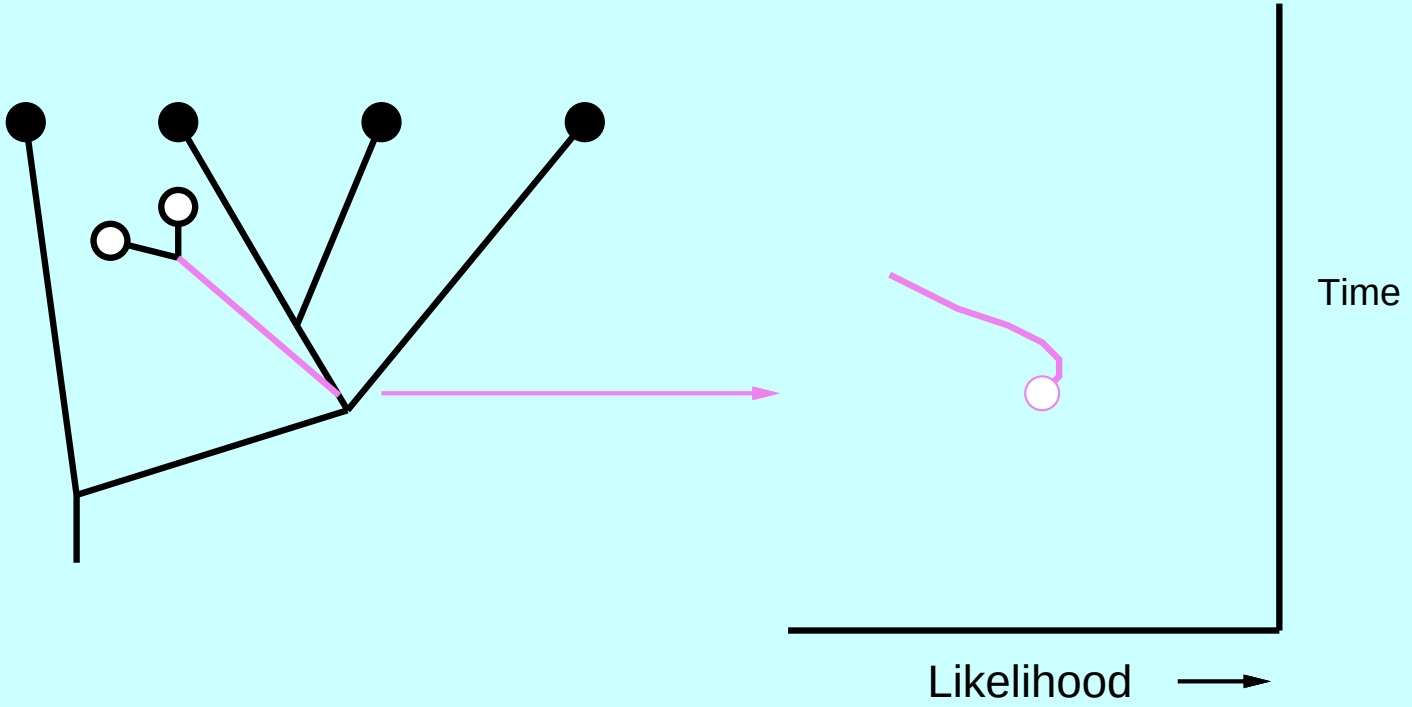
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



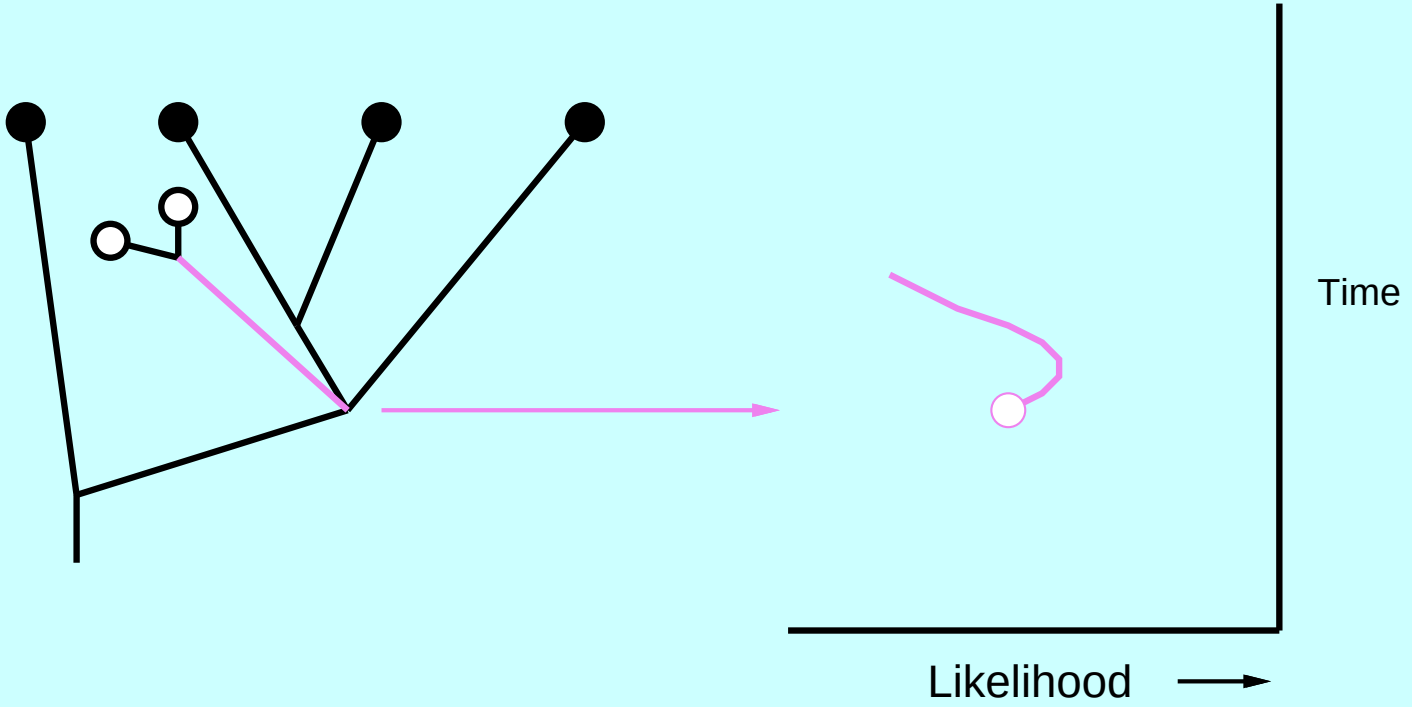
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



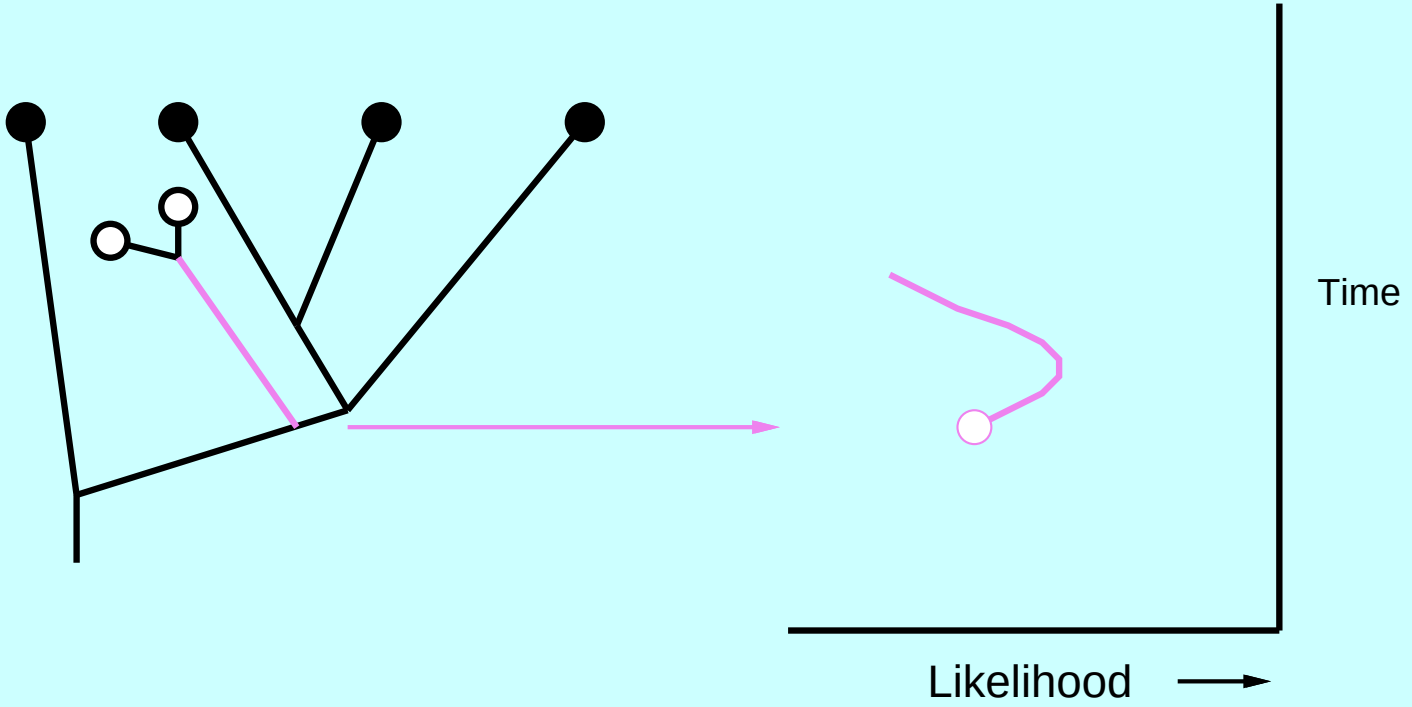
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



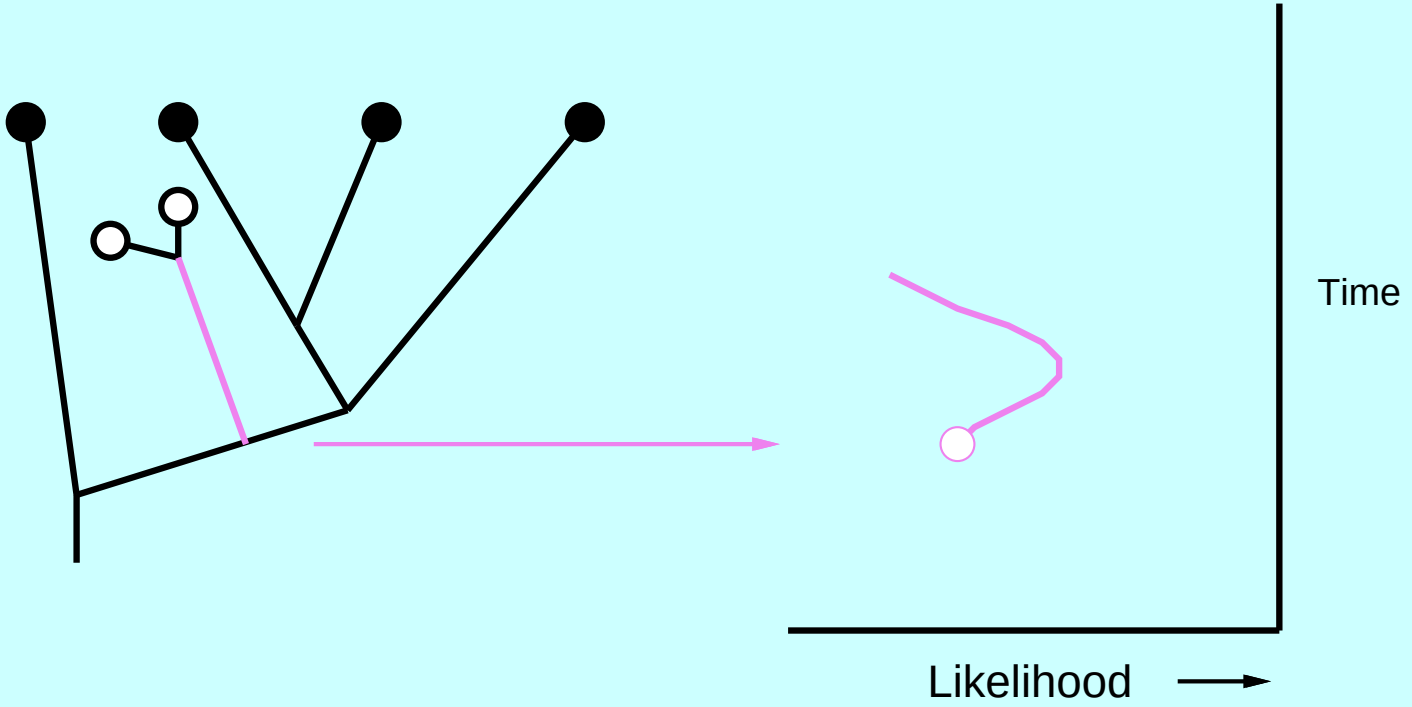
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



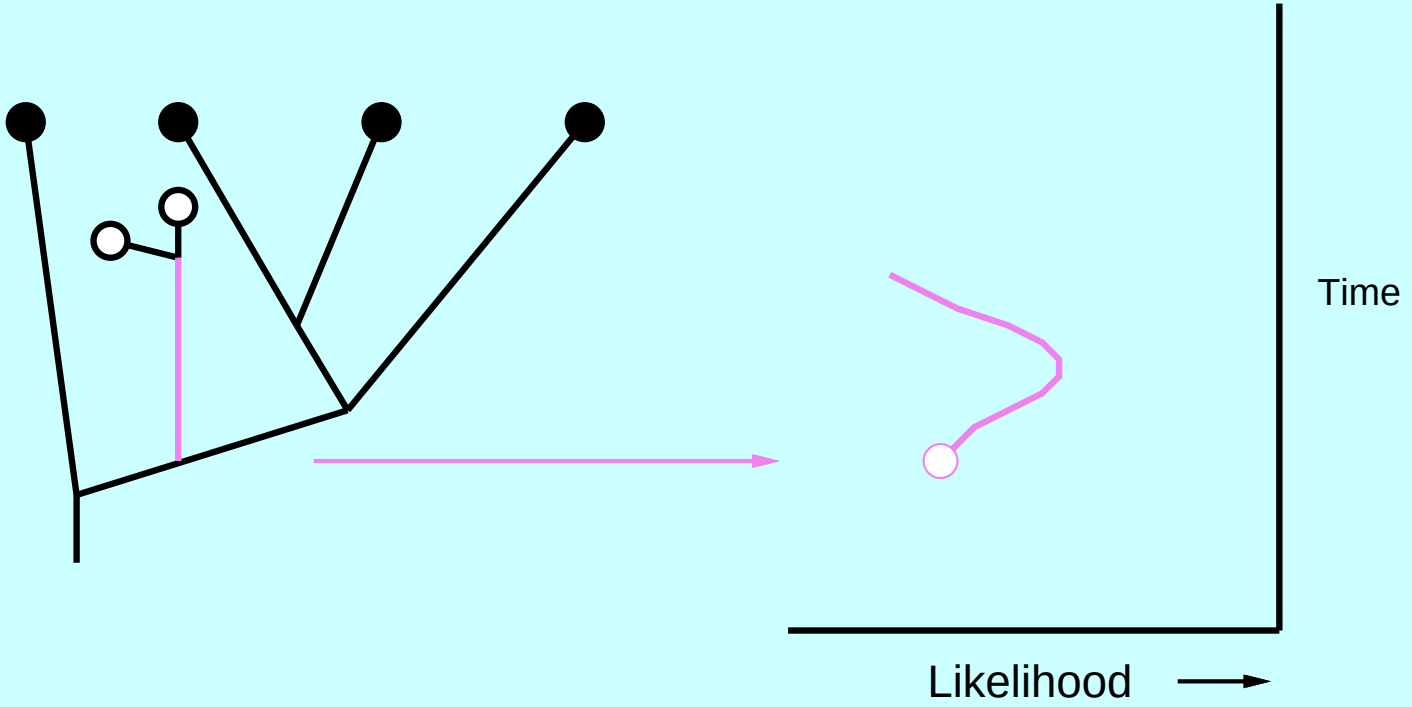
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



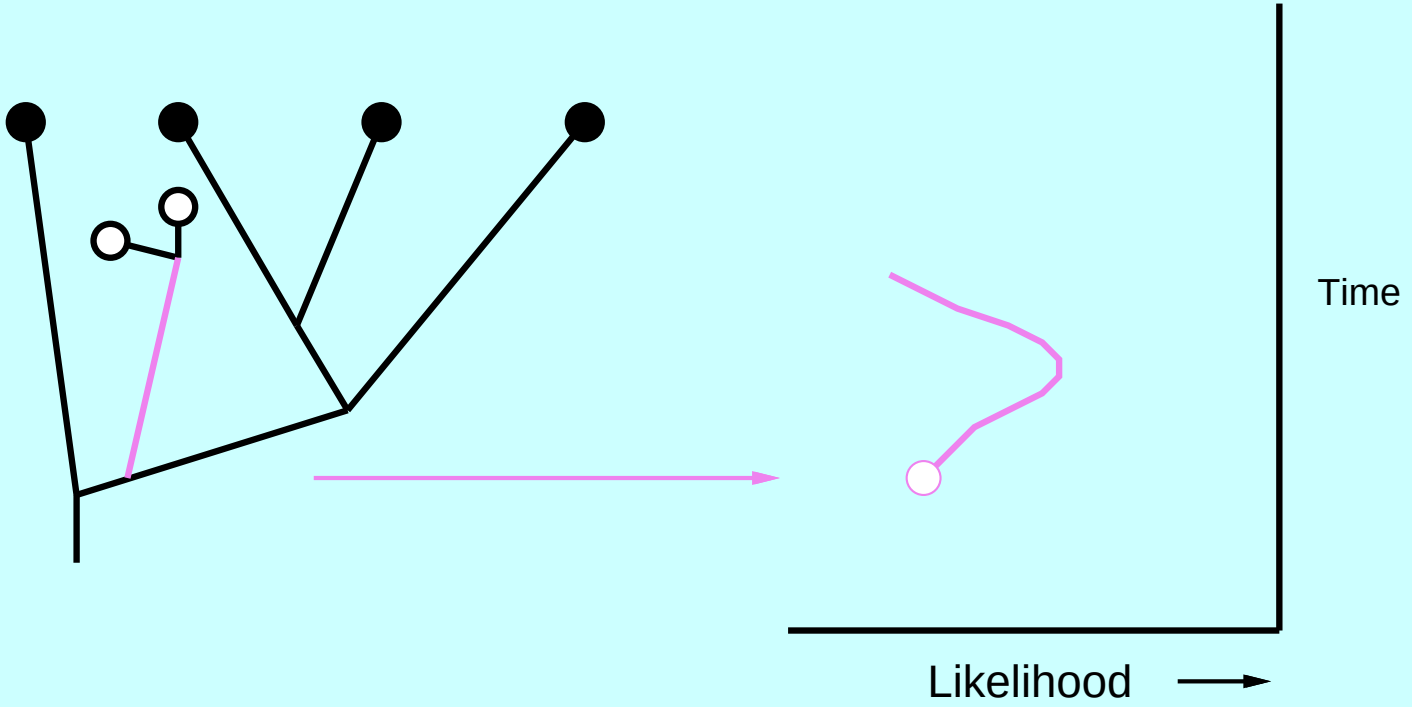
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



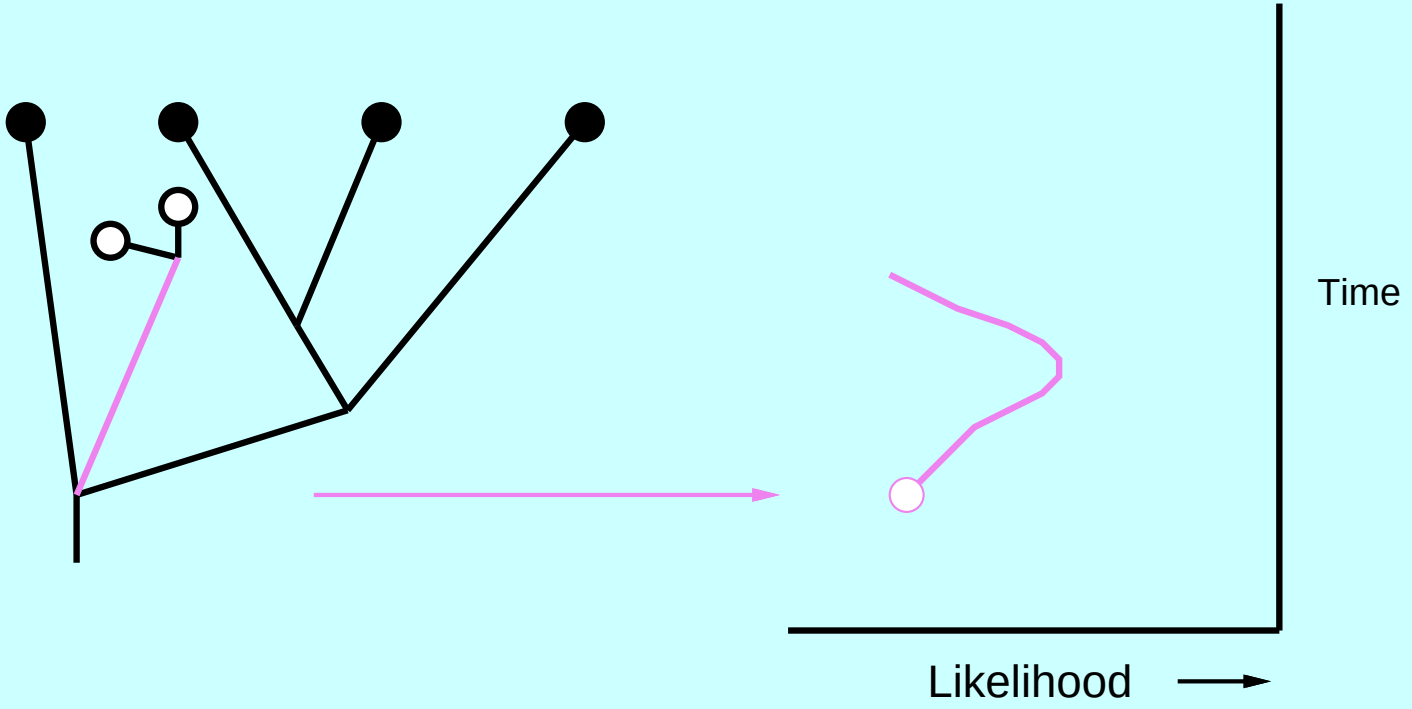
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



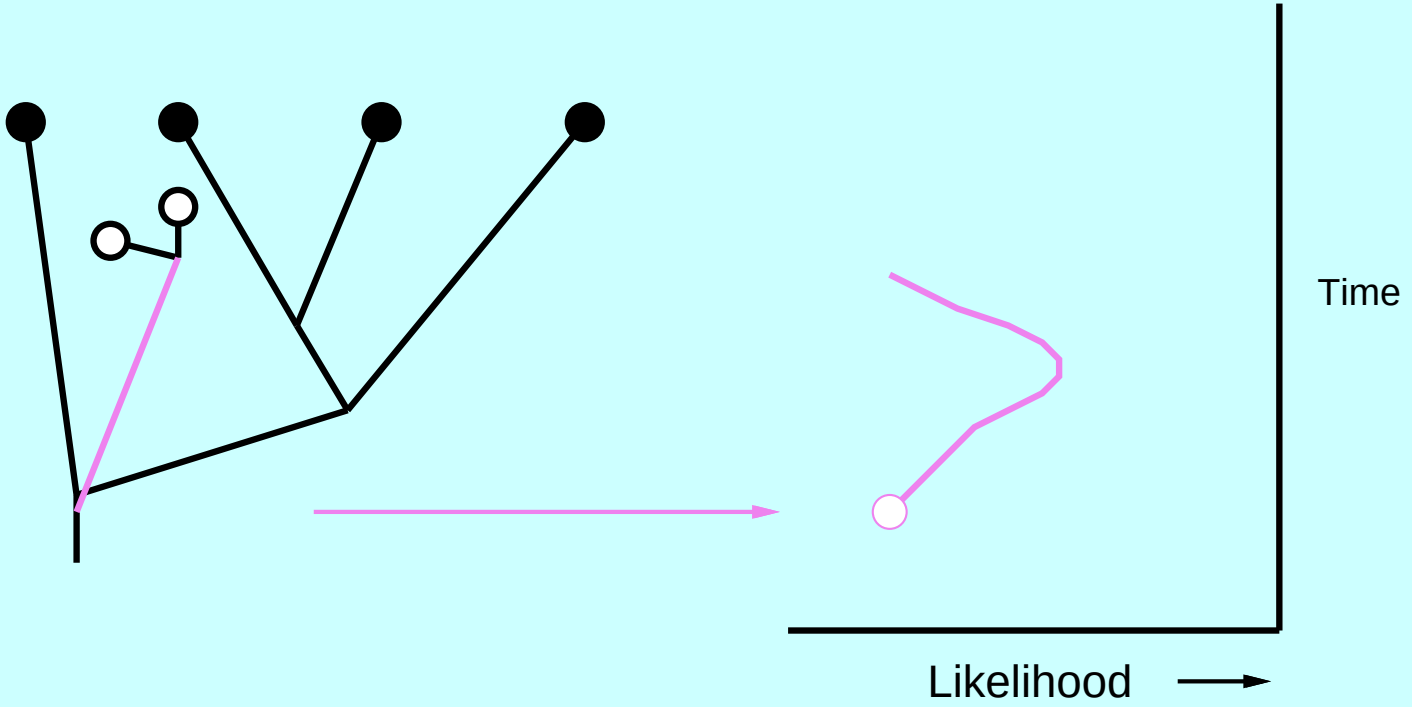
Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



Use fossil and present-day morphology, covariances, tree, also stratigraphic models

Using fossils



Use fossil and present-day morphology, covariances, tree, also stratigraphic models

A qualification

- The present method takes the molecular tree as known.

A qualification

- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.

A qualification

- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.
- Pyron and Ronquist both use a more comprehensive “total evidence” approach of allowing the morphological data to influence Bayesian inference of the tree.

A qualification

- The present method takes the molecular tree as known.
- Uncertainty in it could be modelled by doing the analysis multiple times on bootstrap samples (or Bayesian posterior samples) of the tree estimates.
- Pyron and Ronquist both use a more comprehensive “total evidence” approach of allowing the morphological data to influence Bayesian inference of the tree.
- I suspect this will have little effect if there is a lot of molecular data, so I am sticking with this approach.

The algebra

If $\mathbf{T}$ is the covariances of $\mathtt{n}$ tips on the tree, and $\mathbf{V}$ is the (unknown) covariances of the Brownian motion of the $\mathtt{p}$ characters, the log-likelihood of a set of characters (stacked as a vector) $\mathbf{x}$ is

$$\ln L \;=\; -(np/2)\ln(2\pi) - (1/2)\ln|\mathbf{T}\otimes\mathbf{V}| - (1/2)(\mathbf{x}-\mu)^t(\mathbf{T}\otimes\mathbf{V})^{-1}(\mathbf{x}-\mu)$$

If $\mathbf{C}$ is an $(\mathtt{n}-1)\times\mathtt{n}$ set of contrasts, each orthogonal to the grand mean, such that $\mathbf{CTC}^t$ is an $\mathtt{n}-1$-dimensional identity matrix, then taking the density of the transformed data $\mathbf{y} = \mathbf{C}\,\mathbf{x}$, this has expectation vector $\mathbf{0}$:

$$\ln L \;=\; K - (1/2)\ln|\mathbf{I}_{n-1}\otimes\mathbf{V}| - (1/2)\mathbf{y}^t(\mathbf{I}_{(n-1)}\otimes\mathbf{V})^{-1}\mathbf{y}$$

(where K collects the constant stuff).

... simplifying ...

This can also be expressed as

$$\ln L \;=\; K - ((n-1)/2)\ln|\mathbf{V}| - (1/2)\mathrm{tr}\left(\mathbf{SV}\right)^{-1})$$

where

$$\mathbf{S} \;=\; \sum_i \mathbf{y}^{(i)}\left(\mathbf{y}^{(i)}\right)^t$$

is the $\mathsf{p}\times\mathsf{p}$ sum of squares matrix of characters across contrasts. Inferring the Brownian motion phylogenetic covariances by maximum likelihood we find that

$$\widehat{\mathbf{V}} \;=\; \mathbf{S}/(\mathsf{n}-1)$$

which leads to

$$\ln \mathsf{L} \;=\; \mathsf{K}' \;-\; ((\mathsf{n}-1)/2)\ln|\widehat{\mathbf{V}}| \;-\; \frac{\mathsf{p}}{2}\prod_{\mathsf{i}=1}^{\mathsf{n}-1}\ln\left(\mathsf{v}_1^{(\mathsf{i})}+\mathsf{v}_2^{(\mathsf{i})}\right)$$

where the $\mathsf{v}_1^{(\mathsf{i})}$ and $\mathsf{v}_2^{(\mathsf{i})}$ are the (augmented) branch lengths for the ith contrast.

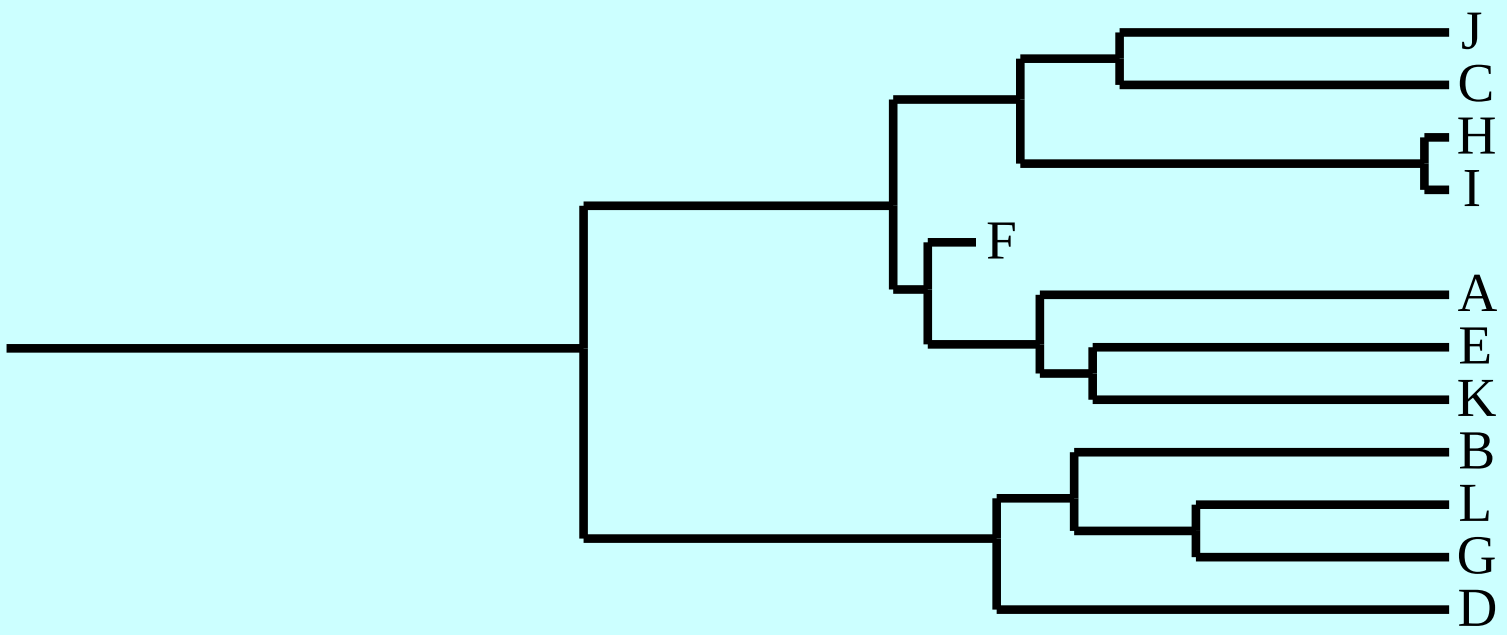
A simple result

The upshot is that to find the maximum likelihood placement of a fossil lineage, we

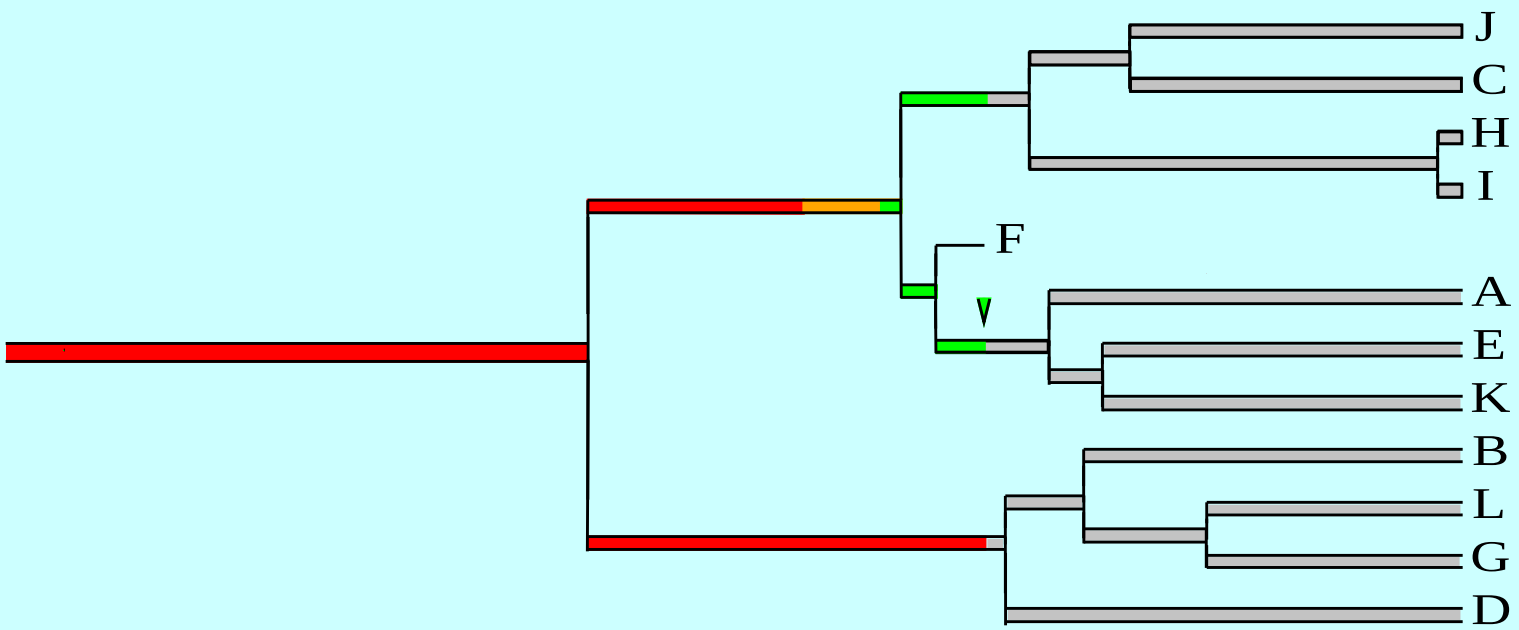
- Hook it up somewhere
- Obtain the contrasts for that tree
- Infer the phylogenetic covariances of the characters from the contrasts
- The log-likelihood for this placement is (a constant plus) $-(n-1)/2$ times the log of the determinant of the covariance matrix, minus a penalty which depends on the sum of the logs of the standard deviations of the contrasts.

So we minimize the determinant to find the best placement. We can consider whether we can do likelihood ratio tests, too, at least for placement within a single branch.

An example: the true tree with F a fossil species

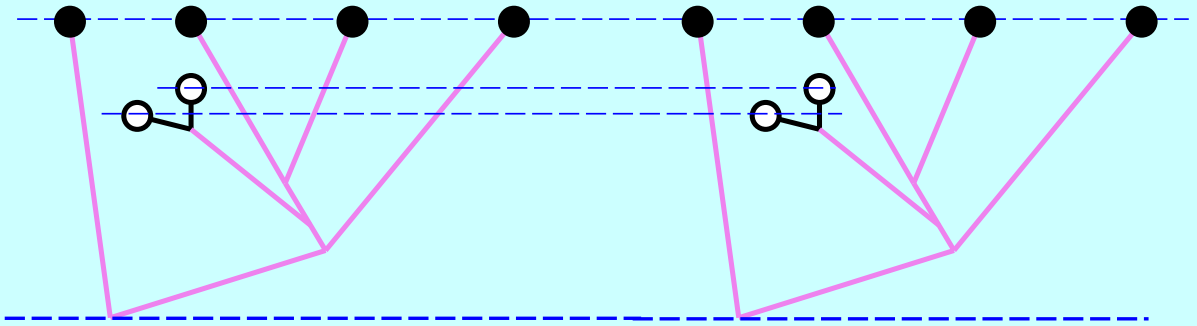


Traffic-light colors shows where fossil can be placed



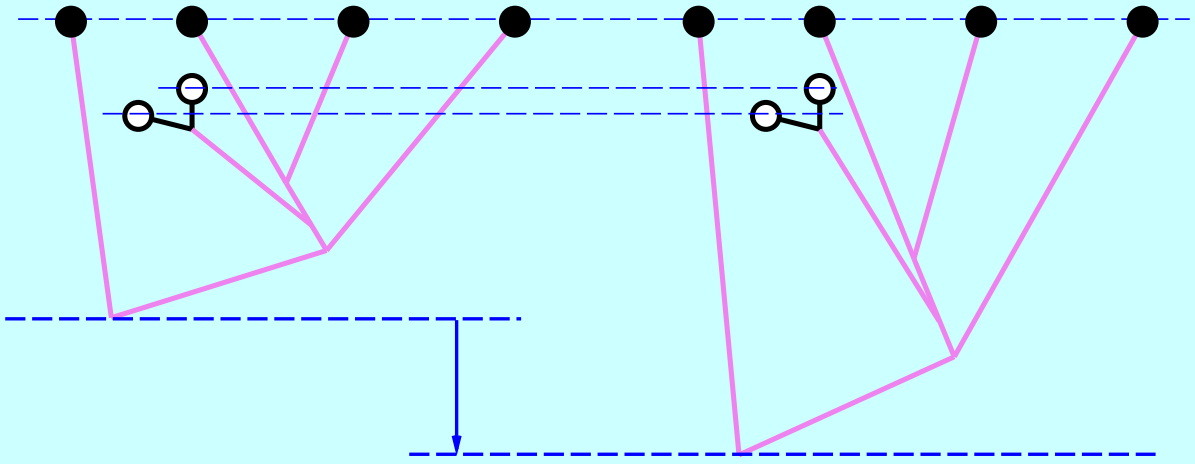
Green = within 1 log-likelihood unit, Orange = within 2 units, Red = lower than that. Green arrow is the ML placement. Gray placements are ruled out by date of the fossil.

Calibrating the molecular clock



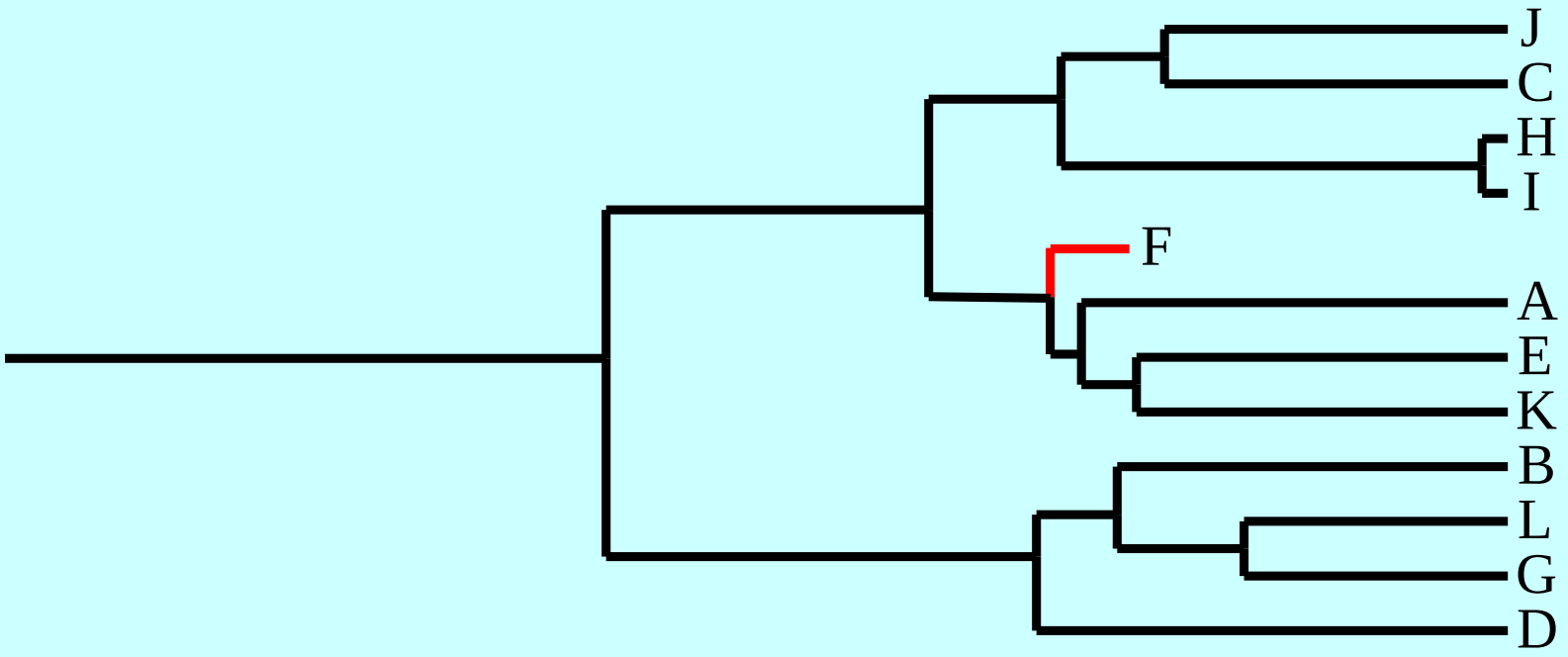
Molecular trees don’t usually have branch lengths on a time scale, and we need that. How to infer the calibration of the clock?

Calibrating the molecular clock



There will be two quantities to infer, the scaling of the molecular tree on the time scale, and the placement of the connection to the fossil. We make an ML estimate and accept other values that are not rejected by a Likelihood Ratio Test with 2 degrees of freedom.

Calibrating the molecular clock



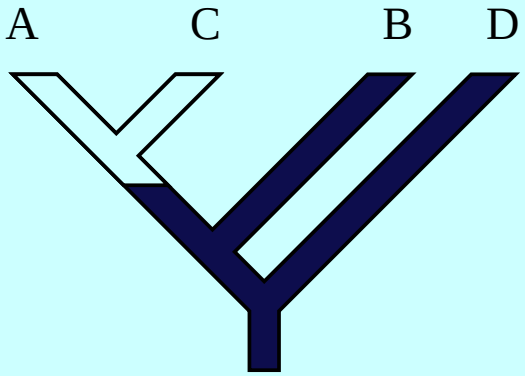
For example if (not a real example) the placement of F turned out to be as shown, with the branch length shown in red, that in turn scales the whole molecular tree, as we know the time of F.

Part 3

A threshold model for 0/1 characters

Current methods for statistical treatment of 0/1 characters

Pagel (1994) and Lewis (2001) treat such data with

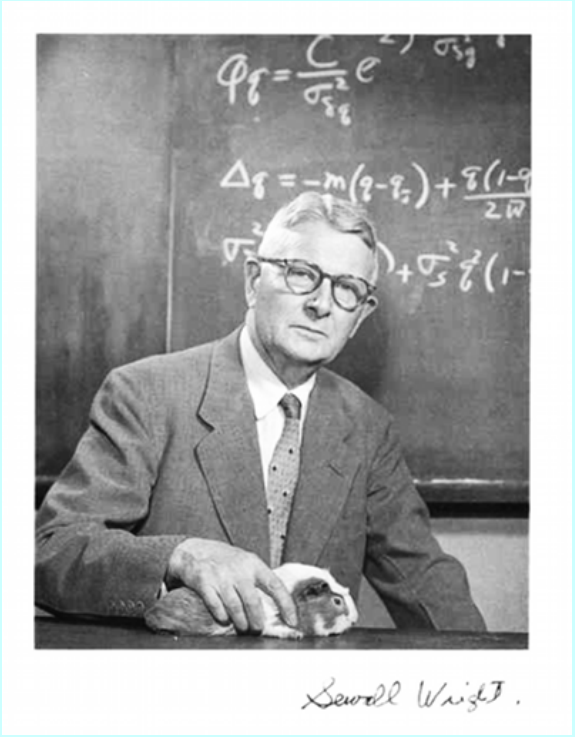


$$0 \underset{\beta\, dt}{\overset{\alpha\, dt}{\rightleftharpoons}} 1$$

Pagel allows inference of whether change is correlated, on a known tree. Lewis infers the tree, but does not allow for correlations among characters. Neither takes into account contributions to a 0/1 character from multiple underlying loci.

The threshold model

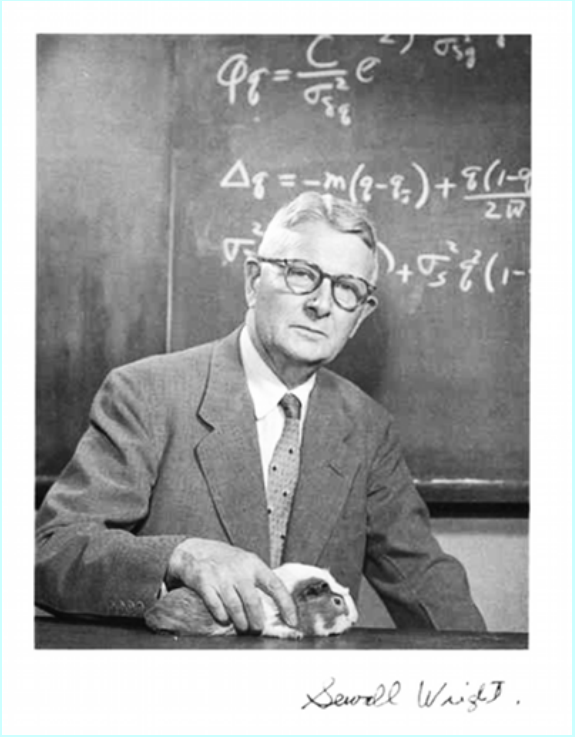
A relevant model was invented in 1934 by



Sewall Wright (1889-1988)
shown here in 1954

The threshold model

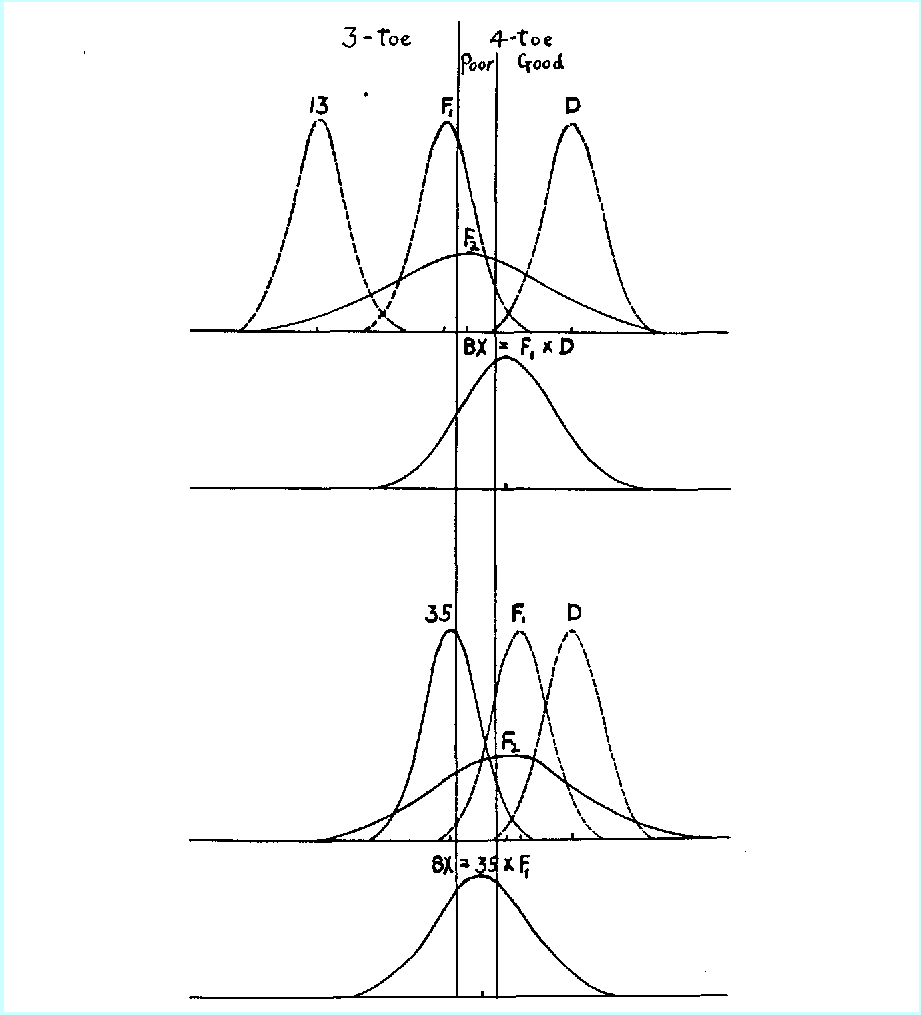
A relevant model was invented in 1934 by



Sewall Wright (1889-1988)

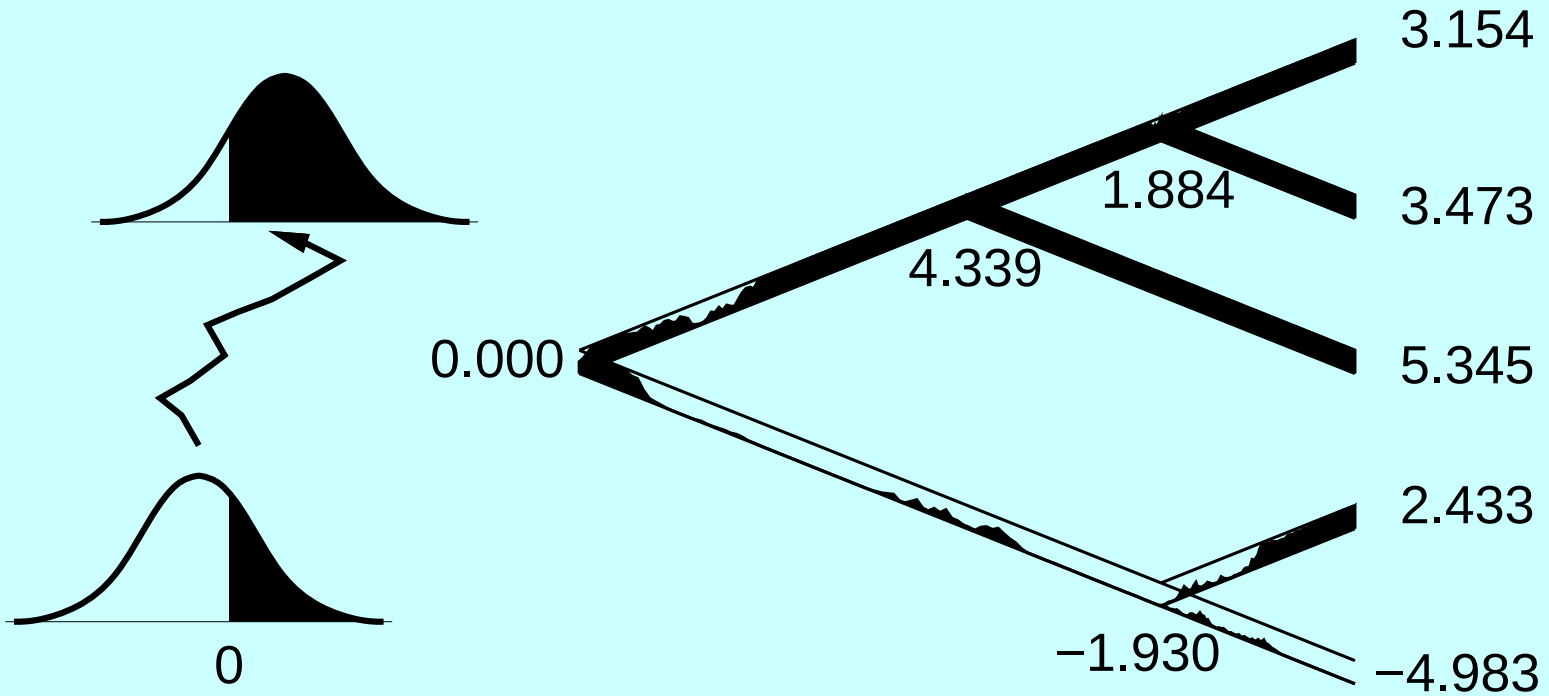
rumor has it he then absent-mindedly
erased the board with the guinea pig

The threshold model



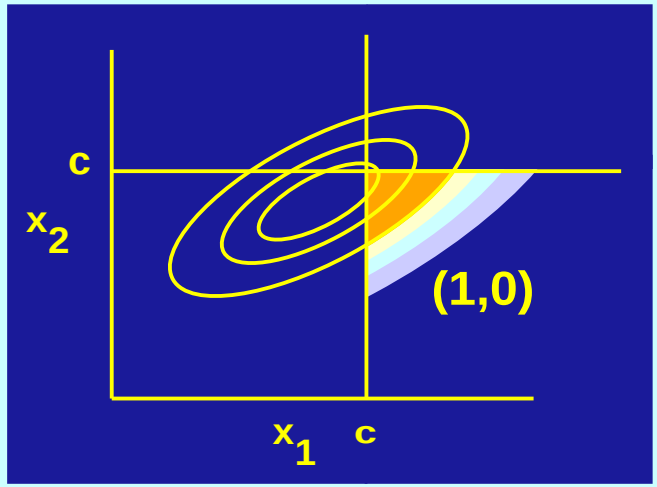
Sewall Wright (1934), guinea pig digit number
(from Wright's follow-up 1934 second paper)

The threshold model on a tree



Computing the likelihood

With two species, one character:



Disadvantages:

Quite hard to compute likelihoods: need to compute area in a corner of a correlated multivariate normal distribution.

With 5 species, one character:

$$\begin{aligned} L &= \text{Prob}\,(1,1,0,1,1) \\ &= \int_0^\infty \int_0^\infty \int_{-\infty}^0 \int_0^\infty \int_0^\infty \varphi(x_1, x_2, x_3, x_4, x_5 \mid \text{Tree})\, dx_1\, dx_2\, dx_3\, dx_4\, dx_5 \end{aligned}$$

Likelihoods under the threshold model on a tree

To compute the likelihood for a tree under the threshold model with p characters, want to compute:

$$\mathsf{L} \;=\; \int_{\mathsf{c}}^{\infty} \int_{-\infty}^{\mathsf{c}} \cdots \int_{\mathsf{c}}^{\infty} \; |\mathbf{V}|^{-1} (2\pi)^{-\mathsf{np}/2}$$

$$\times \exp\left(-\tfrac{1}{2}\,(\mathbf{x}-\mu)^{\mathsf{t}}\,\mathbf{V}^{-1}\,(\mathbf{x}-\mu)\right)\; \mathsf{dx}_{11}\,\mathsf{dx}_{12}\ldots\mathsf{dx}_{\mathsf{np}}$$

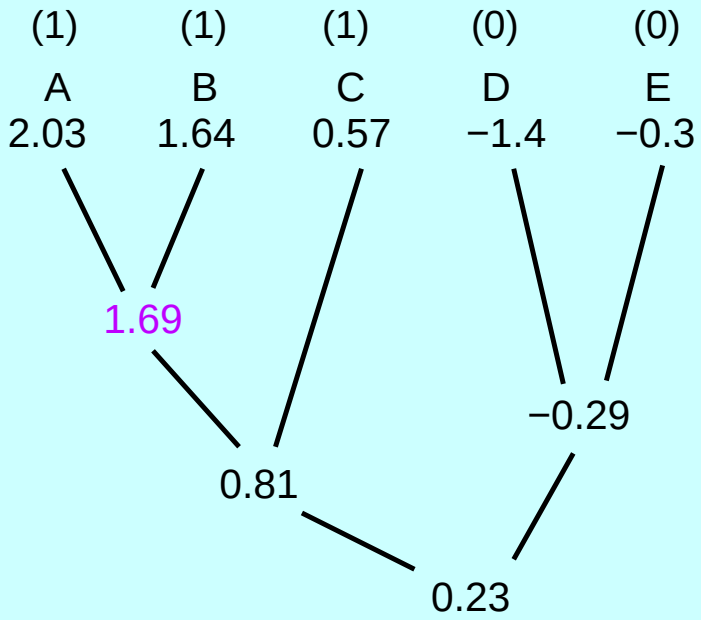
where μ is the appropriate vector of means, and

$$\mathbf{V} \;=\; \mathbf{A} \otimes \mathbf{T}$$

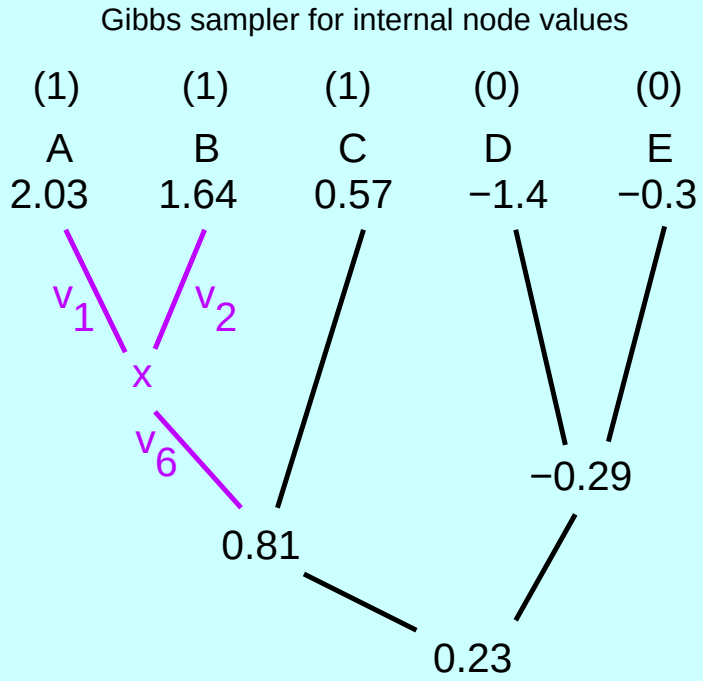
involves the tree and the “evolutionary" covariance matrix of the characters.

In other words, the probability density of the (unknown) liabilities gets integrated over the region of their values that corresponds to the observed discrete characters.

MCMC on liabilities



MCMC on liabilities: Gibbs sampling in the interior

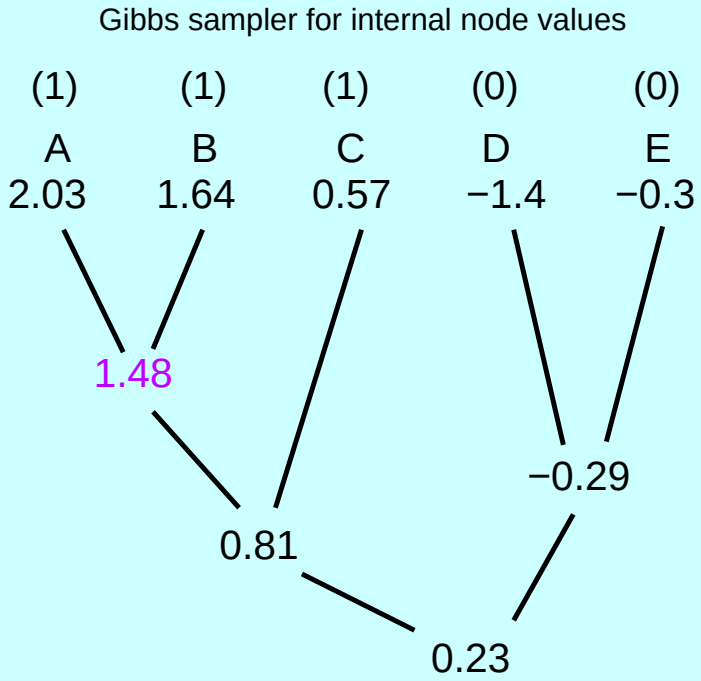


x drawn from normal distribution,

$$\text{mean} = \frac{(1/v_1)\,2.03 + (1/v_2)\,1.64 + (1/v_6)\,0.81}{(1/v_1) + (1/v_2) + (1/v_6)}$$

$$\text{var} = \frac{1}{(1/v_1) + (1/v_2) + (1/v_6)}$$

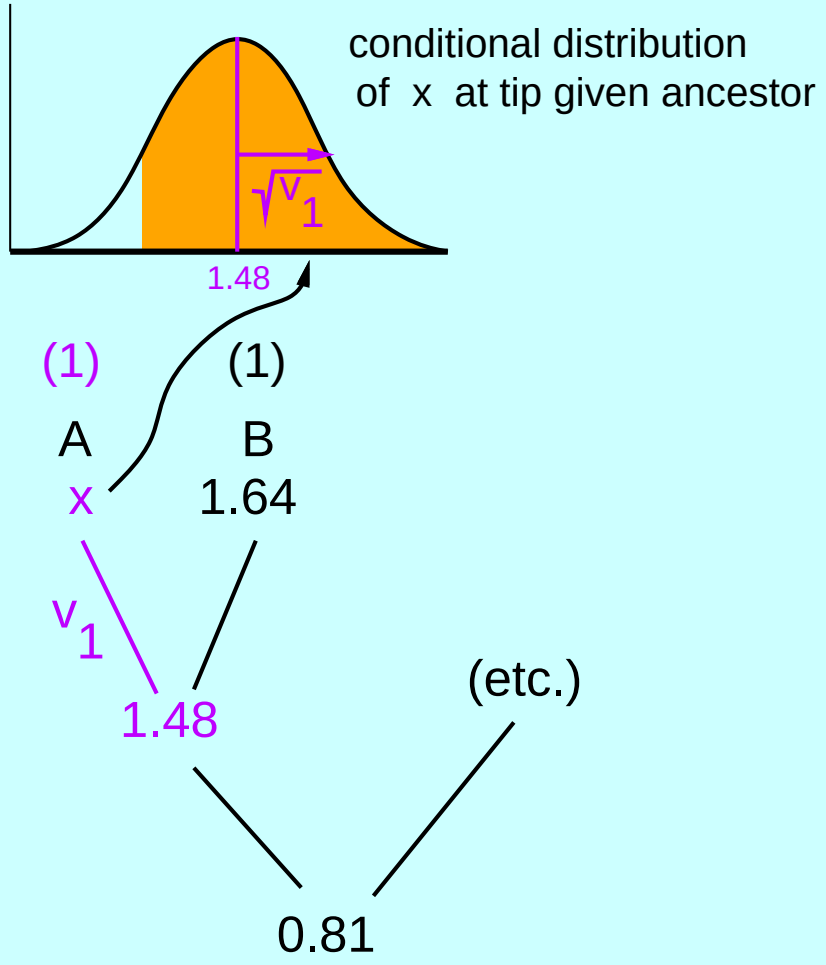
MCMC on liabilities: result of Gibbs sampling



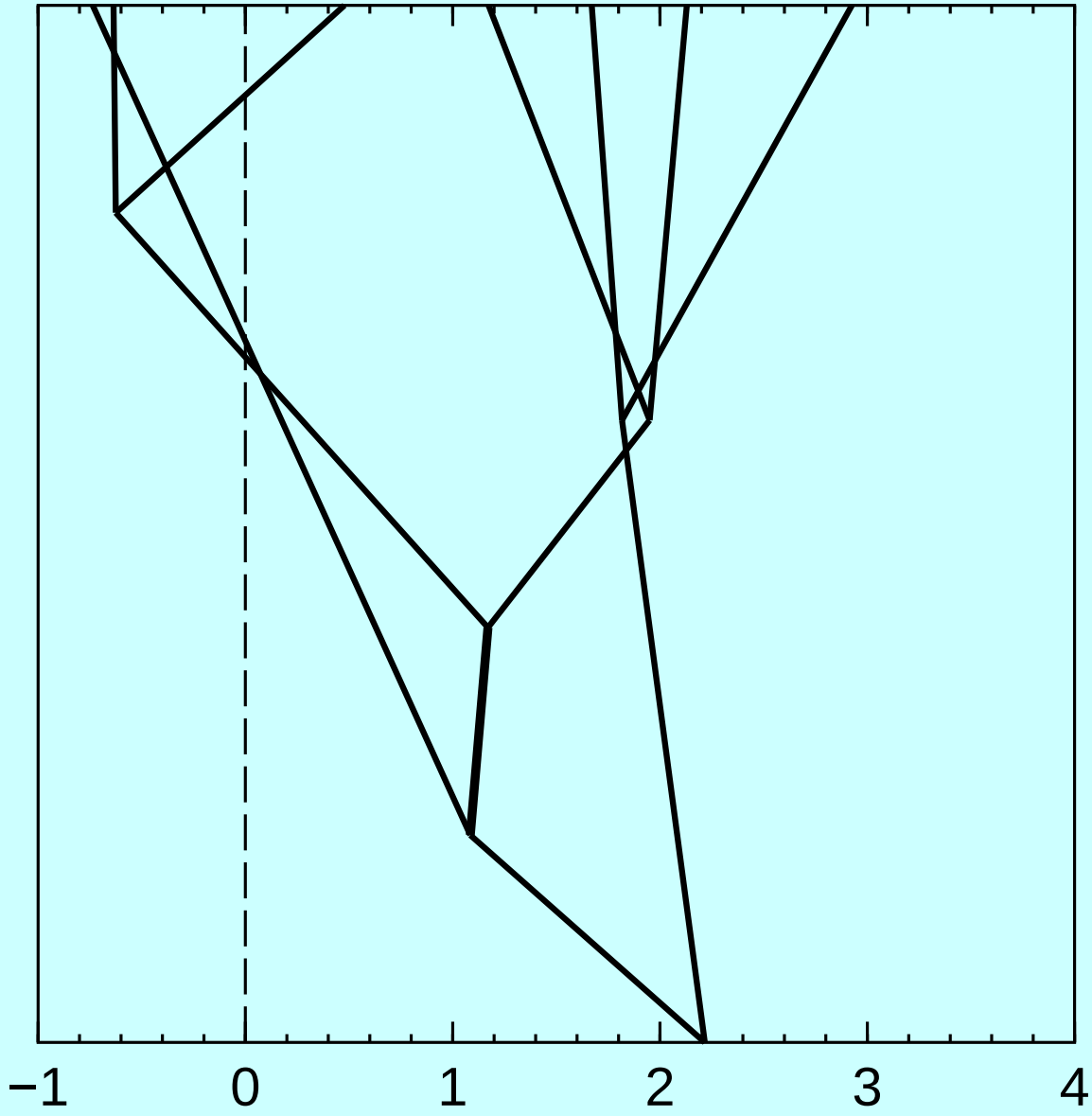
MCMC on liabilities: rejection at tips

How to update the liability at a tip?

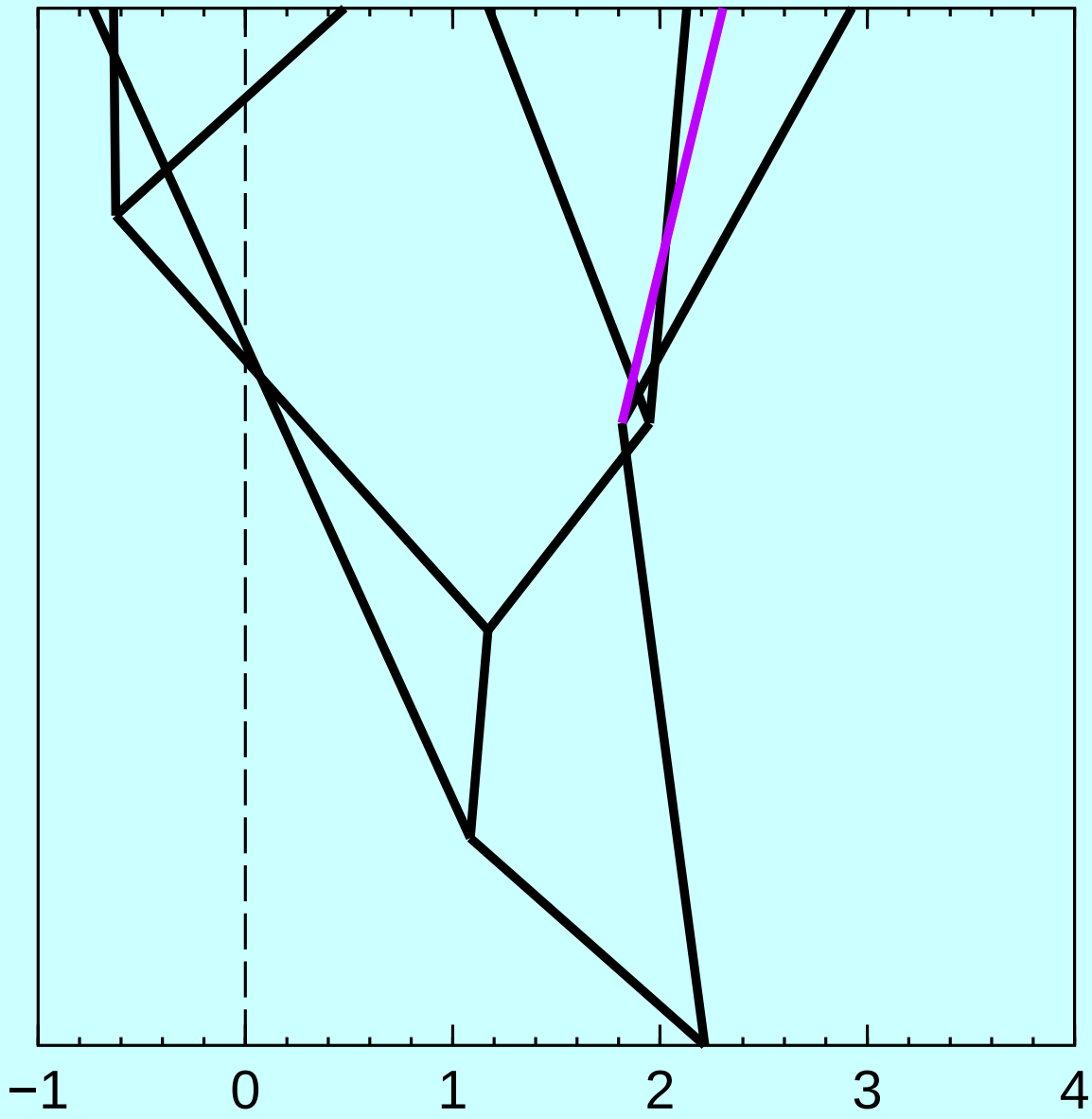
(must condition on ancestor and observed phenotype)



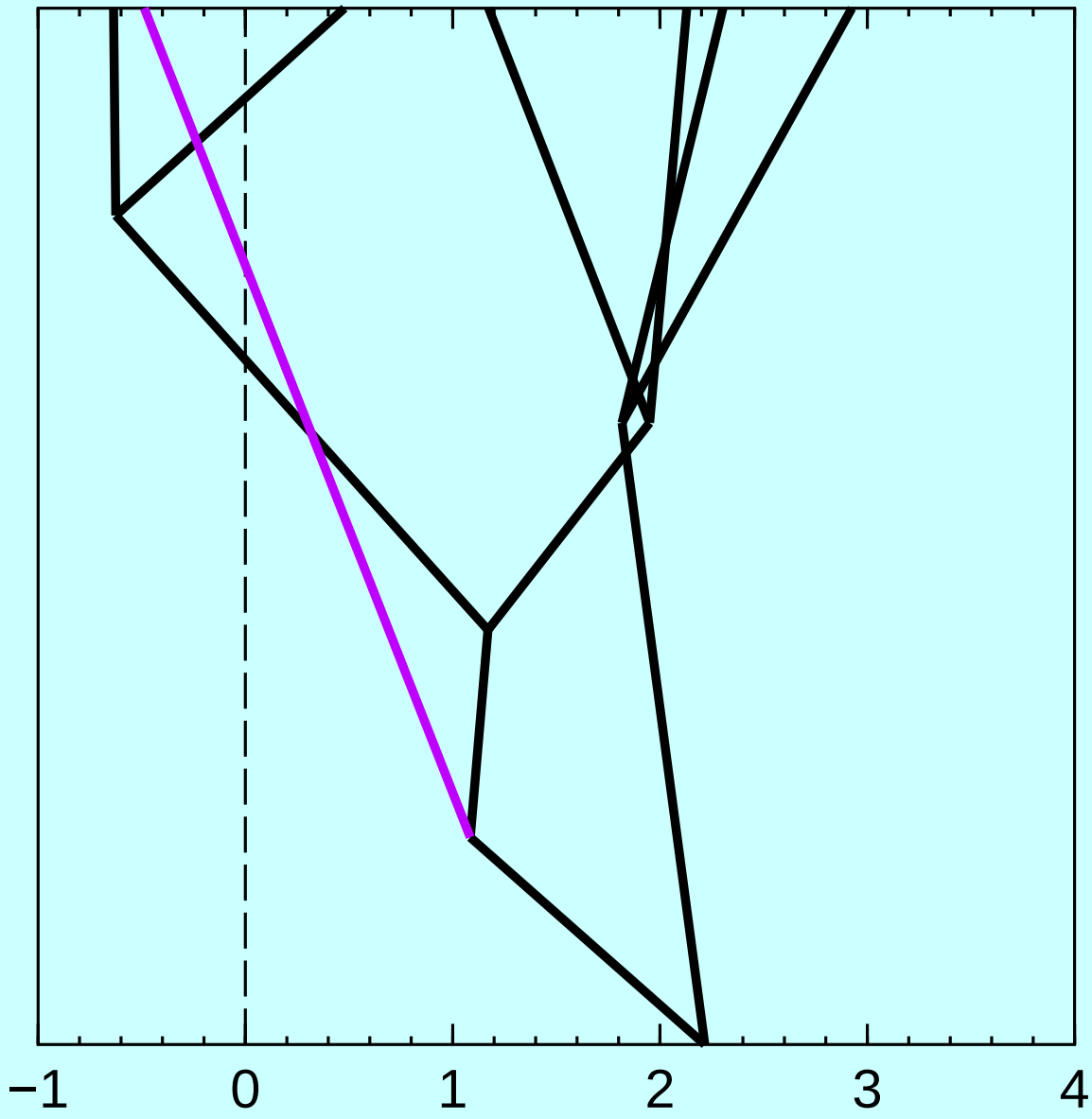
An example



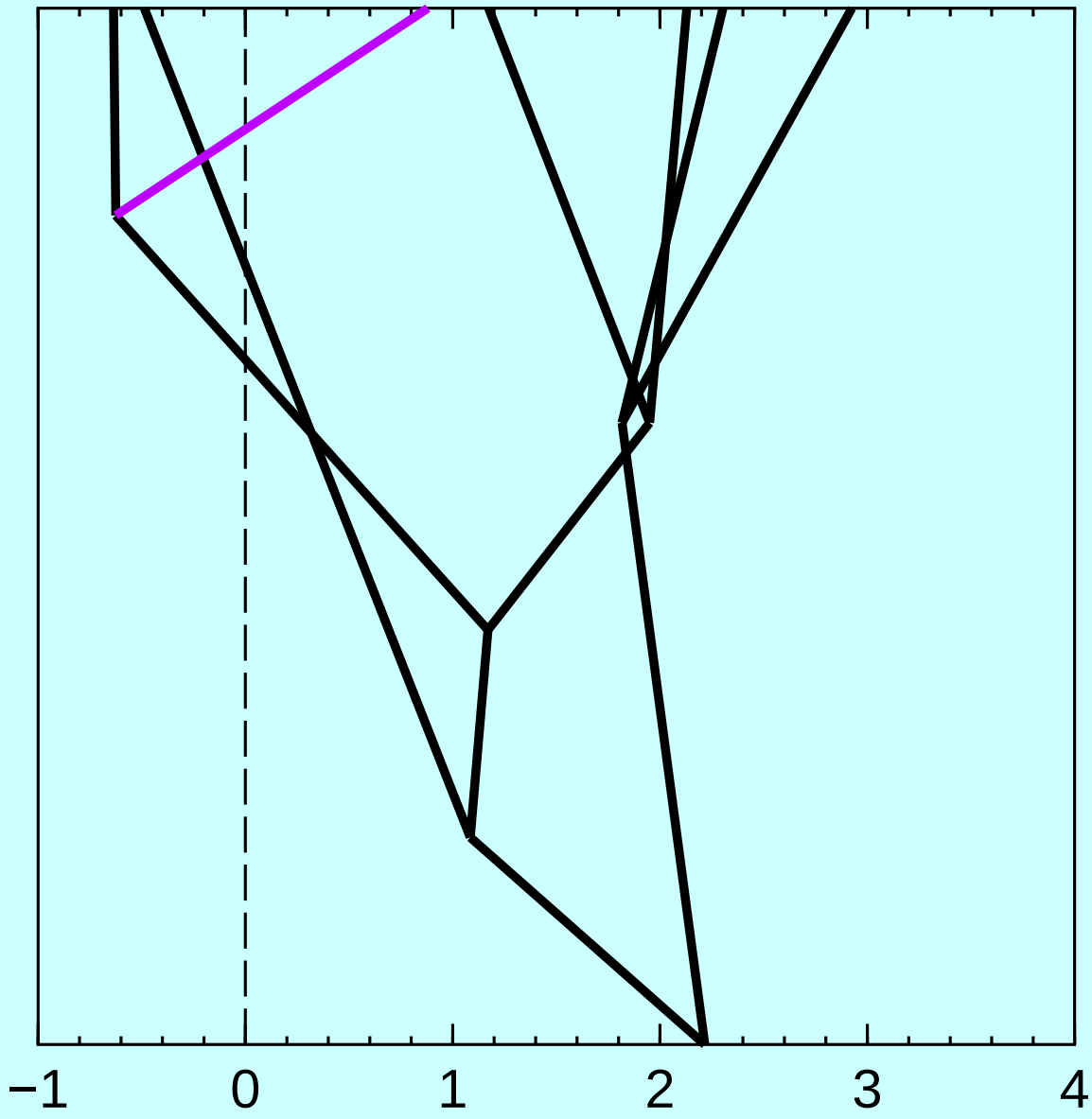
An example



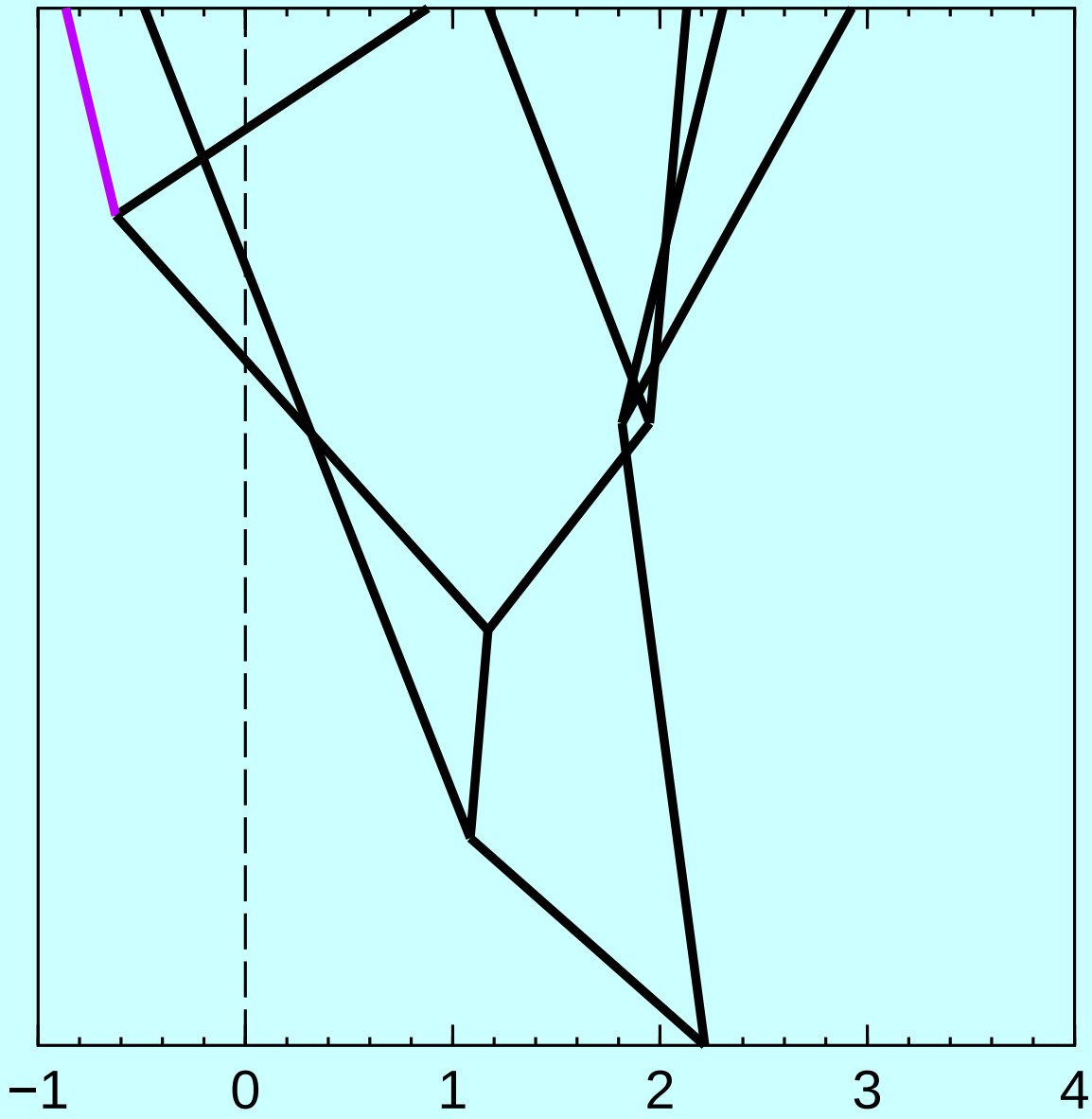
An example



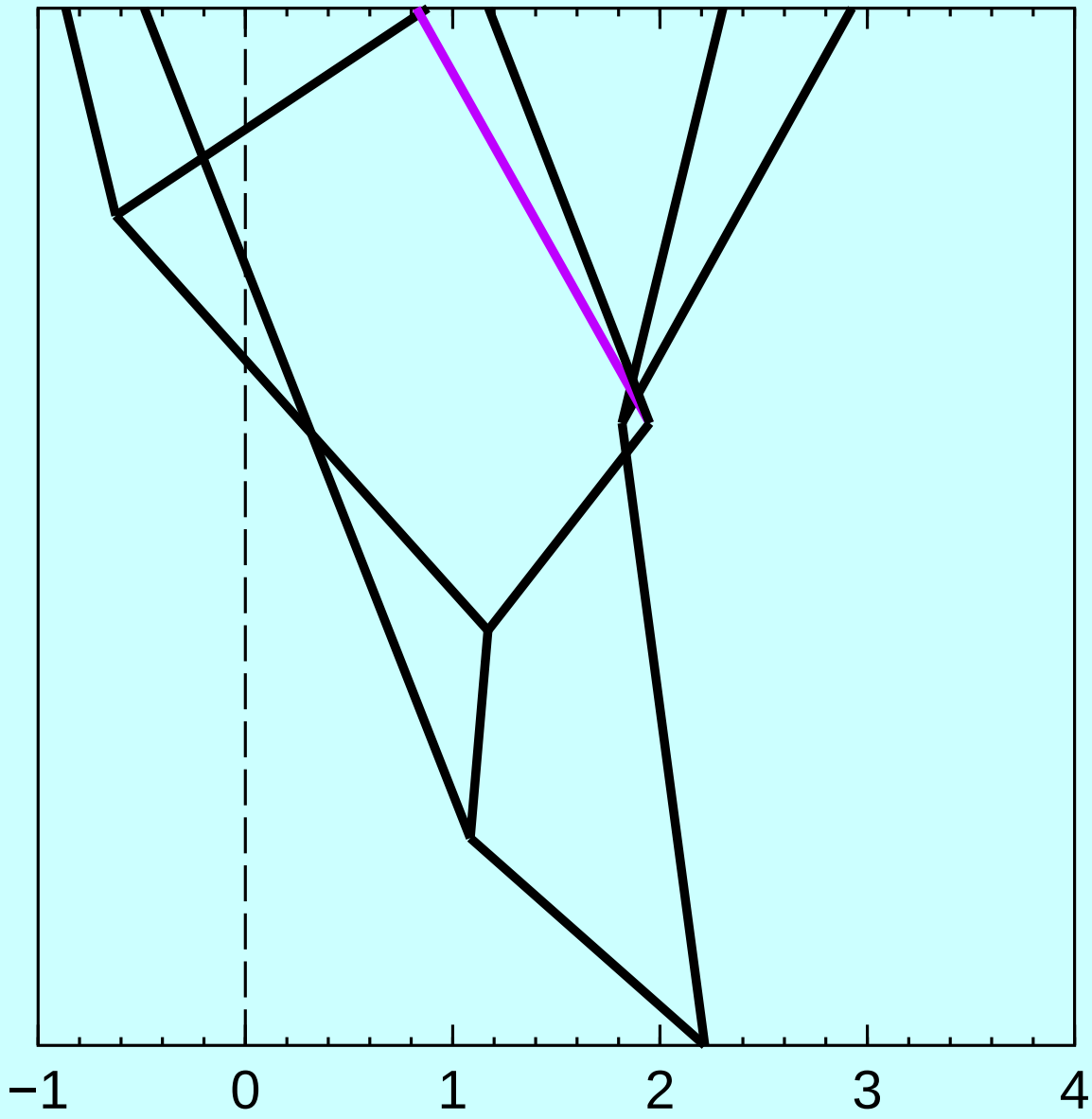
An example



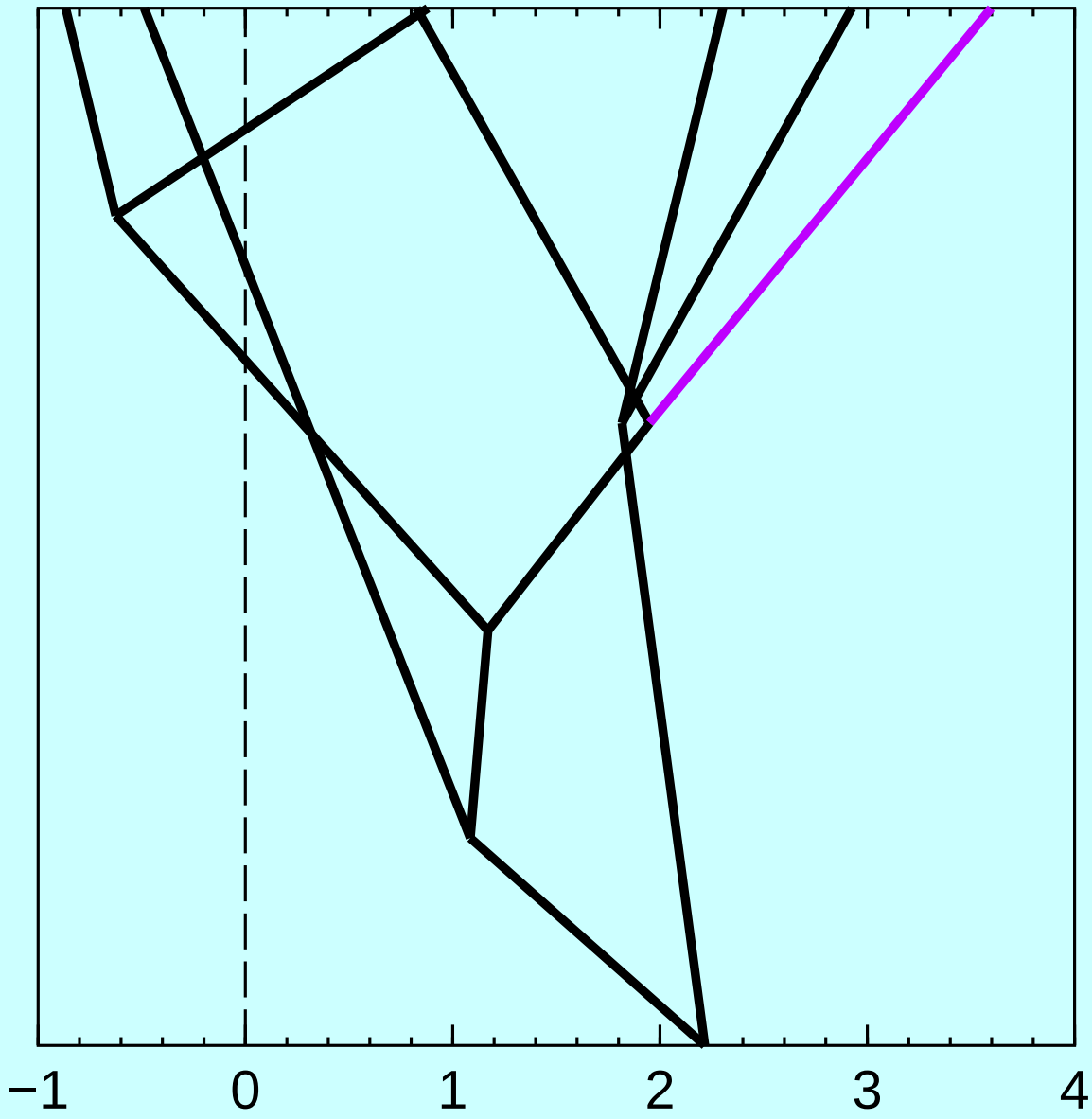
An example



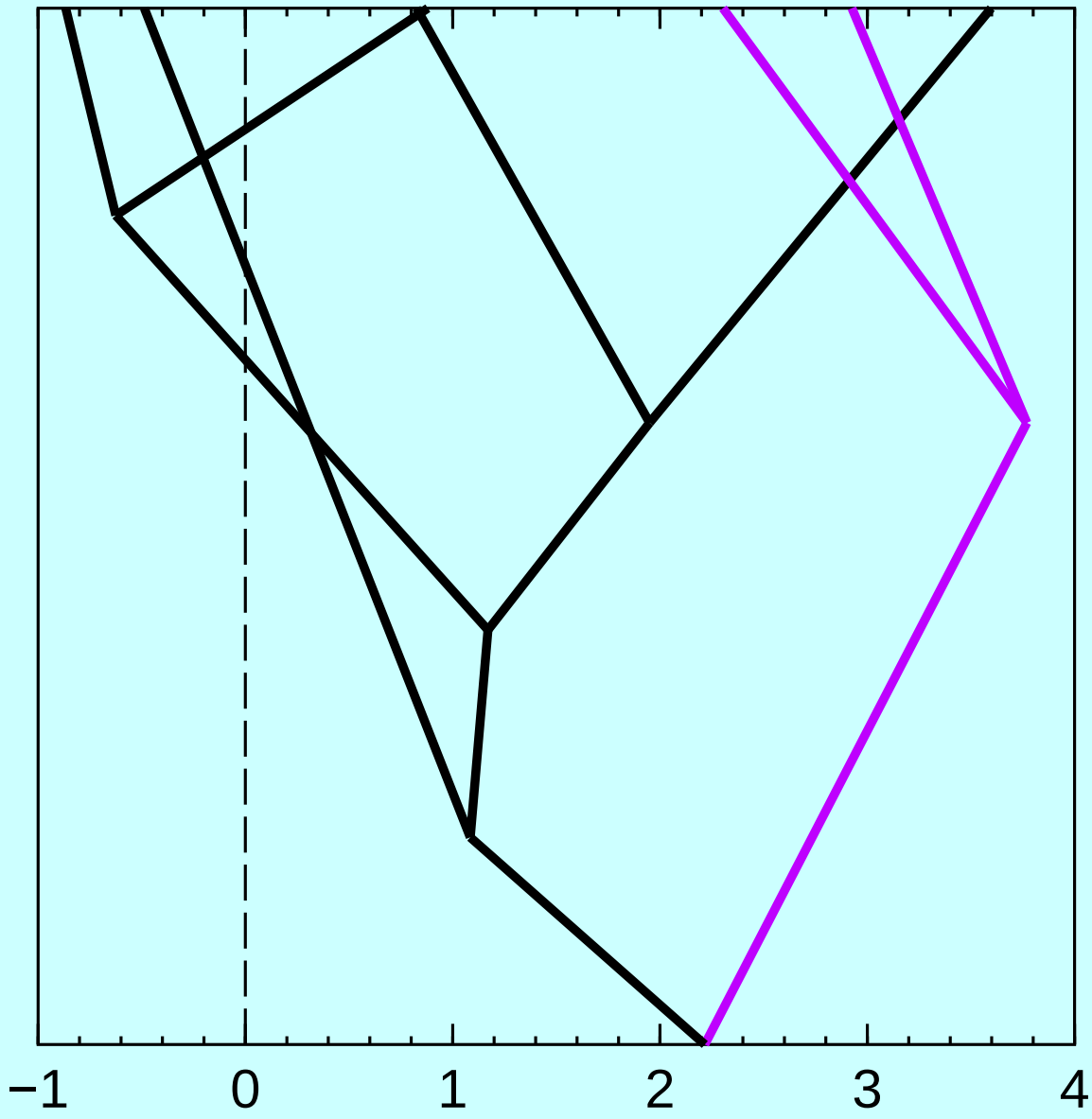
An example



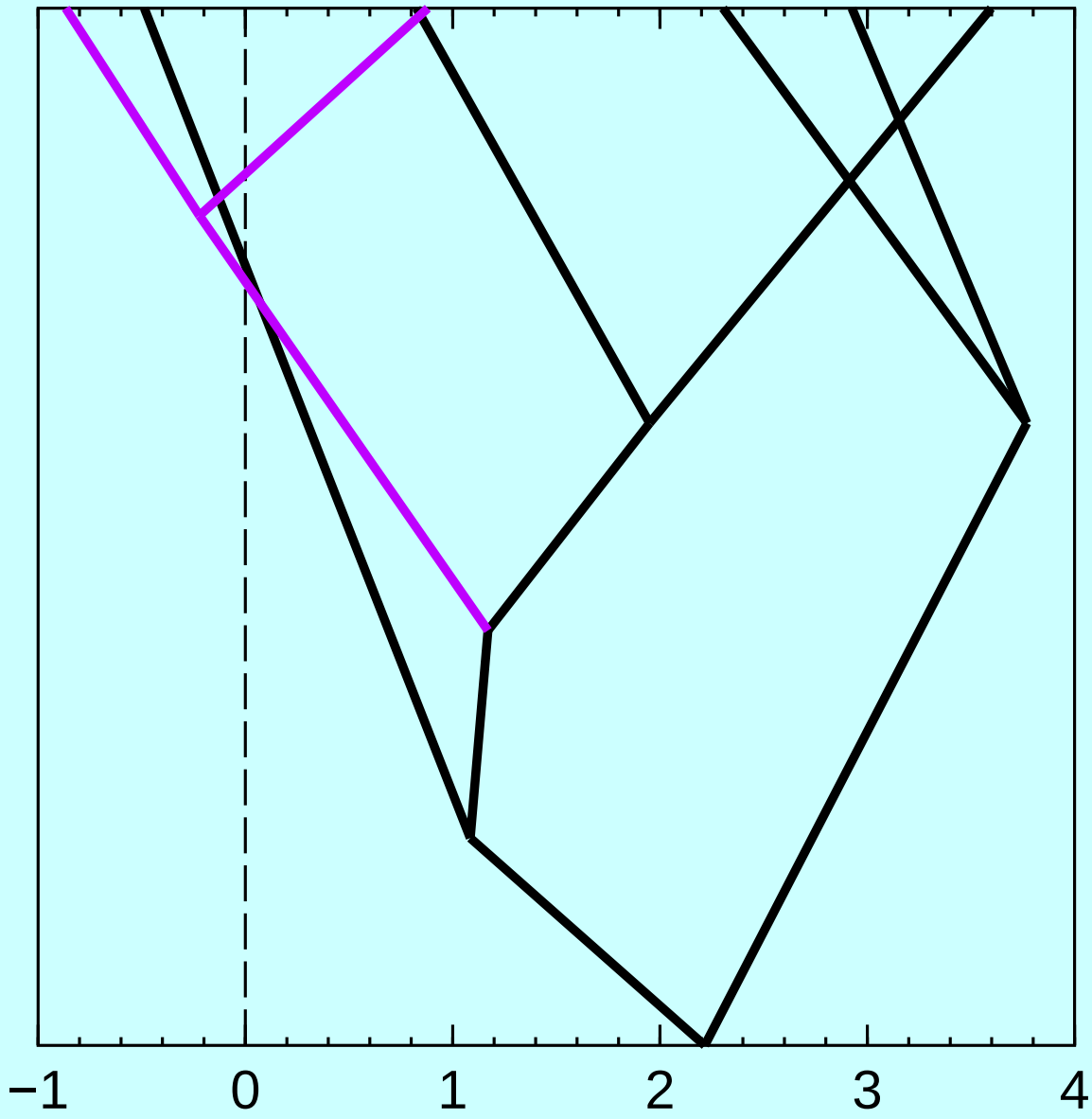
An example



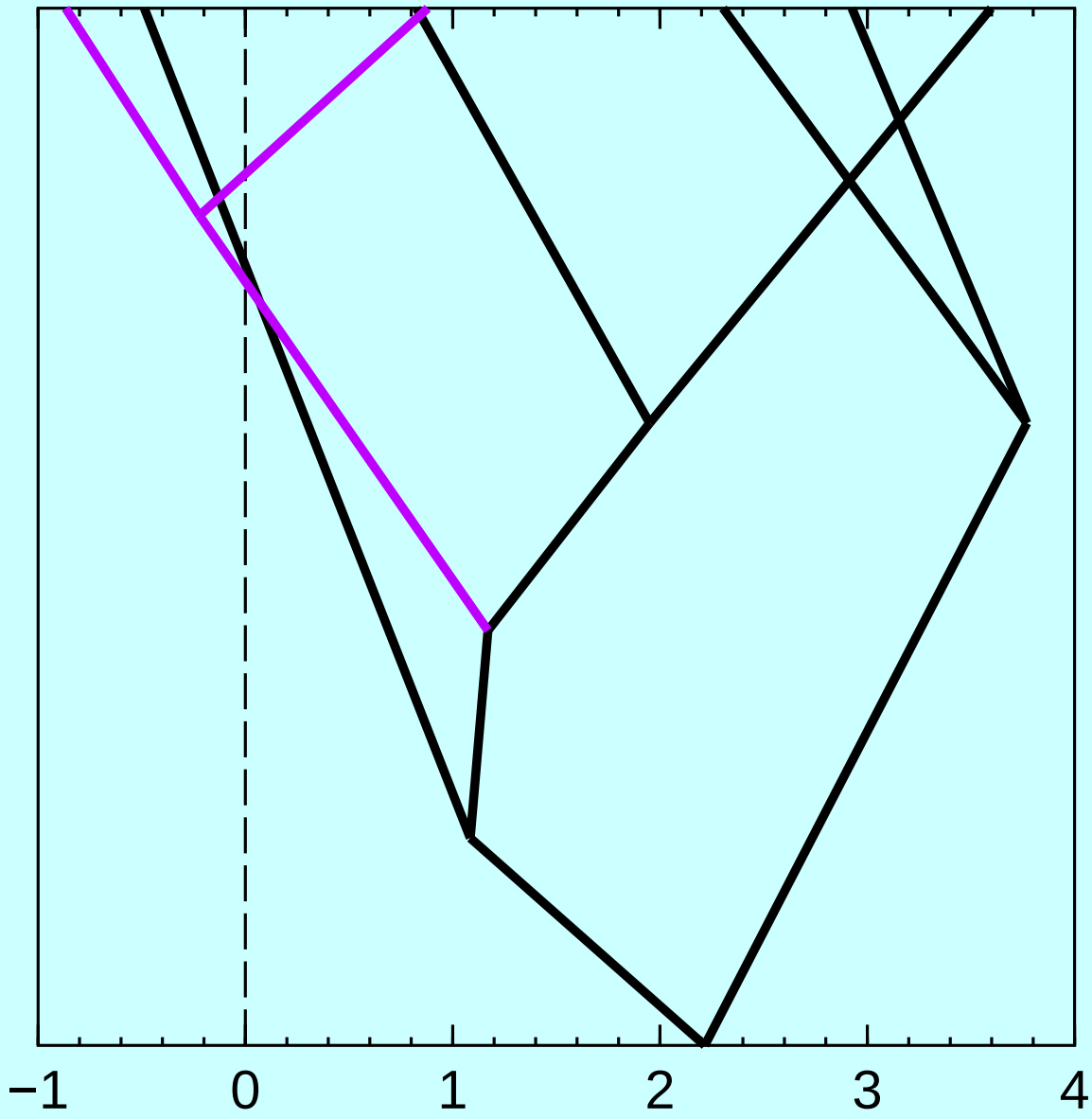
An example



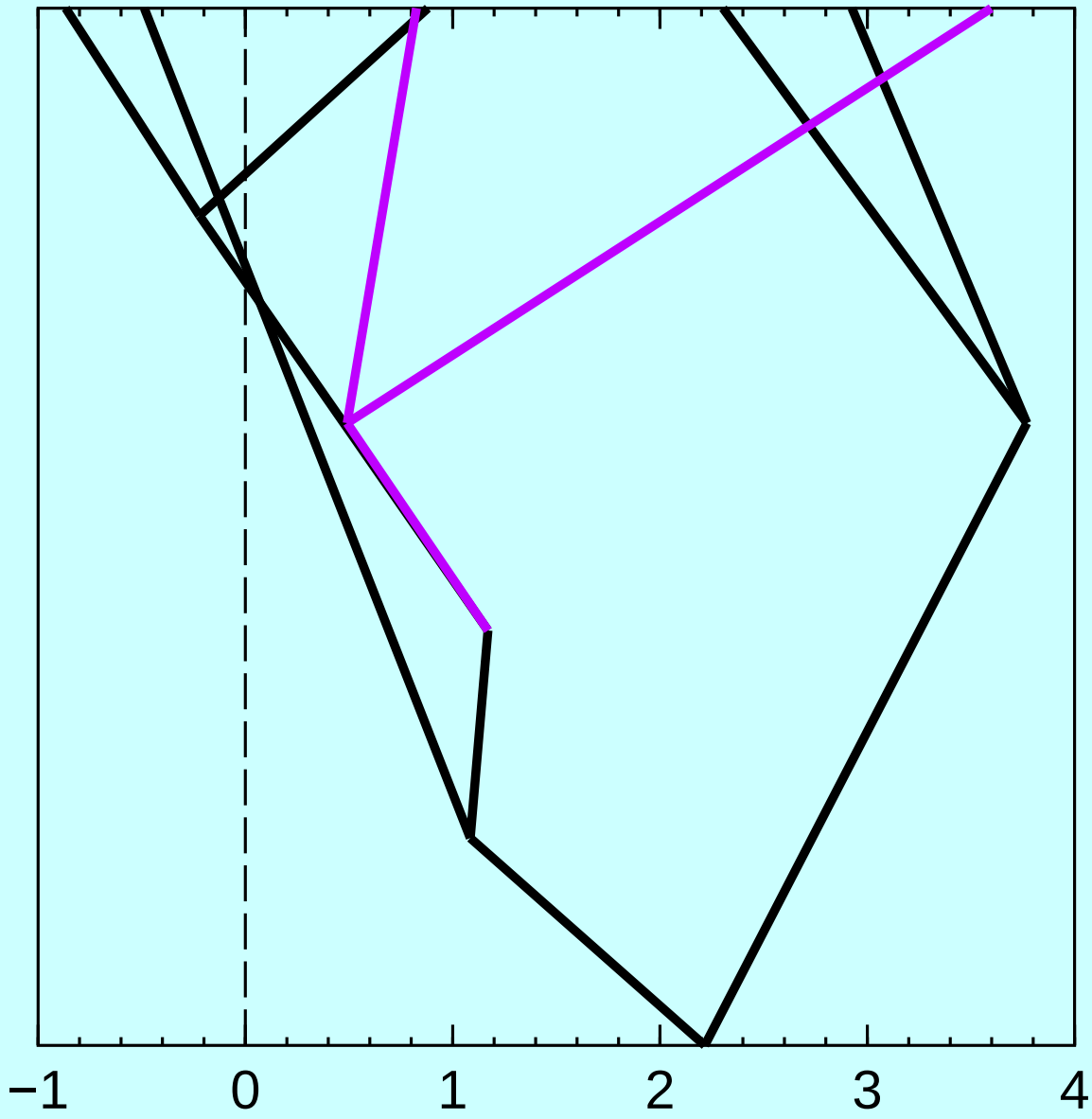
An example



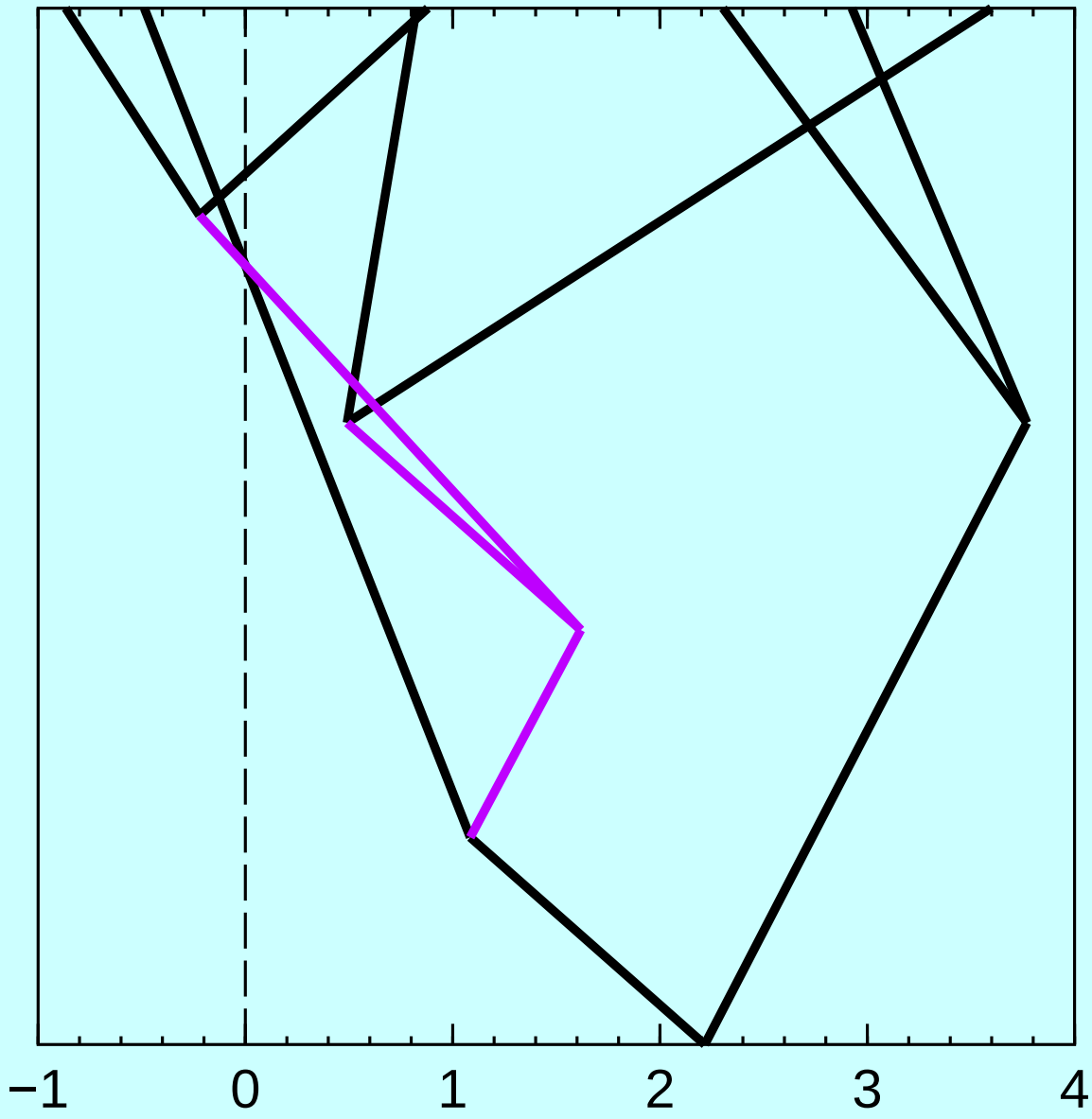
An example



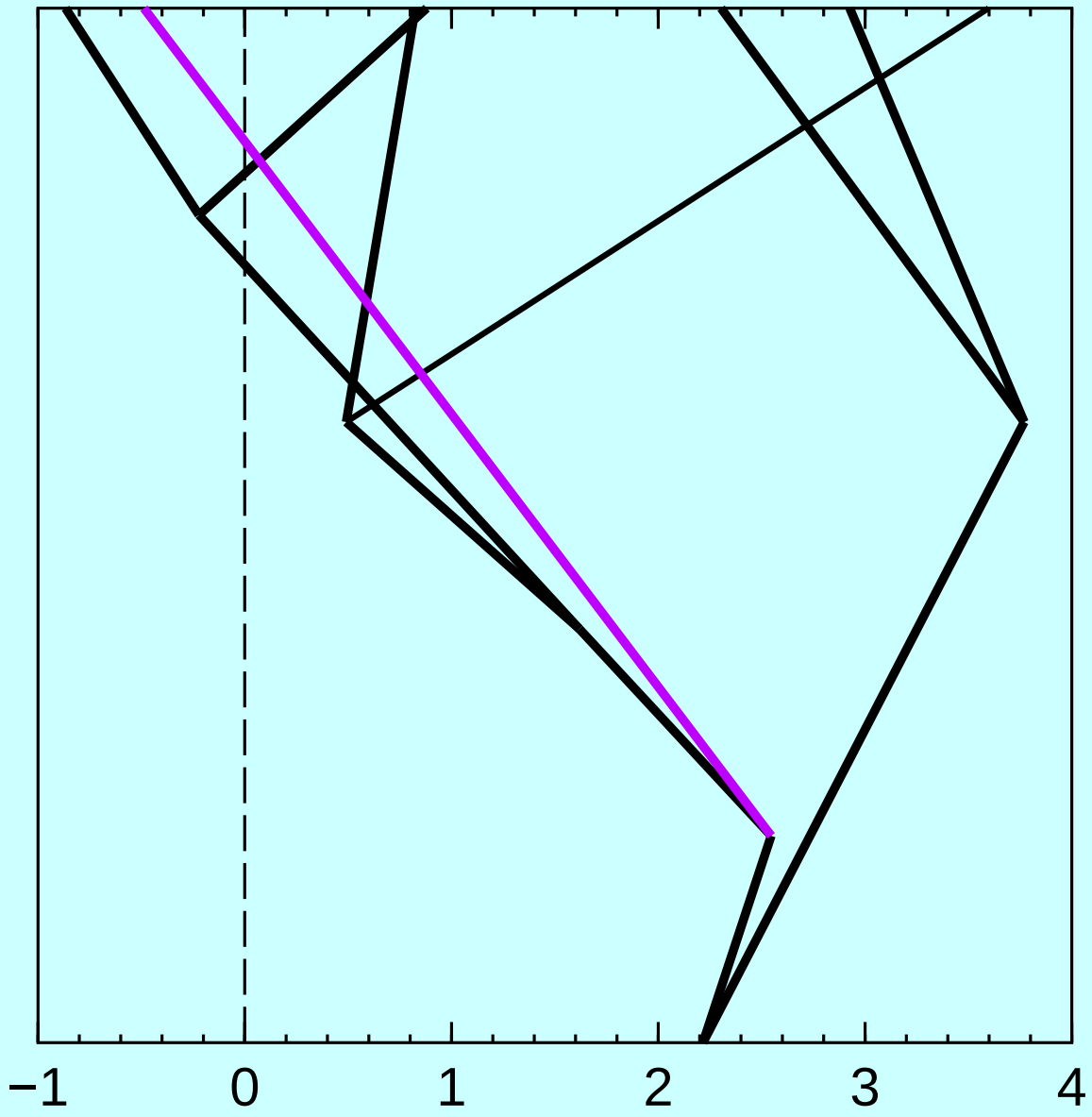
An example



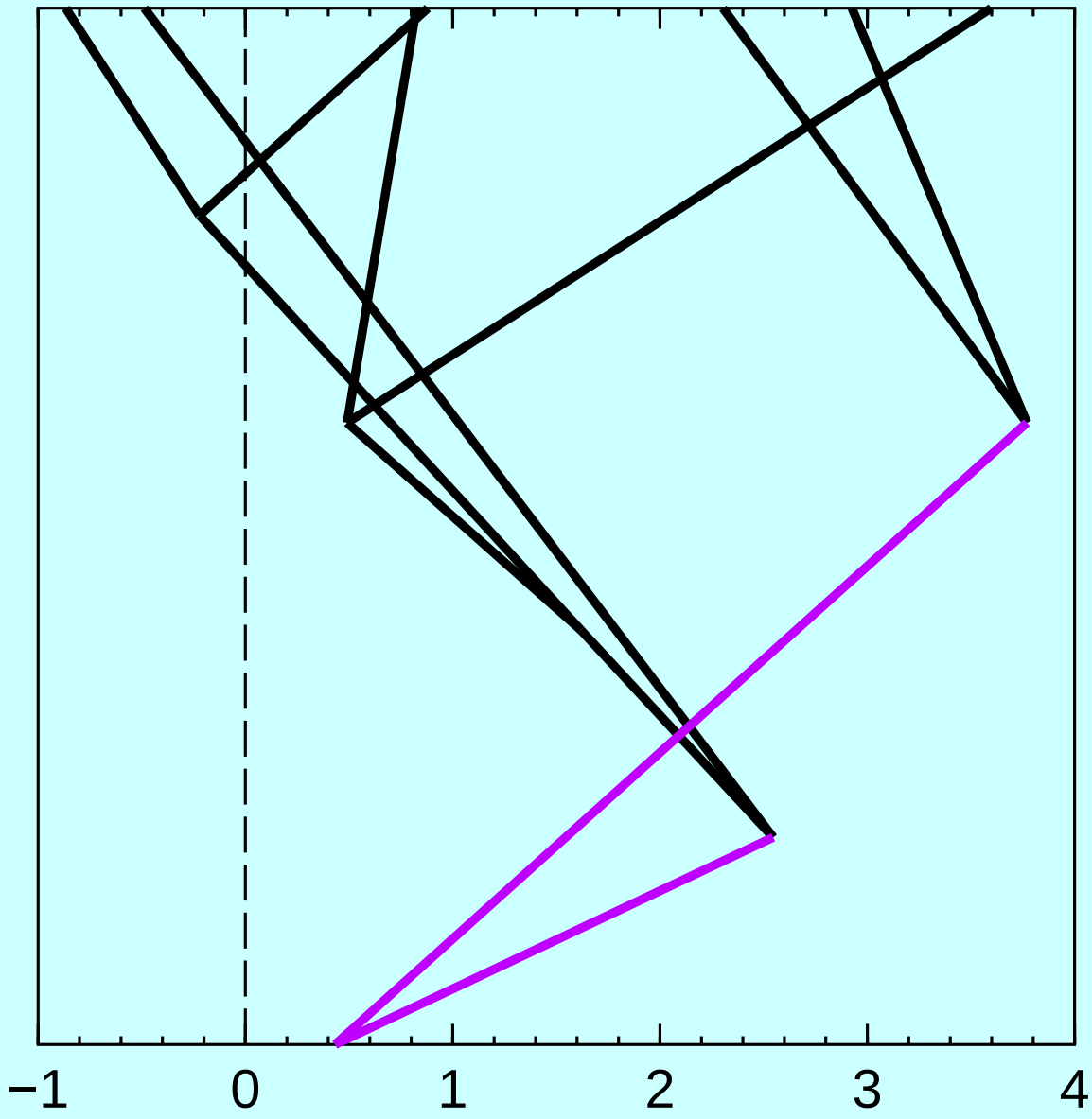
An example



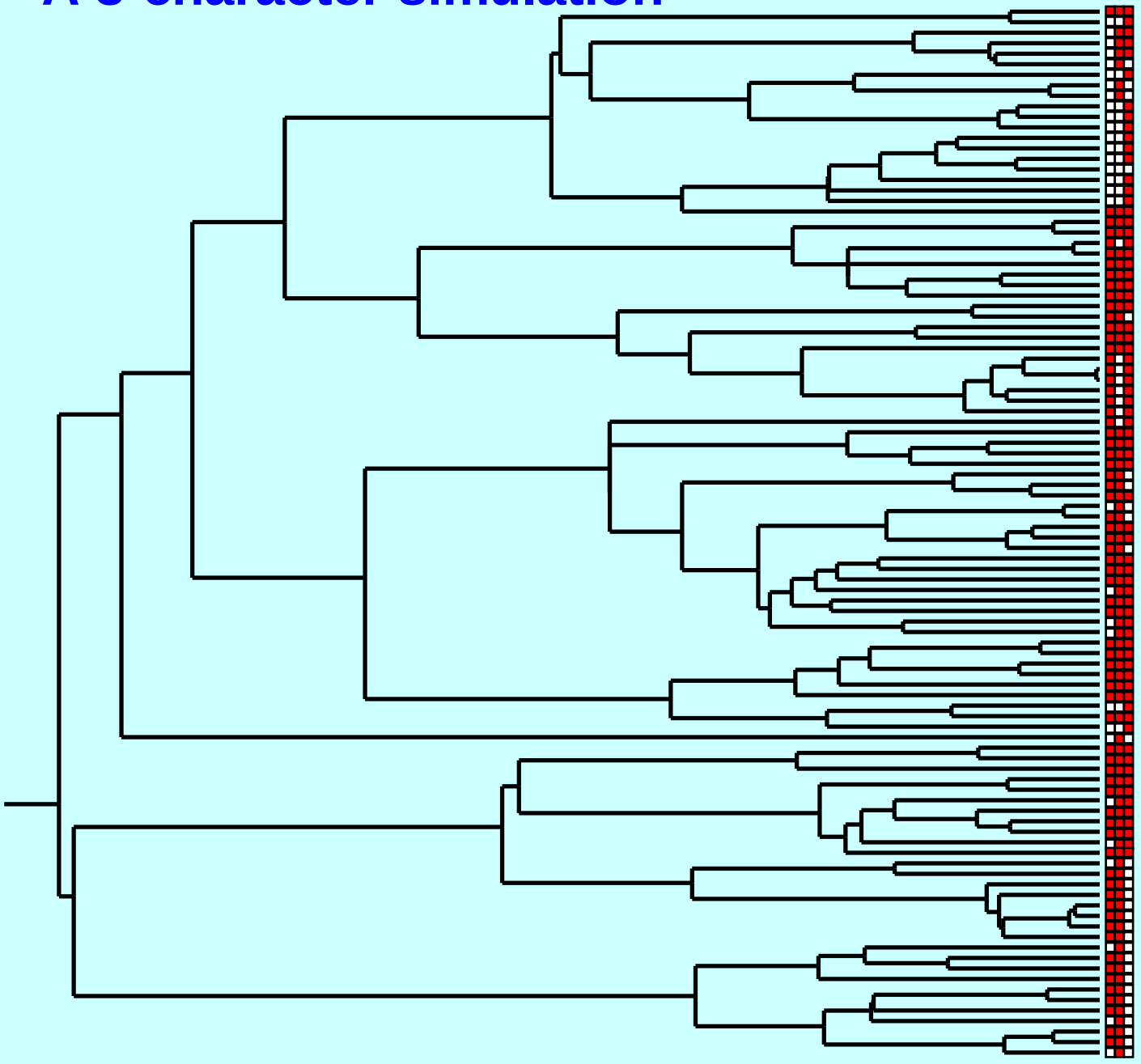
An example



An example



A 3-character simulation



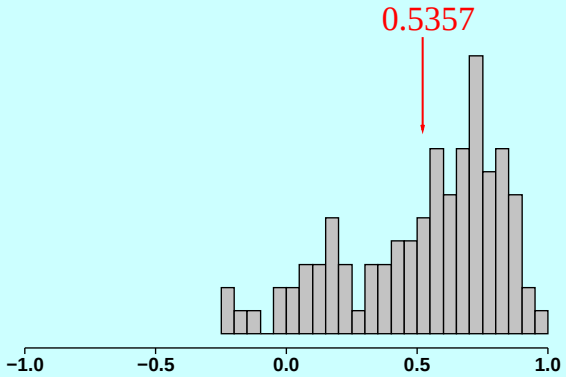
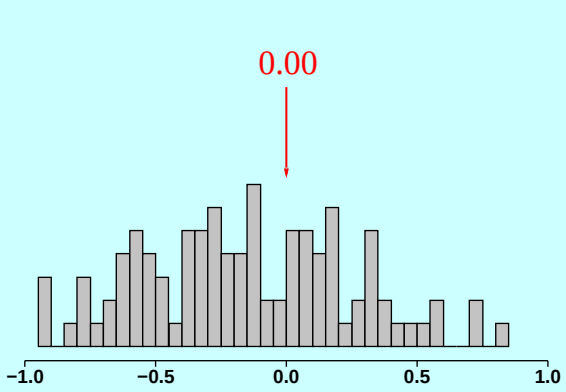
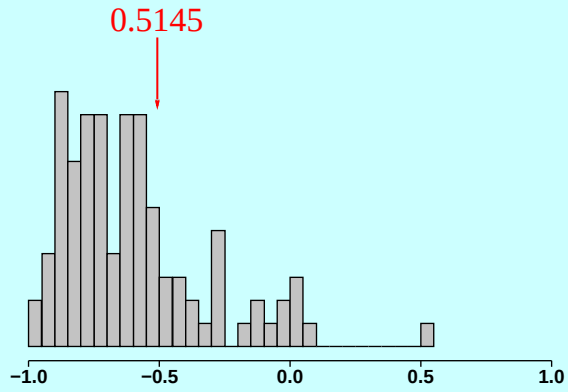
A 3-character simulation

For these true covariances:

$$\begin{bmatrix} 1.64 & 0.8 & 0 \\ 0.8 & 1.36 & -0.6 \\ 0 & -0.6 & 1 \end{bmatrix}$$

100 data sets with 100-species trees were analyzed.

Inferred correlation coefficients

	character 1	character 2
character 1		
character 2		
character 3		

Extensions

- To take uncertainty in the tree into account ...

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or
- ... use Bayesian posterior samples of trees, and in either case

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or
- ... use Bayesian posterior samples of trees, and in either case
- ... do the analysis once for each of these trees

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or
- ... use Bayesian posterior samples of trees, and in either case
- ... do the analysis once for each of these trees
- It is also possible to do an overall Bayesian inference if we have priors on the covariances that we are happy with. But this may turn out much the same.

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or
- ... use Bayesian posterior samples of trees, and in either case
- ... do the analysis once for each of these trees
- It is also possible to do an overall Bayesian inference if we have priors on the covariances that we are happy with. But this may turn out much the same.
- For within-species variation, where we have finite samples of specimens for each species, we can extend these methods using machinery similar to my 2008 paper on comparative methods with within-species variation.

Extensions

- To take uncertainty in the tree into account ...
- ... use bootstrap sampled trees, or
- ... use Bayesian posterior samples of trees, and in either case
- ... do the analysis once for each of these trees
- It is also possible to do an overall Bayesian inference if we have priors on the covariances that we are happy with. But this may turn out much the same.
- For within-species variation, where we have finite samples of specimens for each species, we can extend these methods using machinery similar to my 2008 paper on comparative methods with within-species variation.
- I have not talked about the Ornstein-Uhlenbeck model of character change but will in the next hour, where we will cover adaptive peaks, multiple peaks, and moving peaks. *OK, just kidding!*

What about QTLs?

- In principle, in the long run, we can integrate this work with QTL inference.

What about QTLs?

- In principle, in the long run, we can integrate this work with QTL inference.
- Perhaps this would enable us to use phylogenetic information to not only identify QTLs, but to see them change across species, including some QTLs causing variation within some species, some within others.

What about QTLs?

- In principle, in the long run, we can integrate this work with QTL inference.
- Perhaps this would enable us to use phylogenetic information to not only identify QTLs, but to see them change across species, including some QTLs causing variation within some species, some within others.
- In principle, this could even allow us to infer on which of two correlated characters the selection really acted (or to what extent and in what direction on both).

What we can ... and cannot ... infer

- BUT ... we have limited power from any one sample of species. Biologists must learn to accept that, and find ways to propagate that uncertainty through the analysis that flow from these inferences. We cannot (ever!) have a Fly-On-The-Wall account of evolution.

What we can ... and cannot ... infer

- BUT ... we have limited power from any one sample of species. Biologists must learn to accept that, and find ways to propagate that uncertainty through the analysis that flow from these inferences. We cannot (ever!) have a Fly-On-The-Wall account of evolution.
- Furthermore we must always be sensitive to the limits of our models – as we expand the tree to less related groups, the models are called severely into question.

The Reunion

- For the last 40-50 years population-genetic work within species has been (mostly) isolated from work on molecular evolution between species.
- Now we are in a gradual Reunion of these two lines of work (*not a New Synthesis, though*) as observations can be made that connect them (coalescents across species boundaries, Ds/Dn inferences, etc.)
- As this happens, Russ Lande's vision will become more and more of a reality – quantitative genetics will become directly relevant to multi-species evolutionary biology.

References

Lande, R. 1976. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30 (2):** 314-334. **[One of Russ's major papers using the constant-variances approximation]**

Stebbins, G. L. 1950. Variation and Evolution in Plants. Columbia University Press, New York. **[Describes selective covariance and cites Tedin (1926) for it]**

Pyron, R. A. 2011. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60:** 466-481. **[Pioneering paper on using morphology and molecules to place fossils and date divergences]**

Ronquist, F., S. Klopfstein, L. Vilhelmsen, S. Schulmeister, D. L. Murray, and A. P. Rasnitsyn. 2012. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology*, Early Access, 10.1093/sysbio/sys058 **[Another pioneering paper on using morphology and molecules to place fossils and date divergences]**

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* 50: 913-925. **[Uses 0/1 stochastic process to infer morphological phylogenies]**

References

Pagel, M. 1994. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London Series B Biological Sciences.* **255:** 37-45. **[0/1 stochastic model for discrete characters]**

Wright, S. 1934. An analysis of variability in number of digits in an inbred strain of guinea pigs. *Genetics* **19:** 506-536. **[The threshold model for discrete traits]**

Falconer, D. S. 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* **29:** 51-76. **[Threshold model applied to human diseases]**

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* **19:** 445-471. **[Review with mention of usefulness of threshold model]**

Felsenstein, J. 2002. Quantitative characters, phylogenies, and morphometrics.pp. 27-44 in “Morphology, Shape, and Phylogenetics", ed. N. MacLeod. Systematics Association Special Volume Series 64. Taylor and Francis, London. **[Review repeating 1988 material and going into some more detail on the question of threshold models.]**

References

Felsenstein, J. 2004. *Inferring Phylogenies.* Sinauer Associates, Sunderland, Massachusetts. **Mentions this model and also sample size issues in contrasts method.**

Felsenstein, J. 2005. Using the quantitative genetic threshold model for inferences between and within species. *Philosophical Transactions of the Royal Society of London, series B* **360:** 1427-1434. **[The threshold project in a slightly earlier version]**

Felsenstein, J. 2008. Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *American Naturalist* **171:** 713-725. **[On how to allow for finite sample size and within-species phenotypic variation in comparative methods]**

Felsenstein, J. 2012. A comparative method for both discrete and continuous characters using the threshold model. *American Naturalist* **179:** 145-156. **[The threshold method for discrete characters, allowing covariation with continuous characters to be inferred too]**