

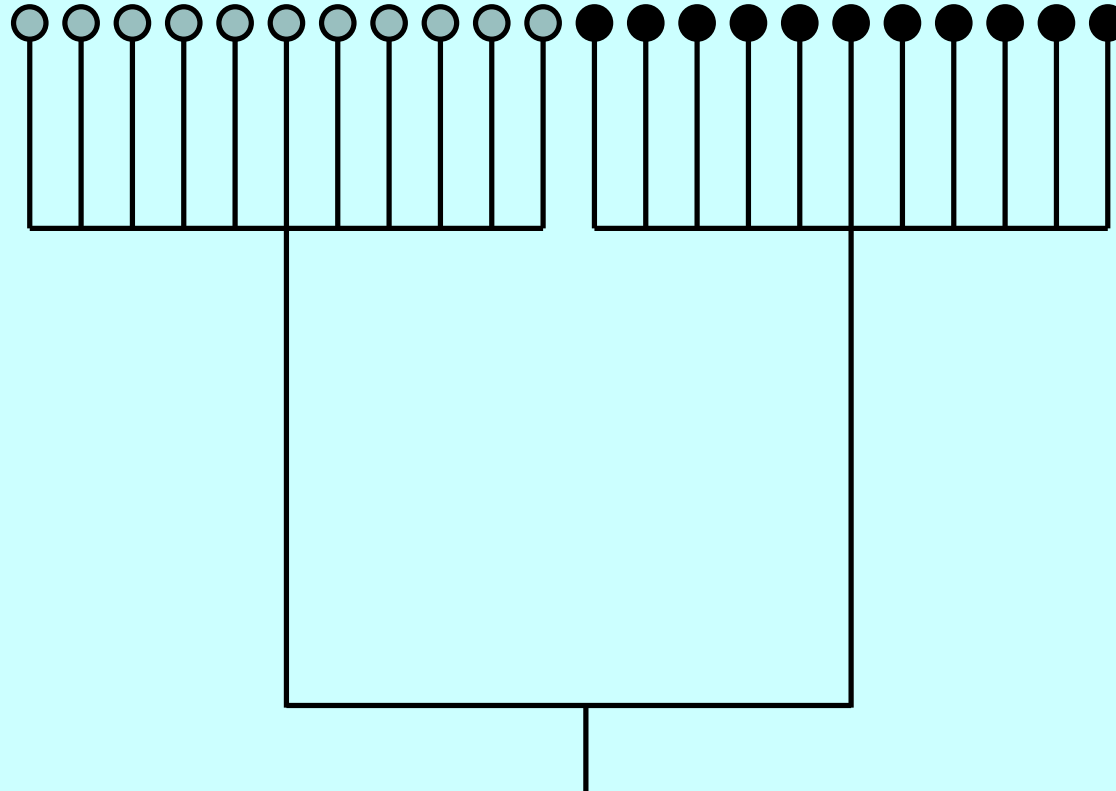
# Comparative method and phylogenies

8 August 2013.

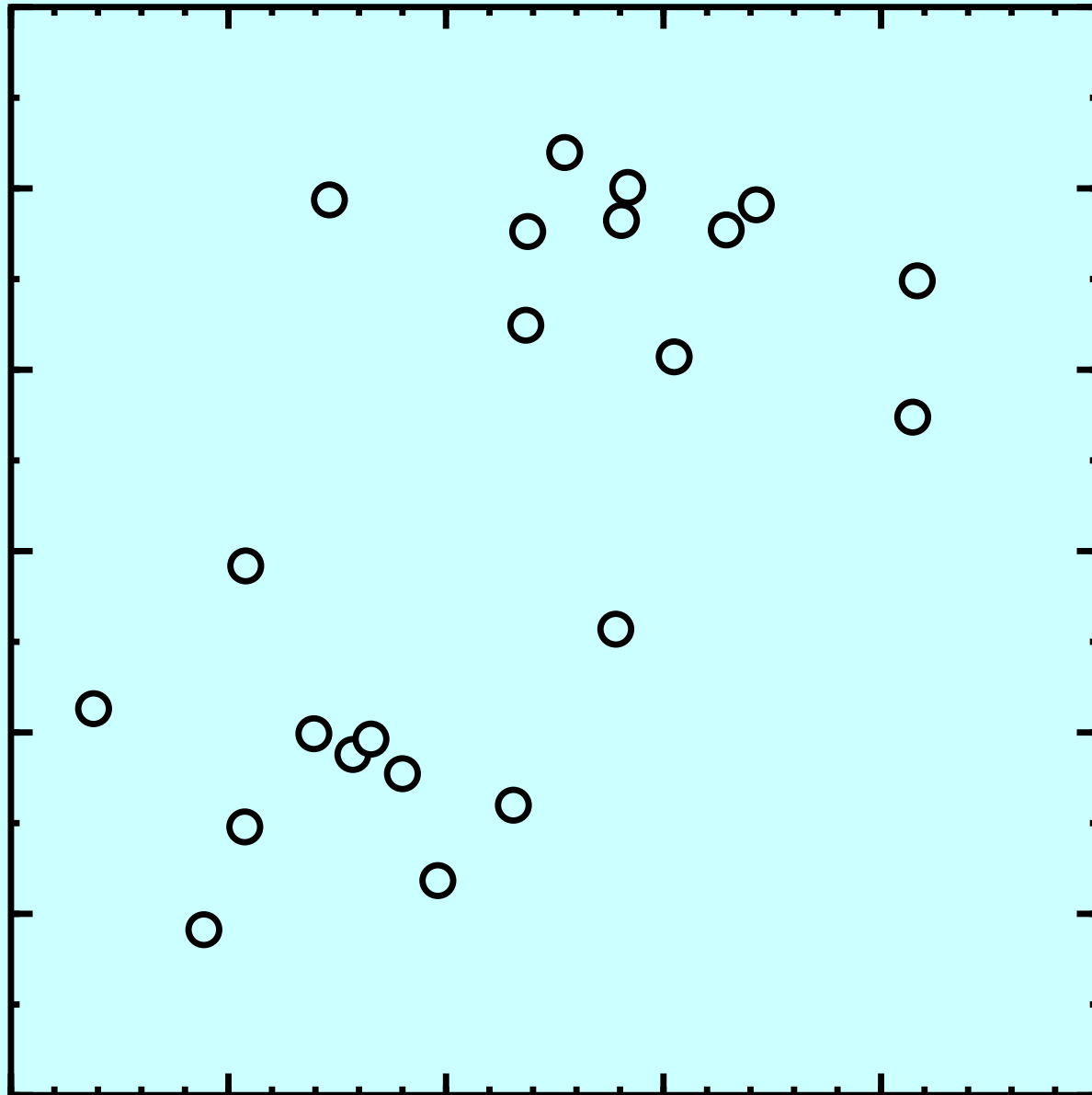
Joe Felsenstein

NESCENT Evol Quant Gen course

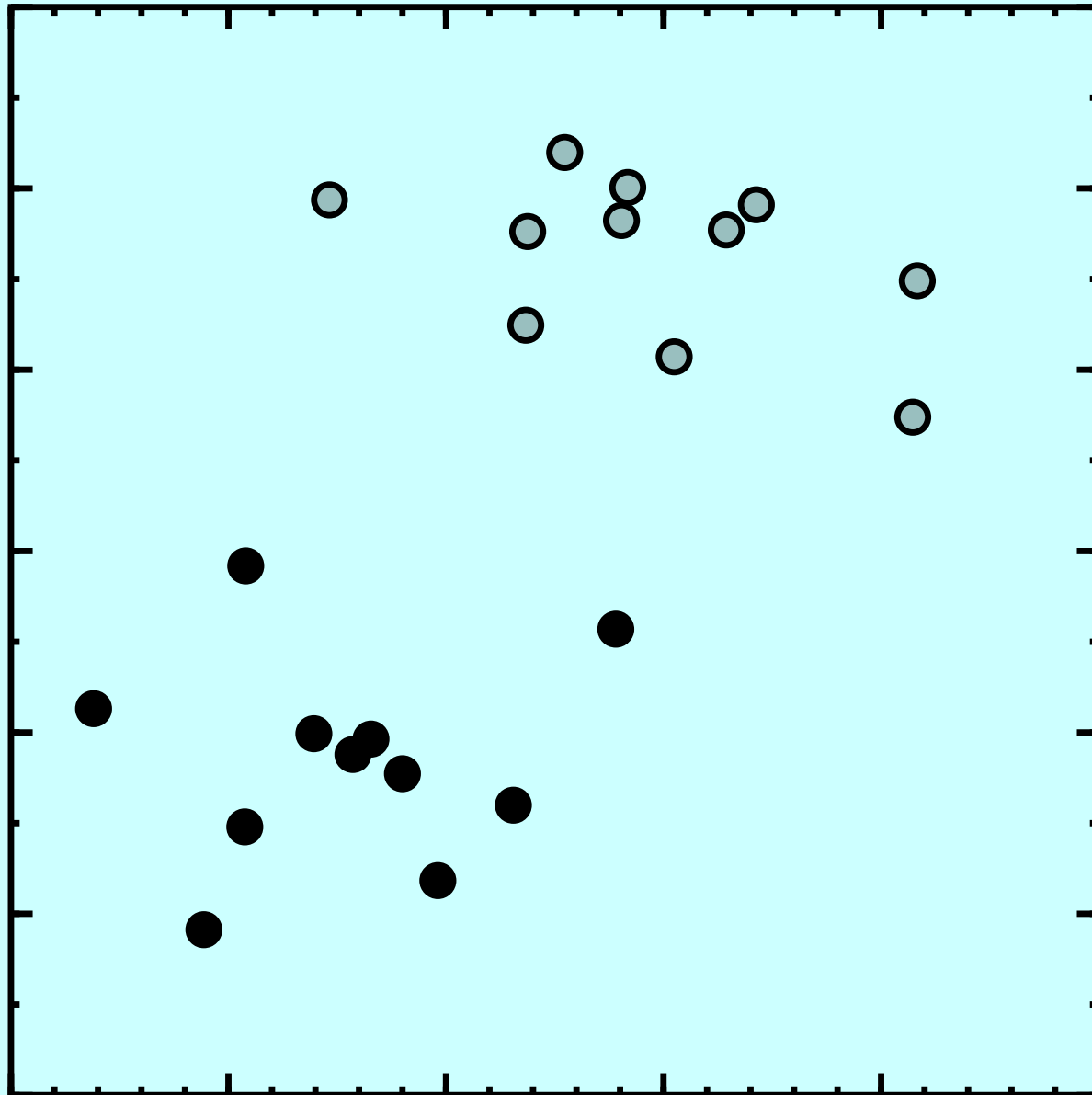
## A simple case to show effects of phylogeny



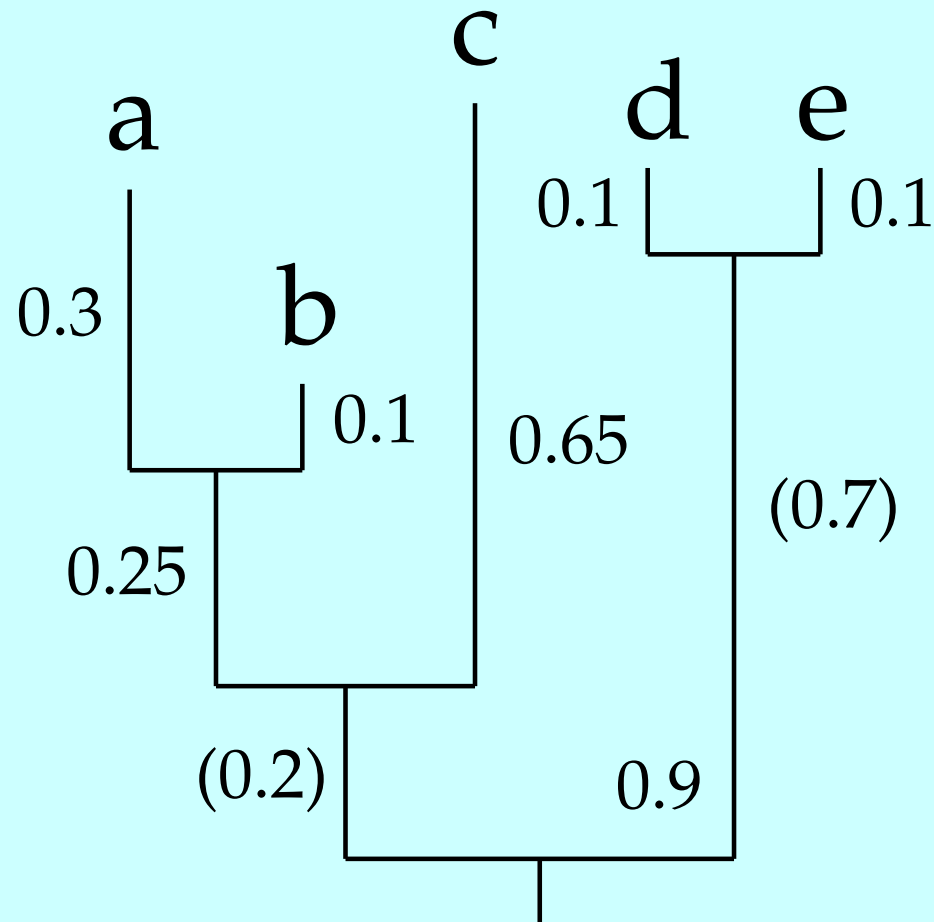
## Two uncorrelated characters evolving on that tree



## Identifying the two clades



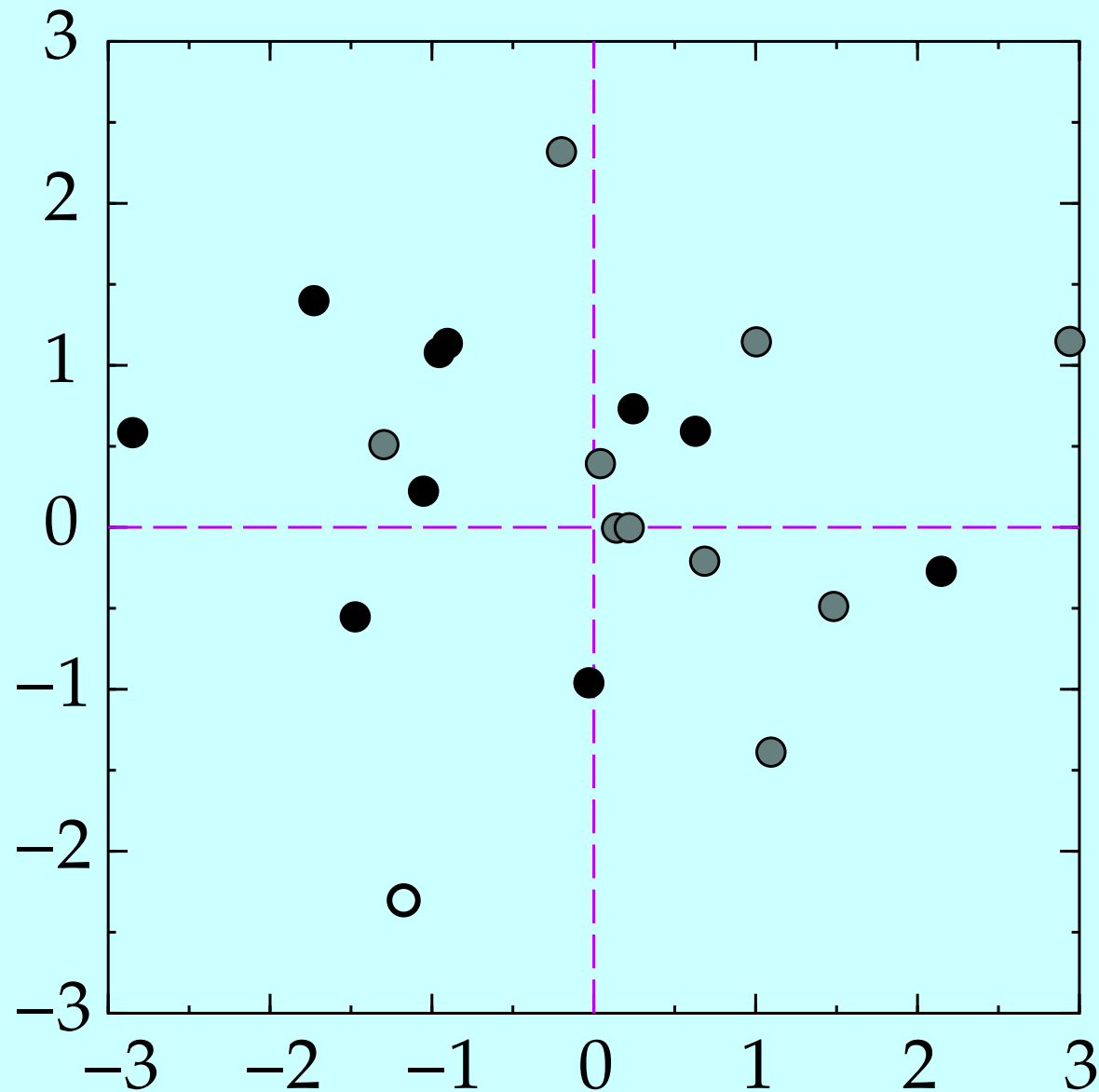
## A tree on which we are to observe two characters



## Contrasts on that tree

Contrast							Variance proportional to
$y_1$	=	$x_a$	—	$x_b$			0.4
$y_2$	=	$\frac{1}{4} x_a$	+	$\frac{3}{4} x_b$	—	$x_c$	0.975
$y_3$	=					$x_d$ — $x_e$	0.2
$y_4$	=	$\frac{1}{6} x_a$	+	$\frac{1}{2} x_b$	+	$\frac{1}{3} x_c$ — $\frac{1}{2} x_d$ — $\frac{1}{2} x_e$	1.11666

## Contrasts for the 20-species two-clade example



## The algebra

If  $\mathbf{T}$  is the covariances of  $n$  tips on the tree, and  $\mathbf{V}$  is the (unknown) covariances of the Brownian motion of the  $p$  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  $(n-1) \times n$  set of contrasts, each orthogonal to the grand mean, such that  $\mathbf{C}\mathbf{T}\mathbf{C}^t$  is an  $n-1$ -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

$$\ln L = K - ((n-1)/2) \ln |\mathbf{V}| - (1/2) \text{tr} (\mathbf{S}\mathbf{V})^{-1})$$

where

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

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

$$\hat{\mathbf{V}} = \mathbf{S}/(n-1)$$

which leads to

$$\ln L = K' - ((n-1)/2) \ln |\hat{\mathbf{V}}|$$

## An example: Smith and Cheverud 2002

Smith, R. J. and J. M. Cheverud. 2002. Scaling of sexual dimorphism in body mass: A phylogenetic analysis of Rensch's Rule in primates. *International Journal of Primatology* **23**(5): 1095-1135.

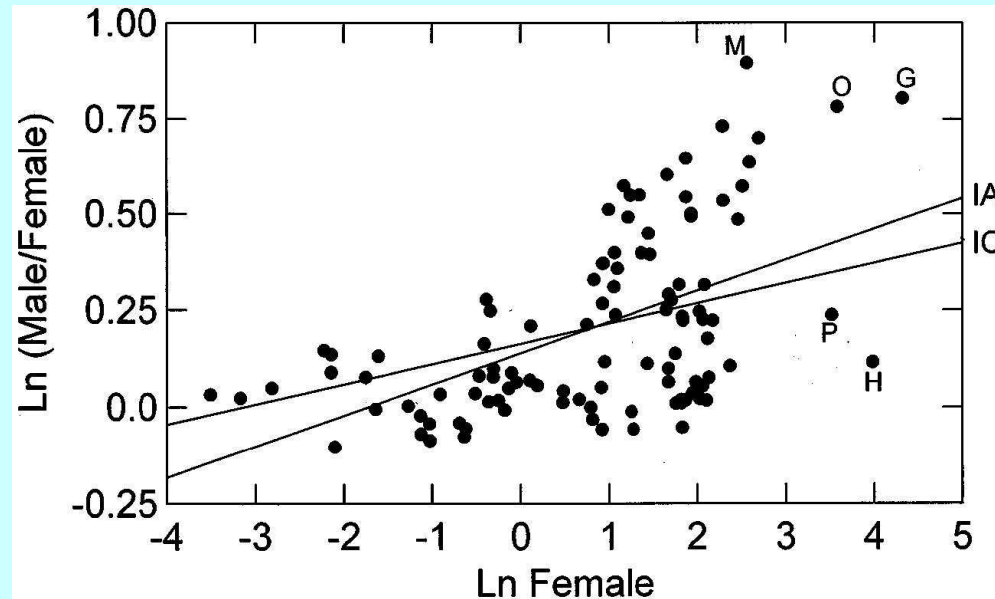


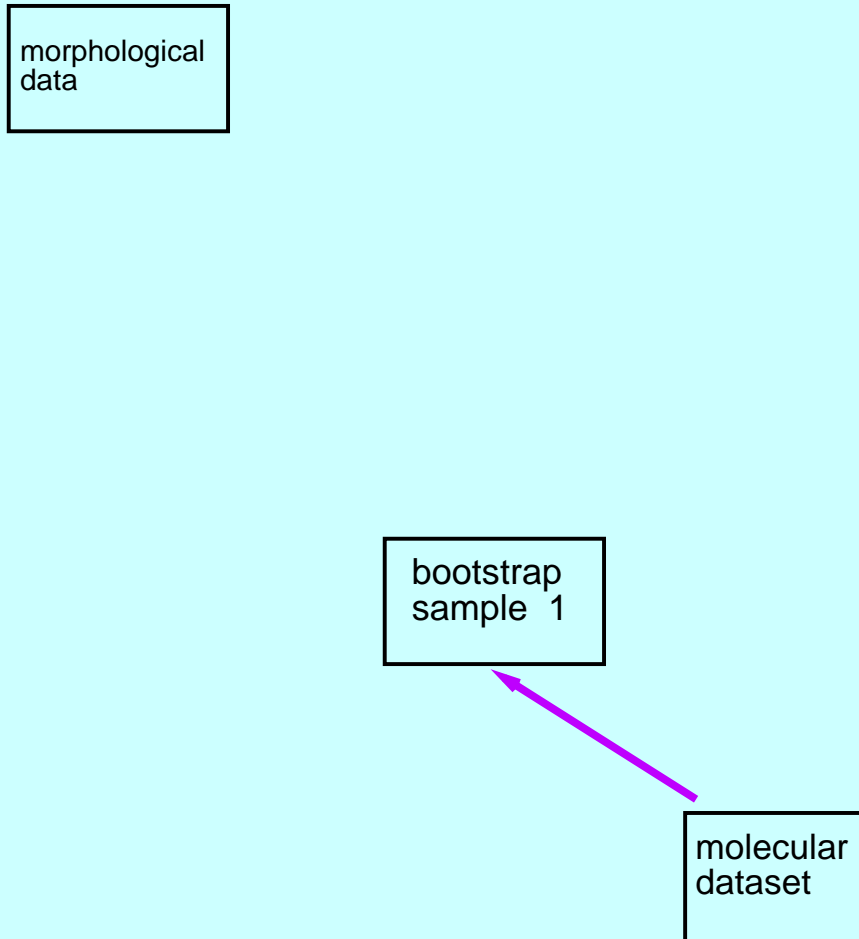
Fig. 1. The interspecific allometric equation (specific regression, identified as IA) and the independent contrasts equation (identified as IC) plotted for 105 primate species in raw data space, transformed to natural logarithms. The interspecific allometric equation is  $\ln y = 0.139 + 0.080(\ln x)$ , with  $r = 0.53$ . The phylogenetically corrected form of this equation, taken from the independent contrasts analysis, is  $\ln y = 0.160 + 0.056(\ln x)$ , with  $r = 0.26$ . The two equations are not significantly different from each other. The identified species are *Mandrillus sphinx* (M), *Pongo pygmaeus* (O), *Gorilla gorilla* (G), *Pan troglodytes* (P), and *Homo sapiens* (H).

# When the tree is noisy: Propagating bootstrap sampling

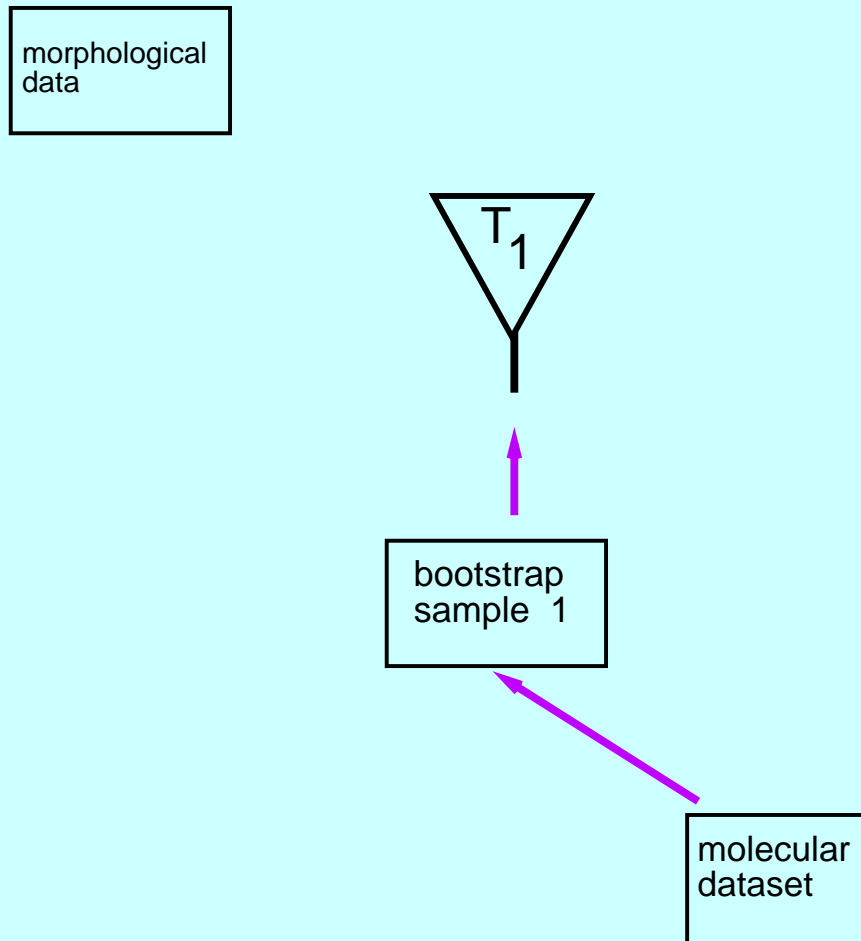
morphological  
data

molecular  
dataset

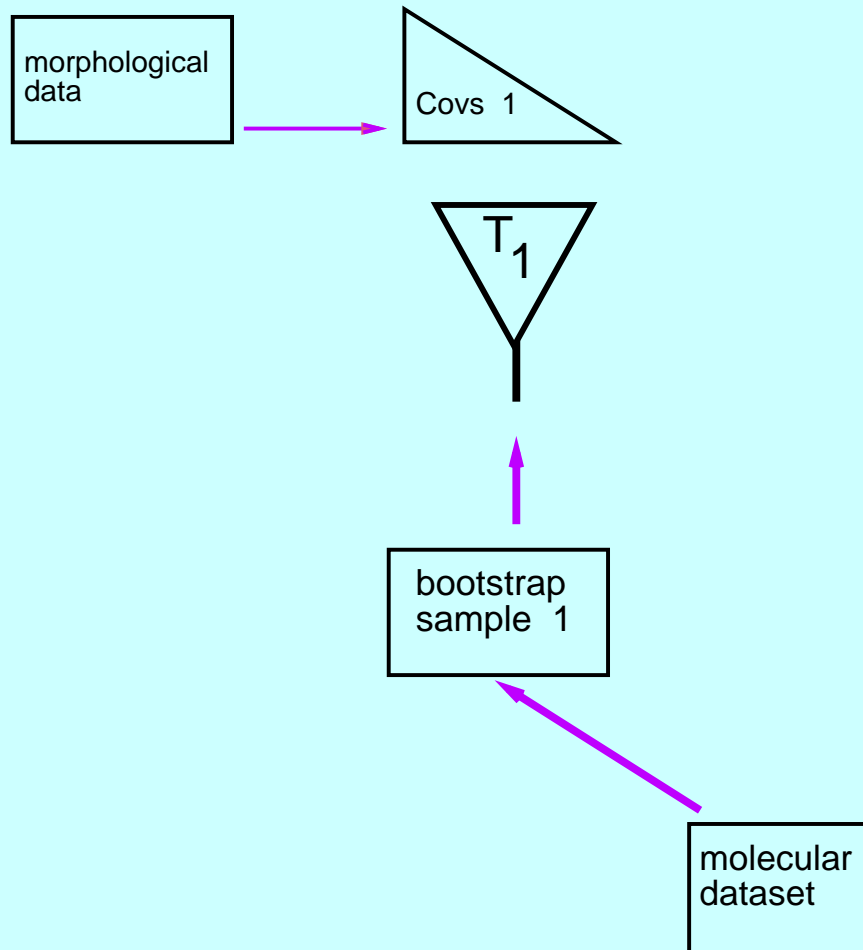
# Propagating bootstrap sampling



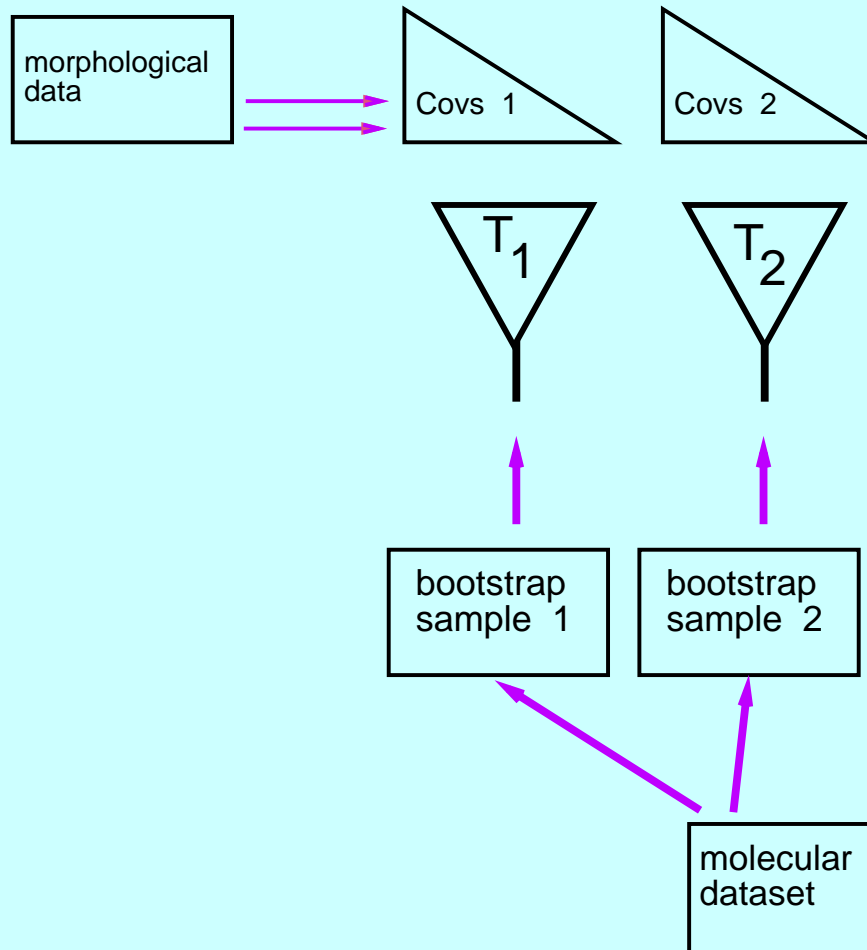
# Propagating bootstrap sampling



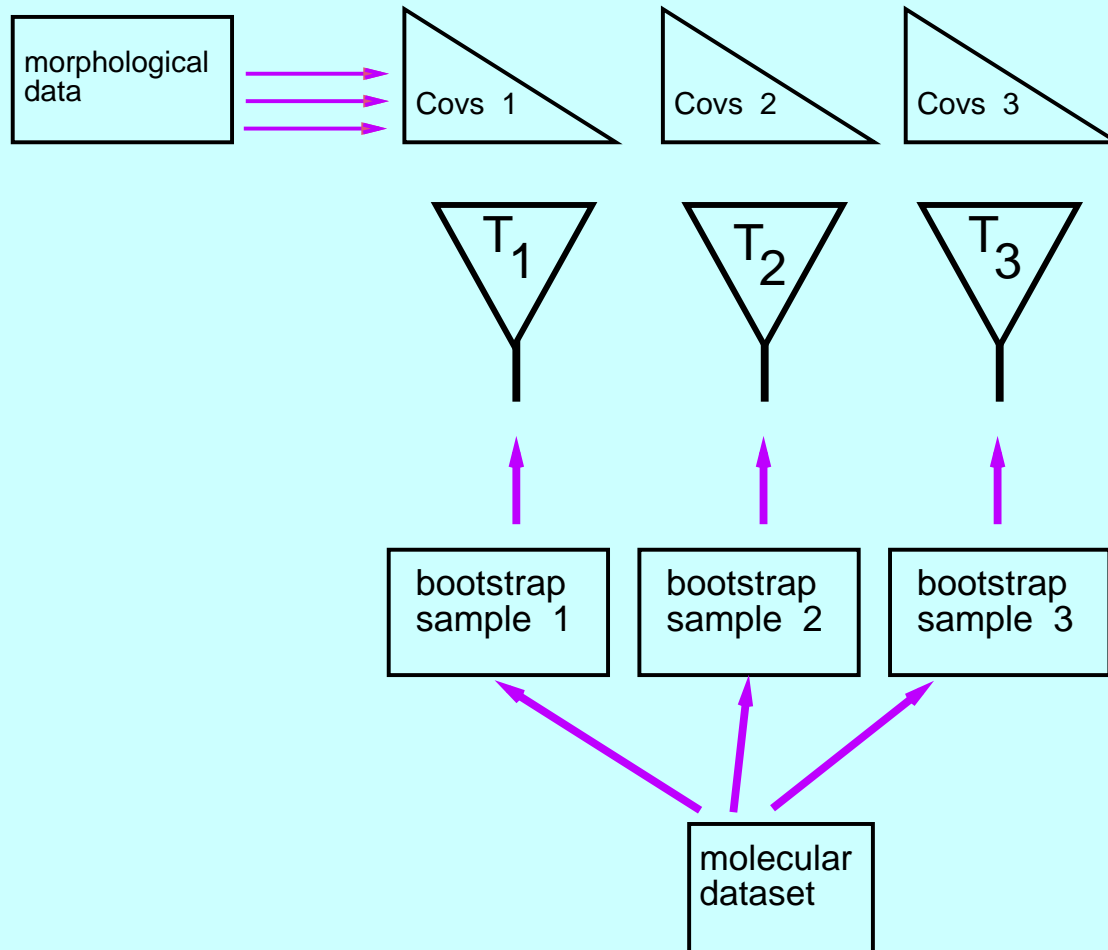
# Propagating bootstrap sampling



# Propagating bootstrap sampling

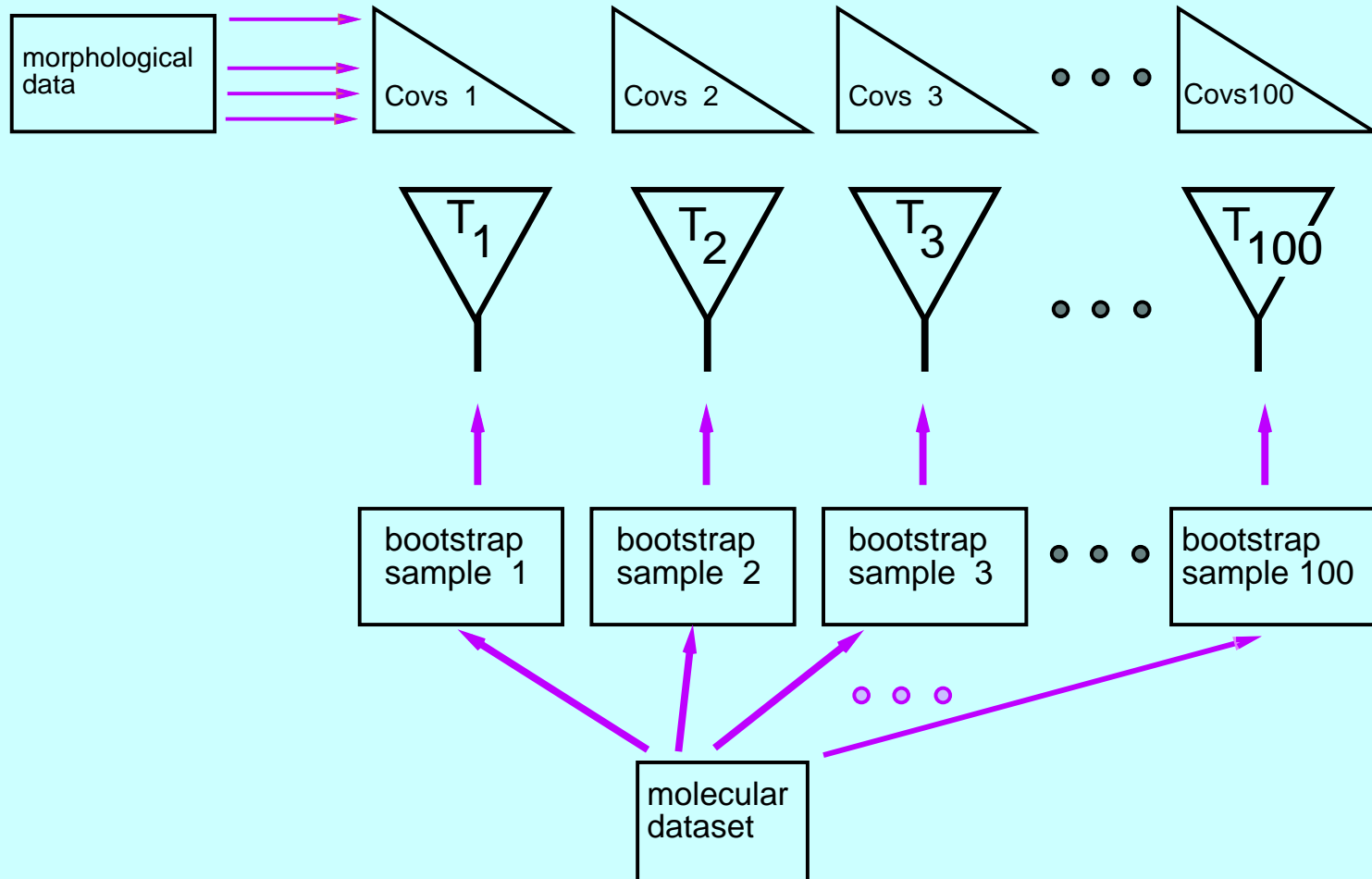


# Propagating bootstrap sampling

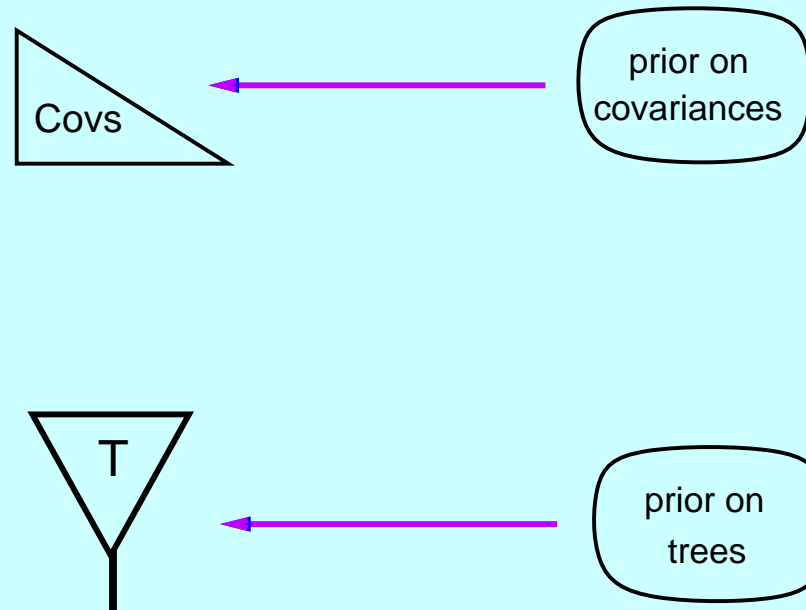




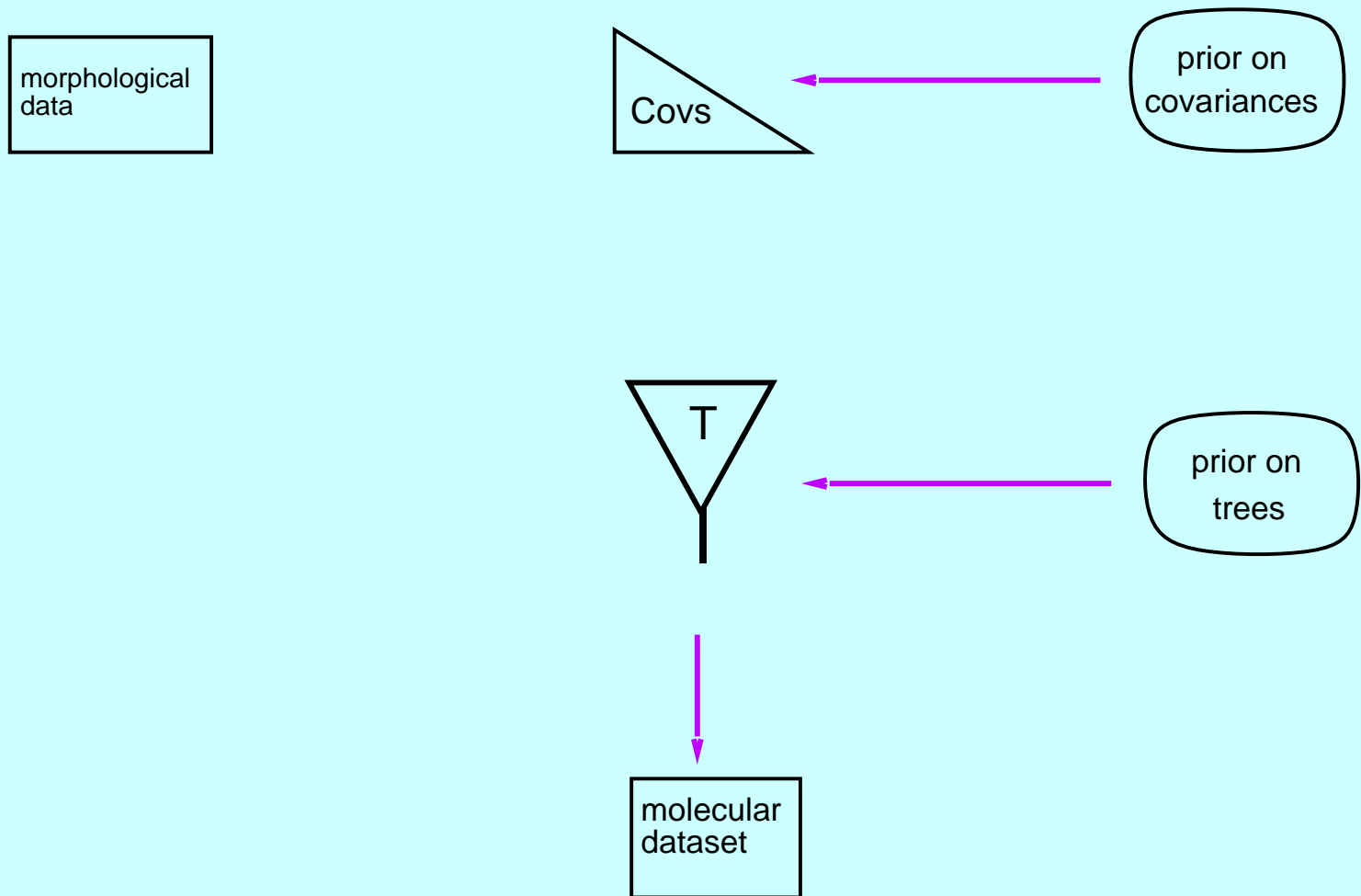
# Propagating bootstrap sampling



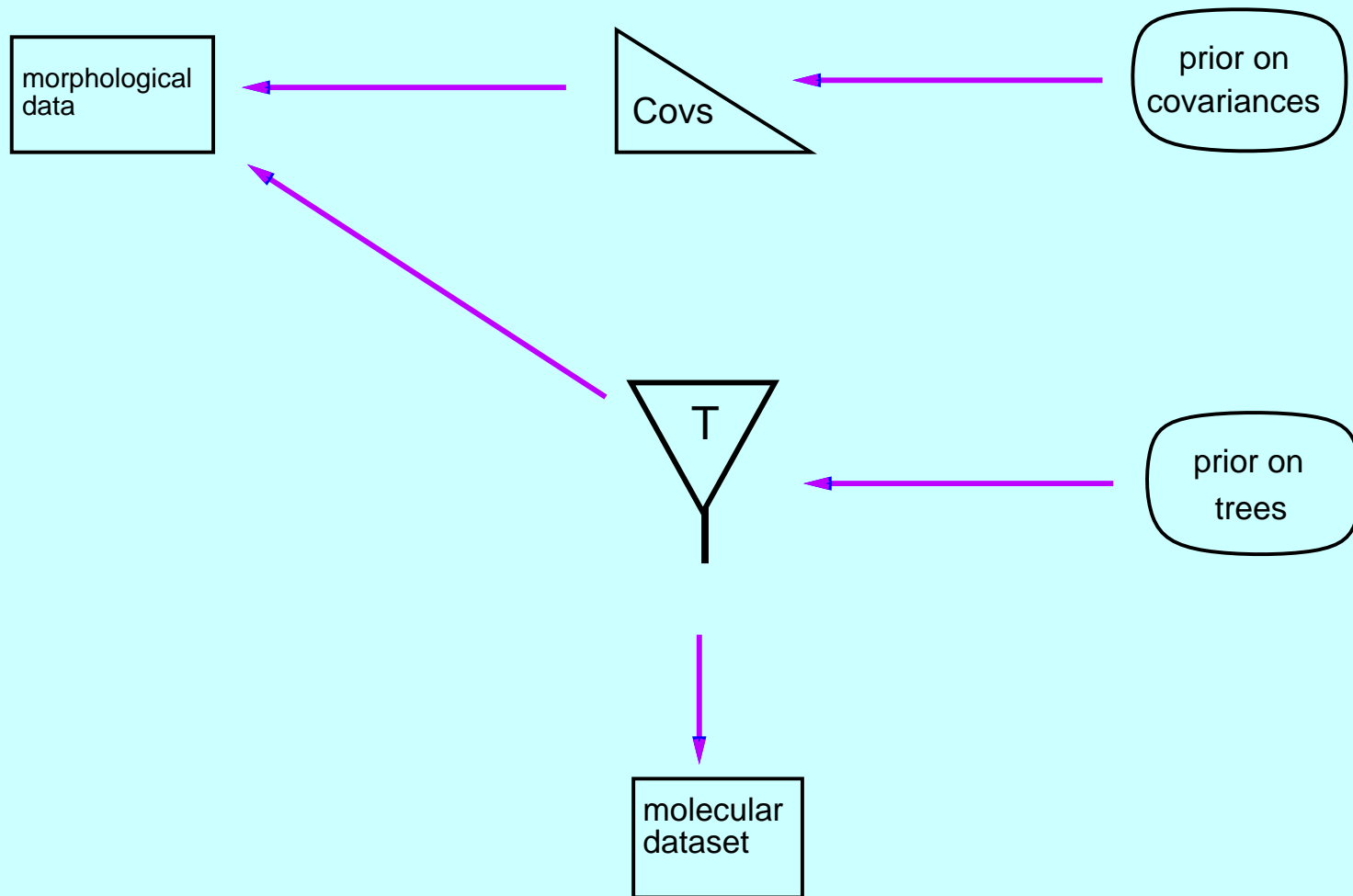
# A Bayesian model



# A Bayesian model

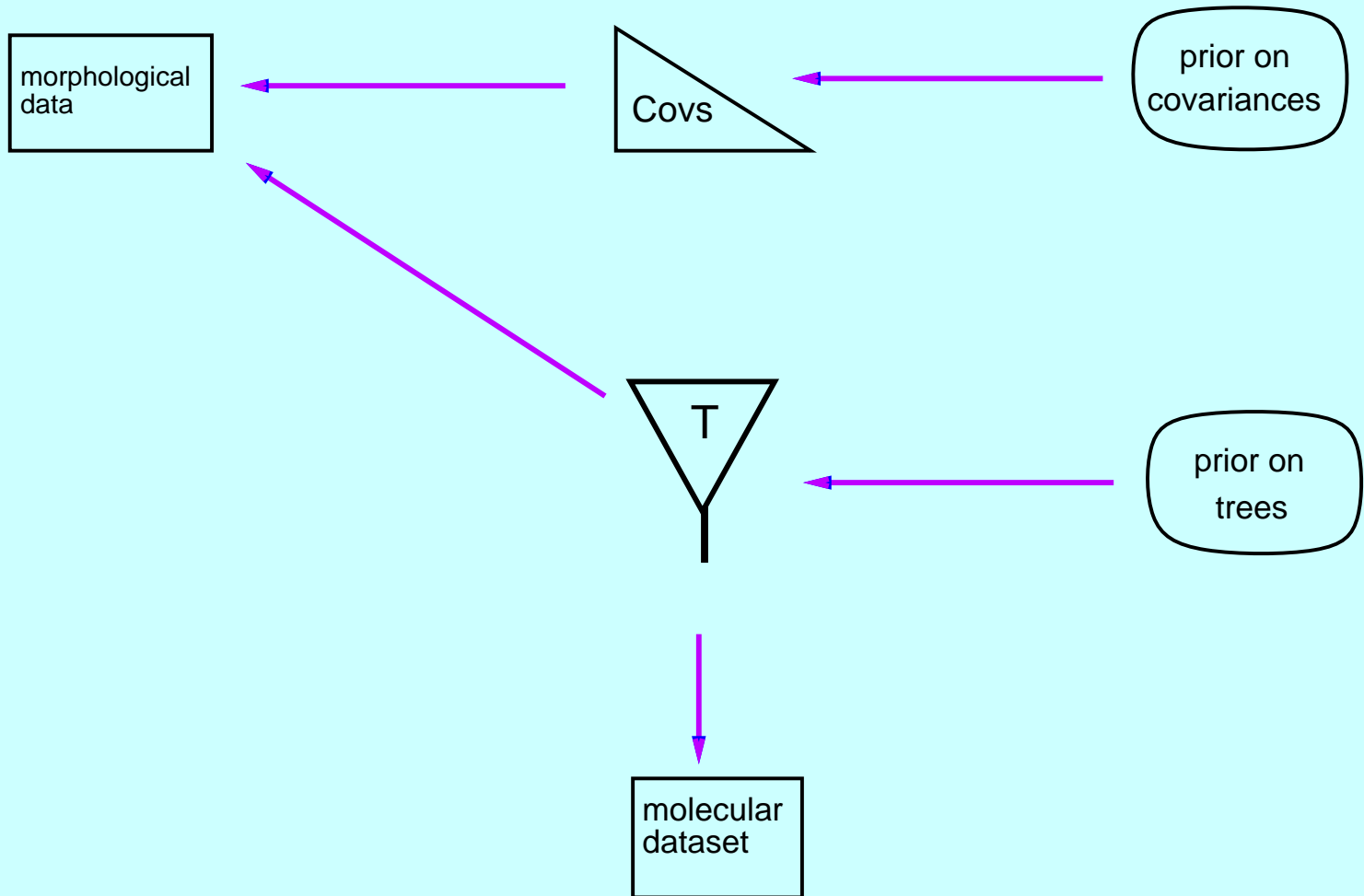


# A Bayesian model

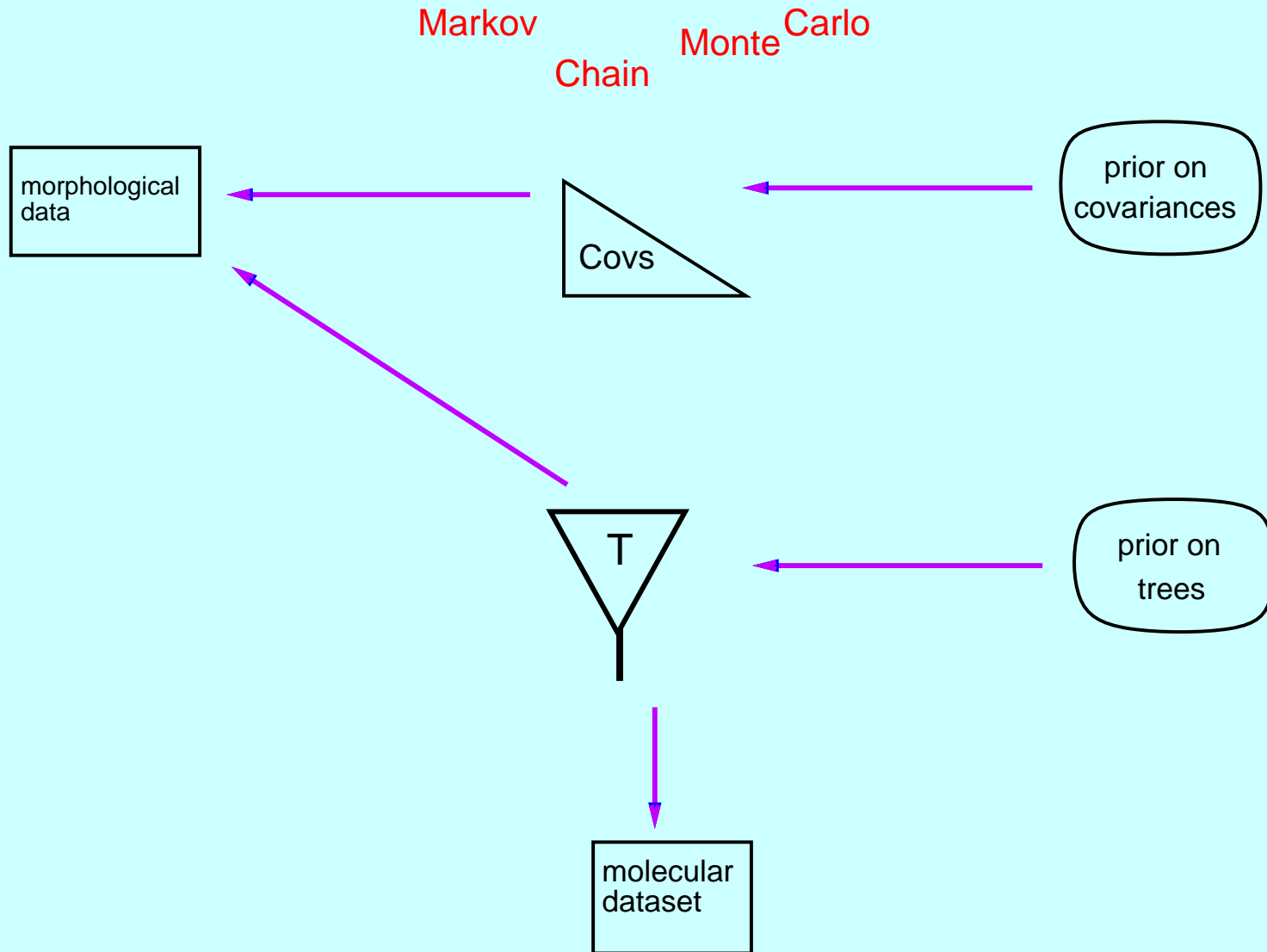


# Bayesian MCMC

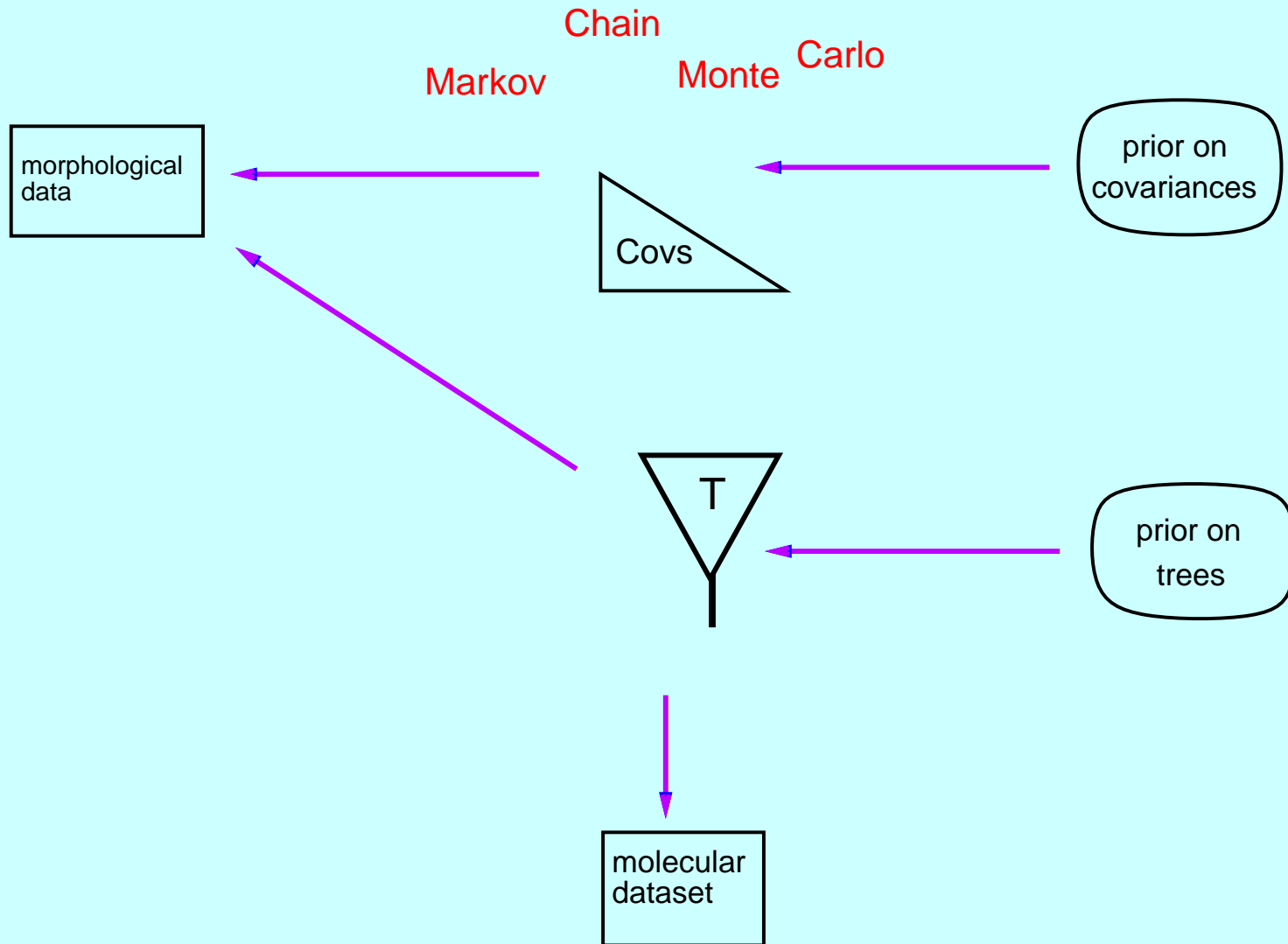
Markov Chain Monte Carlo



# Bayesian MCMC

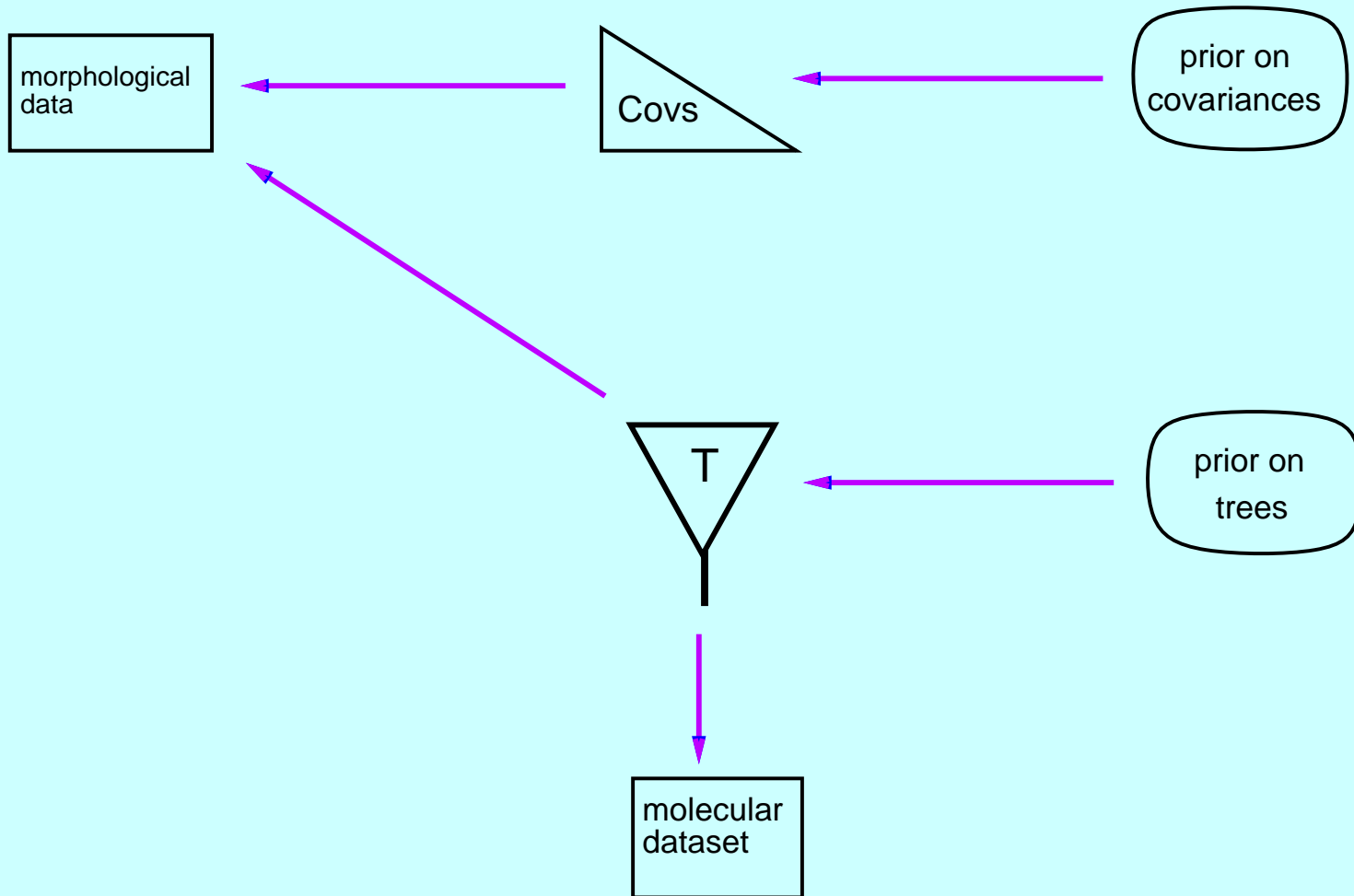


# Bayesian MCMC



# Bayesian MCMC

Markov Chain Monte Carlo





## Some complications

- (As noted above) dealing with uncertainty about the phylogeny

## Some complications

- (As noted above) dealing with uncertainty about the phylogeny
- Small sample size from species means their species means are uncertain. Must use a model with another level of variation – within-species phenotypic variation (Ricklefs and Starck, 1996; Ives et al., 2007; Felsenstein, 2008)

## Some complications

- (As noted above) dealing with uncertainty about the phylogeny
- Small sample size from species means their species means are uncertain. Must use a model with another level of variation – within-species phenotypic variation (Ricklefs and Starck, 1996; Ives et al., 2007; Felsenstein, 2008)
- Rate of change of morphological characters need not be constant on the molecular tree branch lengths.

## Some complications

- (As noted above) dealing with uncertainty about the phylogeny
- Small sample size from species means their species means are uncertain. Must use a model with another level of variation – within-species phenotypic variation (Ricklefs and Starck, 1996; Ives et al., 2007; Felsenstein, 2008)
- Rate of change of morphological characters need not be constant on the molecular tree branch lengths.
- Note – regressions involving contrasts should assume that they all have expectation zero.

## Some complications

- (As noted above) dealing with uncertainty about the phylogeny
- Small sample size from species means their species means are uncertain. Must use a model with another level of variation – within-species phenotypic variation (Ricklefs and Starck, 1996; Ives et al., 2007; Felsenstein, 2008)
- Rate of change of morphological characters need not be constant on the molecular tree branch lengths.
- Note – regressions involving contrasts should assume that they all have expectation zero.
- How to infer the effect of an environmental variable when only its present-day values are known but not its values when the past changes were occurring?

## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled

## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled
- ... particularly if they samples are not independent but on a tree

## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled
- ... particularly if they samples are not independent but on a tree
- ... particularly if the quantitative characters are thresholded



## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled
- ... particularly if they samples are not independent but on a tree
- ... particularly if the quantitative characters are thresholded
- How do we propagate the resulting uncertainty when biologists want “fly on the wall” certainty?

## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled
- ... particularly if they samples are not independent but on a tree
- ... particularly if the quantitative characters are thresholded
- How do we propagate the resulting uncertainty when biologists want “fly on the wall” certainty?
- Expanding to more species may put the model at risk

## Poor inference of covariation – what to do with that?

- Covariances are hard to infer with only (say) 50 species sampled
- ... particularly if they samples are not independent but on a tree
- ... particularly if the quantitative characters are thresholded
- How do we propagate the resulting uncertainty when biologists want “fly on the wall” certainty?
- Expanding to more species may put the model at risk
- Expanding to more characters just adds new parameters to estimate

# References for phylogenetic comparative methods

Felsenstein, J. 1985a. Phylogenies and the comparative method. *American Naturalist* **125**: 1–5. [Introduces the contrasts method]

Felsenstein, J. 1988. Phylogenies and quantitative characters. *Annual Review of Ecology and Systematics* [Suggests using bootstrapping to correct comparative methods for uncertainty about the phylogeny] **19**: 445–471.

Harvey, P. H. and M. D. Pagel. 1991. *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford. [The major book introducing statistical phylogenetic comparative methods]

Grafen, A. 1989. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London, Series B* **326**: 119–157. [Using generalized least squares to evaluate the likelihood for Brownian Motion phylogenies and do comparative methods analysis, without the contrasts methods. Discusses ways of coping with unresolved parts of the phylogeny and with varying evolutionary rates.]

## References, continued

Ricklefs, R. E. and J. M. Starck. 1996. Applications of phylogenetically independent contrasts: A mixed progress report. *Oikos* 77: 167–172.

[Pointing put that small sample size within species is a problem for comparative methods]

Ives, A. R., P. E. Midford, and T. Garland. 2007. Within-species variation and measurement error in phylogenetic comparative methods. *Systematic Biology* 56: 252-270. [Taking small sample size into account when we know the within-species phenotypic covariances]

Felsenstein, J. 2008 Comparative methods with sampling error and within-species variation: contrasts revisited and revised. *American Naturalist* 171: 713–725. [Inferring both between=species evolutionary covariances and within-species phenotypic variation]

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