Models of Continuous Trait Evolution

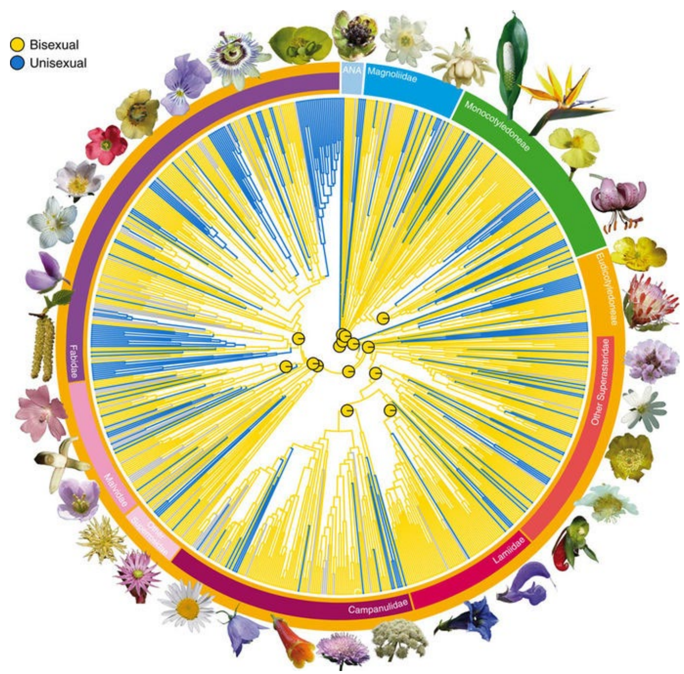
Mapping Traits to Phylogeny

Ancestral State Estimation

-Map traits to phylogeny

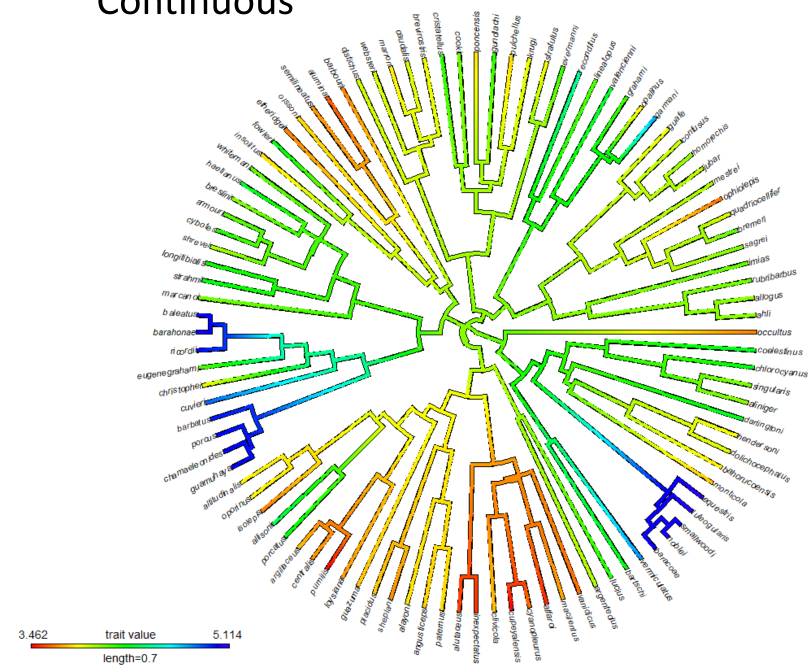
-Estimate trait values at ancestral nodes (and branches)

Discrete



Sauquet et al. *Nature Comm.* (20017)

Continuous



Data from Mahler et al. *Evol.* (2010)

-Accomplished using a *model* of evolutionary change

-How do we define the 'fit' of the data under that model?

Continuous Data: Maximum Likelihood

-One approach uses maximum likelihood

-Using statistical theory, one can ask:

What is the probability of observing my data, given the phylogeny and some evolutionary model?

$$\Pr(\boldsymbol{X} \mid \tau, \theta)$$

Same as: "What is the likelihood of some evolutionary model as observed by *conditioning* the data on the phylogeny under that model?

$$\mathcal{L}(\theta) = \Pr(\boldsymbol{X} \mid \tau, \theta)$$

Use search algorithm to maximize $\mathcal{L}(\theta)$

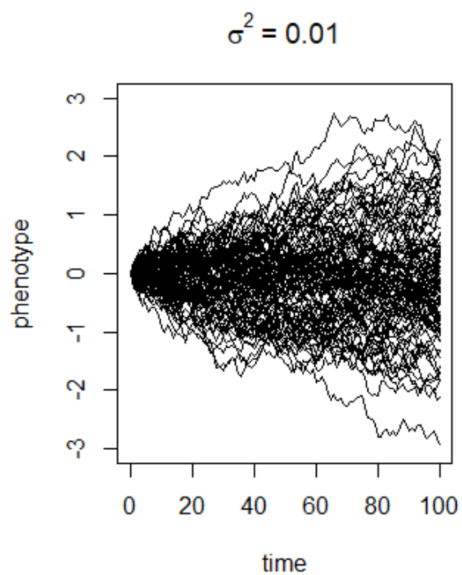
Continuous Data: Brownian Motion

Common *null model* of evolutionary change: Brownian motion

BM embodies the Markov process such that:

Trait changes are independent from time step to time step

Outcome: no change in μ, but $\sigma^2_y \uparrow \propto$ time



Side-note: this is the continuous-trait model equivalent of the Markov process we discussed earlier

Given this *model*, one can calculate the probability of observing the trait data on the phylogeny (or equivalently, the likelihood of the model given the data conditioned and the phylogeny)

Continuous Data: Brownian Motion

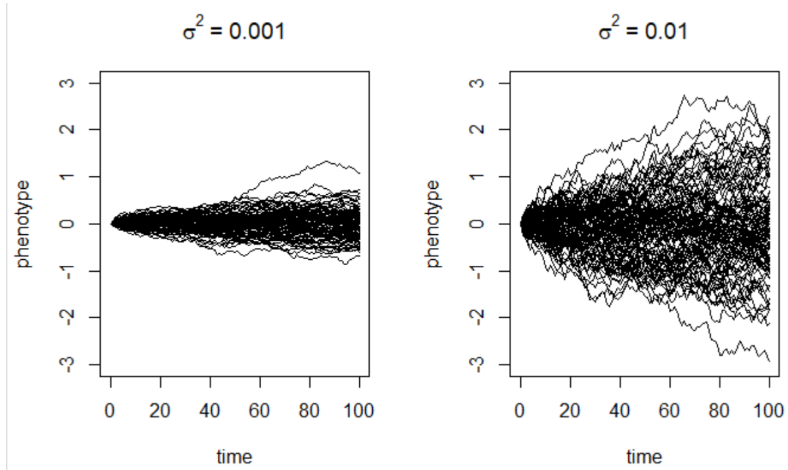
A *null model* of evolutionary change: Brownian motion

$$dY_i(t) = \sigma dB_i(t)$$

Character change — Evolutionary rate — Small random perturbations

$$\sigma^2 = \frac{(\mathbf{Y} - E(\mathbf{Y}))^{\mathbf{t}} \mathbf{C}^{-1} (\mathbf{Y} - E(\mathbf{Y}))}{N}$$

Evolutionary rate of change
(a phylogeny-standardized variance)



$$logL = \log\left[\frac{exp\left\{-\frac{1}{2}[(Y - E(Y))^t \mathbf{V}^{-1}(Y - E(Y))]\right\}}{\sqrt{2\pi^N \times \det(\mathbf{V})}}\right]$$

Note: this is the univariate logL

Brownian Motion: What's in a Likelihood?

-Components of the likelihood: 3 main parts

$$logL = \log\left[\frac{exp\left\{-\frac{1}{2}[(Y-E(Y))^t\mathbf{V}^{-1}(Y-E(Y))]\right\}}{\sqrt{2\pi^N \times \det(\mathbf{V})}}\right]$$

1: $2\pi^N$: A constant

2: $-\frac{1}{2}[(Y-E(Y))^t\mathbf{V}^{-1}(Y-E(Y))]$: Reduces to $\frac{N}{2}$

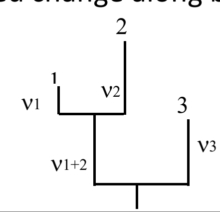
(formally $\frac{Np}{2}$ but for univariate, p=1. Thus for comparing models this is also a constant)

3: det(**V**): error covariance of the model*

<u>The likelihood is thus the residual error of the data under a model</u>

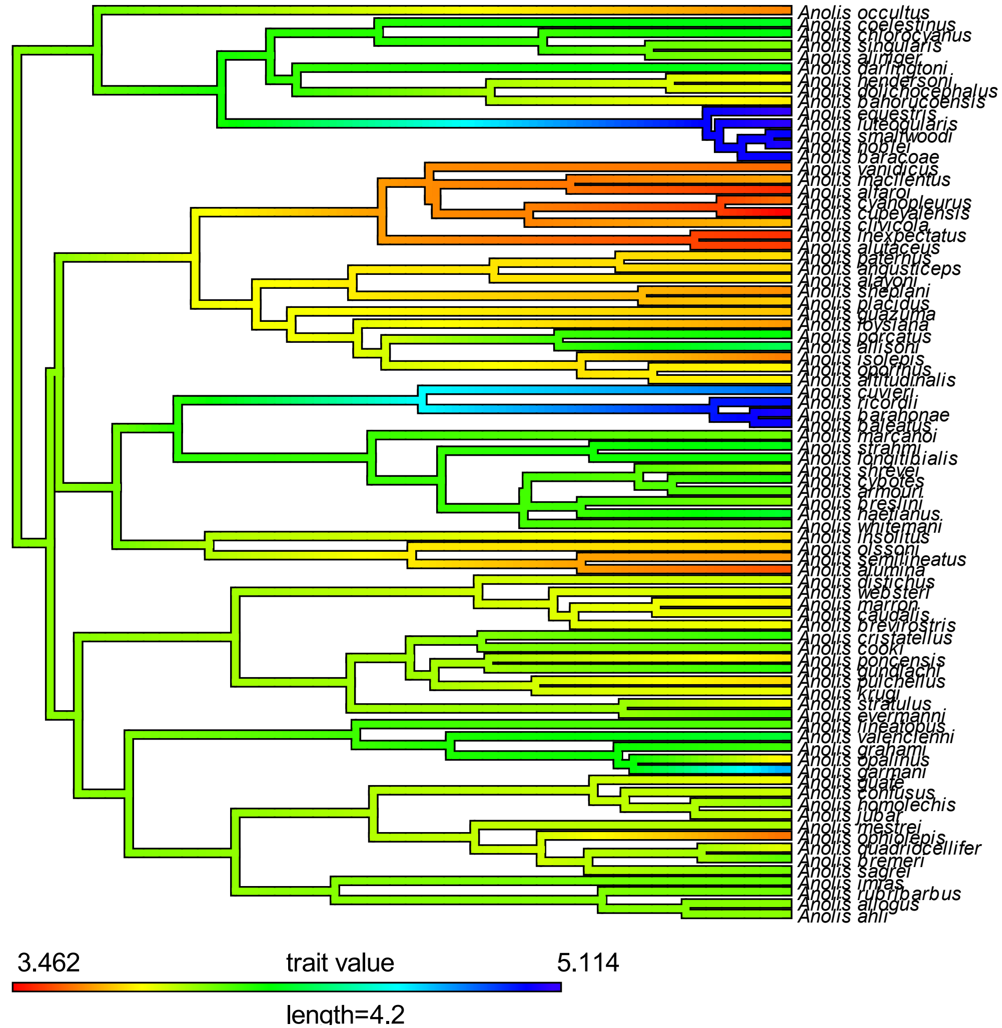
*Determinants of error covariance matrices are measures of the dispersion (generalized variance) of the data. A smaller det(V) means a better fit. HOWEVER: det(V) = 0 does not necessarily mean a 'perfect' fit. Often, there is a singularity issue in the modeling (see multivariate lecture).

Recall: **V**= σ^2**C** the evolutionary model (covariance matrix)
-expected change along branches of phylogeny



$$\mathbf{C} = \begin{pmatrix} v_1 + v_{1+2} & v_{1+2} & 0 \\ v_{1+2} & v_2 + v_{1+2} & 0 \\ 0 & 0 & v_3 \end{pmatrix}$$

Body size evolution in *Anolis* lizards



logL = 5.256010
AIC = -6.512

$\sigma^2 = 0.01823$
E(Y) = 4.0535 (root value)

Questions:

Is this a 'good' fit?

Is neutral evolution an appropriate model of trait change?

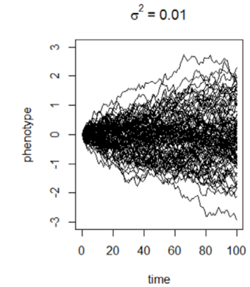
Data: Mahler et al. 2010

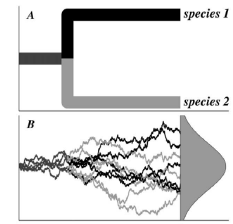
Ornstein-Uhlenbeck (OU) Models

Brownian motion: neutral change under drift (no selection)

$$dY_i(t) = \sigma dB_i(t)$$

Character change — Evolutionary rate — Small random perturbations



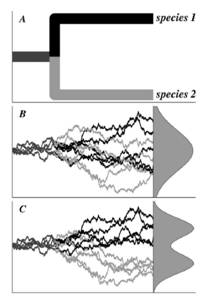


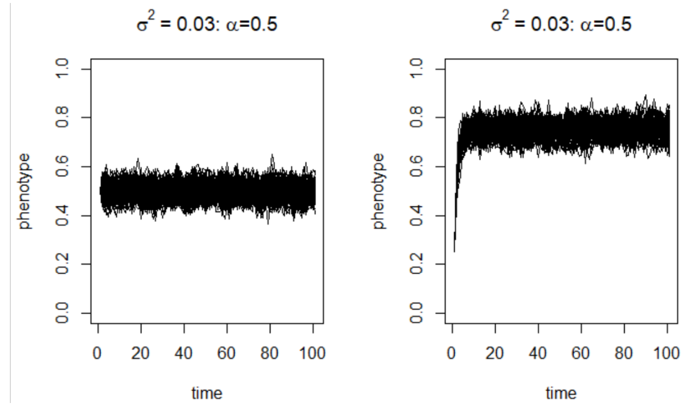
Ornstein-Uhlenbeck (OU): models both drift and selection

-Trait values ‘pulled’ towards optima: Θ (1 Θ :stabilizing; 2+ Θ diversifying selection)

$$dY_i = \alpha(\beta - Y(t)) + \sigma B(t)$$

Character change — Strength of selection — Distance from optimum — Brownian component





This model is fit using a different **V** in the logL!

Compare models: LRT and AIC

Images from Butler and King 2004. Am. Nat. 164: 683-695.

How does one compare different models?

Many approaches; two common ones are:

LRT (likelihood ratio tests)

Test measure that underlies much of parameteric statistical hypothesis testing

$$LRT = -2\log(\frac{L_F}{L_R})$$

LRT tested against X^2 with df = k_F-k_R (difference in model parameters)

AIC (Akaike information criterion)

A measure of model 'fit' relative to the number of parameters required

$$AIC = -2\log L + 2(k + 1)$$

ΔAIC (AIC_F – AIC_R) > 4.0 is strong support for full model

Model Parameters*

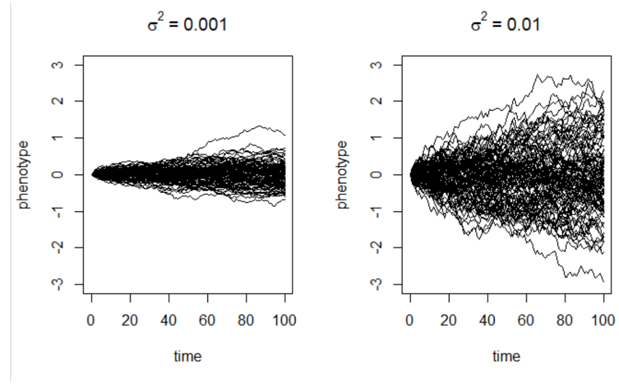
Free parameters differ across models

Brownian motion (BM1): neutral change under drift

-2 parameters: Phylogenetic mean (μ), and rate (σ^2_y)

$$dY_i(t) = \sigma dB_i(t)$$

Character change — Evolutionary rate — Small random perturbations



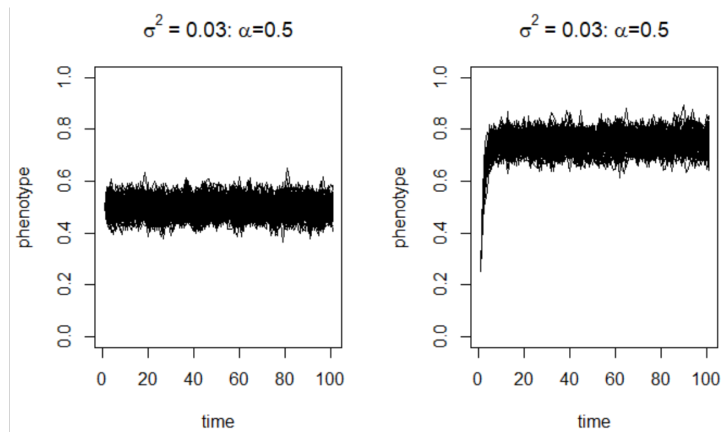
OU1: drift and selection

-3 parameters: Phylogenetic mean (μ), and rate (σ^2_y), selection (α)

$$dY_i = \alpha(\beta - Y(t)) + \sigma B(t)$$

Character change — Strength of selection — Distance from optimum — Brownian component

optima (θ) also specified, but 'linked' to a,
so not part of 'count' of parameters



OUM: multiple optima (multiple α & Θ)

*Univariate only: *many* more parameters for multivariate

Butler and King 2004. Am. Nat. 164: 683-695.

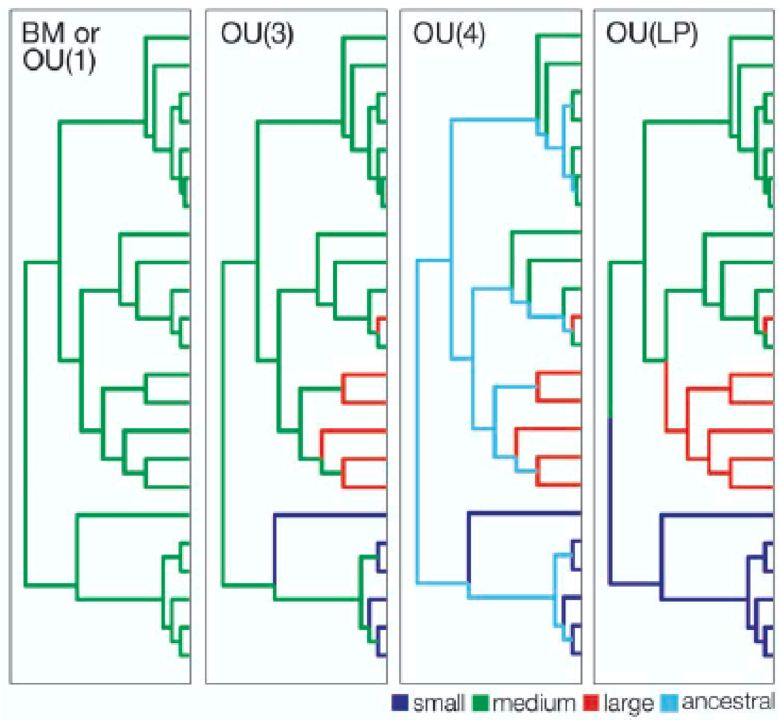
OUM: Multiple Adaptive Optima

What if there is more than one optimum?

OUM: multiple optima (multiple α & Θ)

Problem: how to model? Must define which taxa belong to each optimum

-We 'paint' groups on phylogeny based on biology for hypothesis



BM = OU1: Single group

OU3: optima based on size groups

OU4: size groups + ancestral group

OU(LP): size groups + 'priority' colonizing effects (who was on island first)

Butler and King. 2004. *Am. Nat.*

Comparing Models: Example

How did Anolis body size groups (small, medium, large) evolve?

-5 models: BM, OU_1, OU_3 OU_4 (3 group+anc), OU_{LP} (3 gp + history of colonization)

-OU_{LP} (3 gp + col. hist.) best explains body size evolution

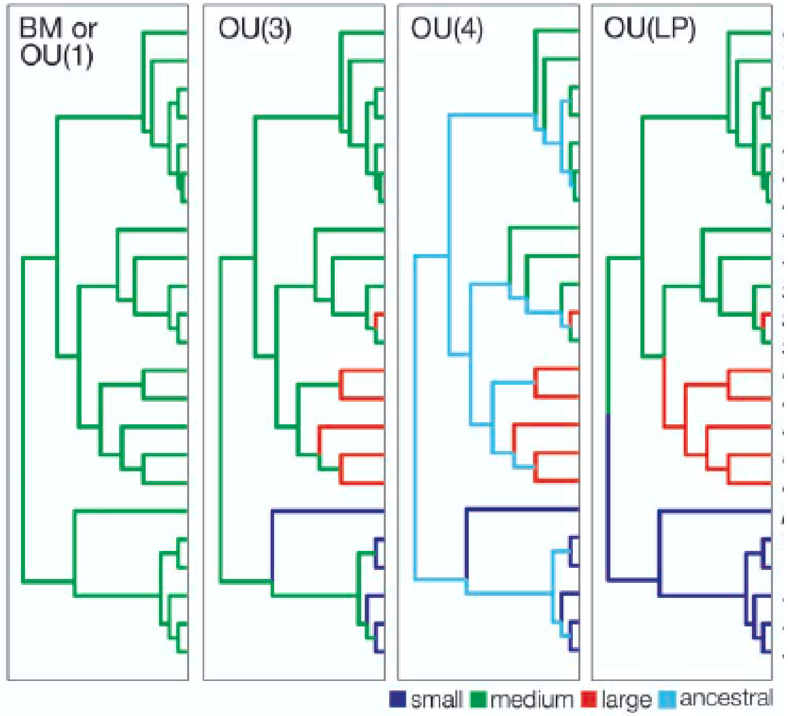


Table 1: Performance of alternative models for body size evolution in the character displacement study

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
$-2 \log \mathcal{L}$	−34.66	−34.66	−40.21	−47.22	−49.69
AIC	−30.66	−26.66	−28.21	−33.22	−37.69
SIC	−28.39	−22.12	−21.40	−25.27	-30.88
LR		0	5.55	12.56	15.03
P value		1	.24	.028	.0046

Butler and King. 2004. *Am. Nat.*

Comparing Models: Example 2

Body size evolution in *Anolis* lizards

BM1:

logL = 5.256010

AIC = -6.512

OU1:

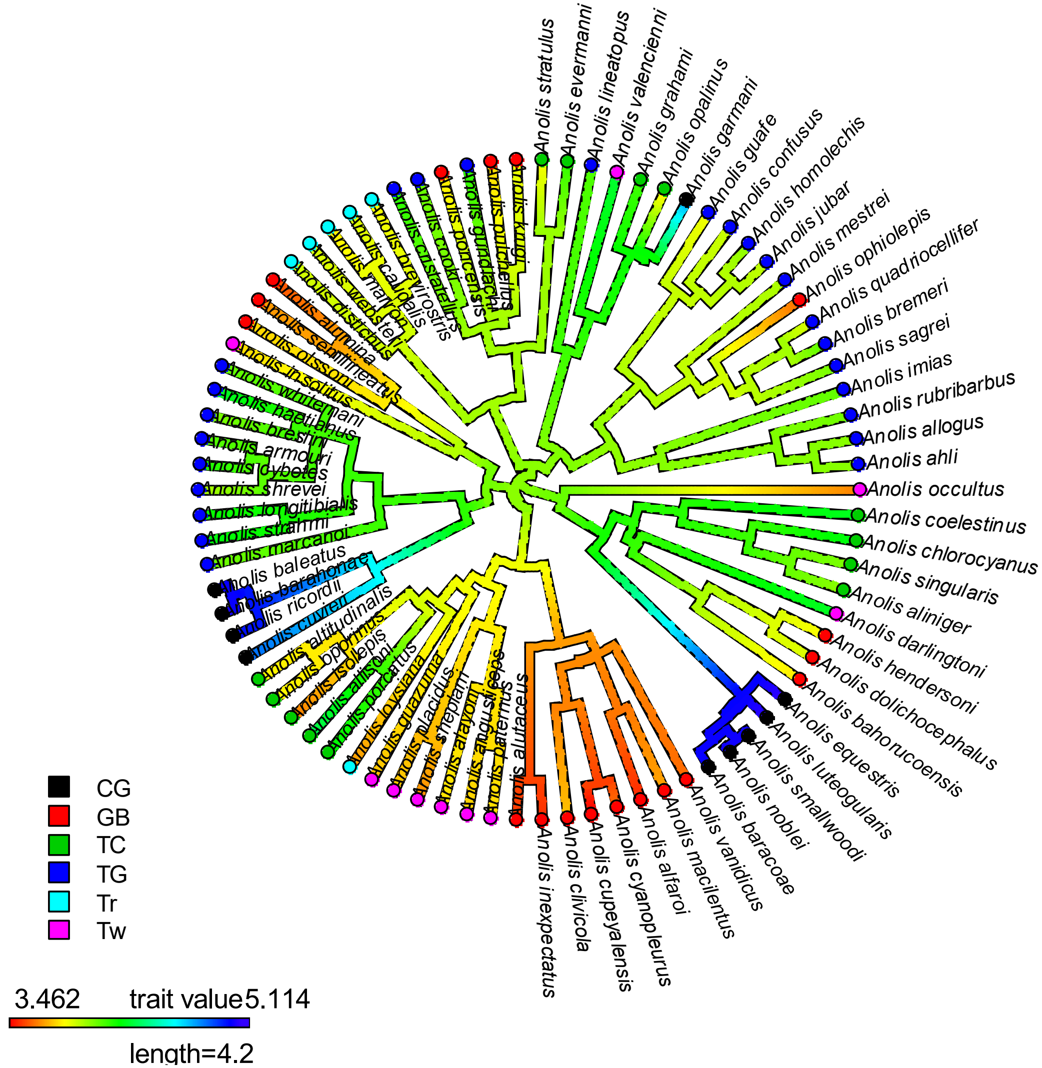
logL = 5.256010

AIC = -4.512

OUM:

logL = 39.4849

AIC = -62.969

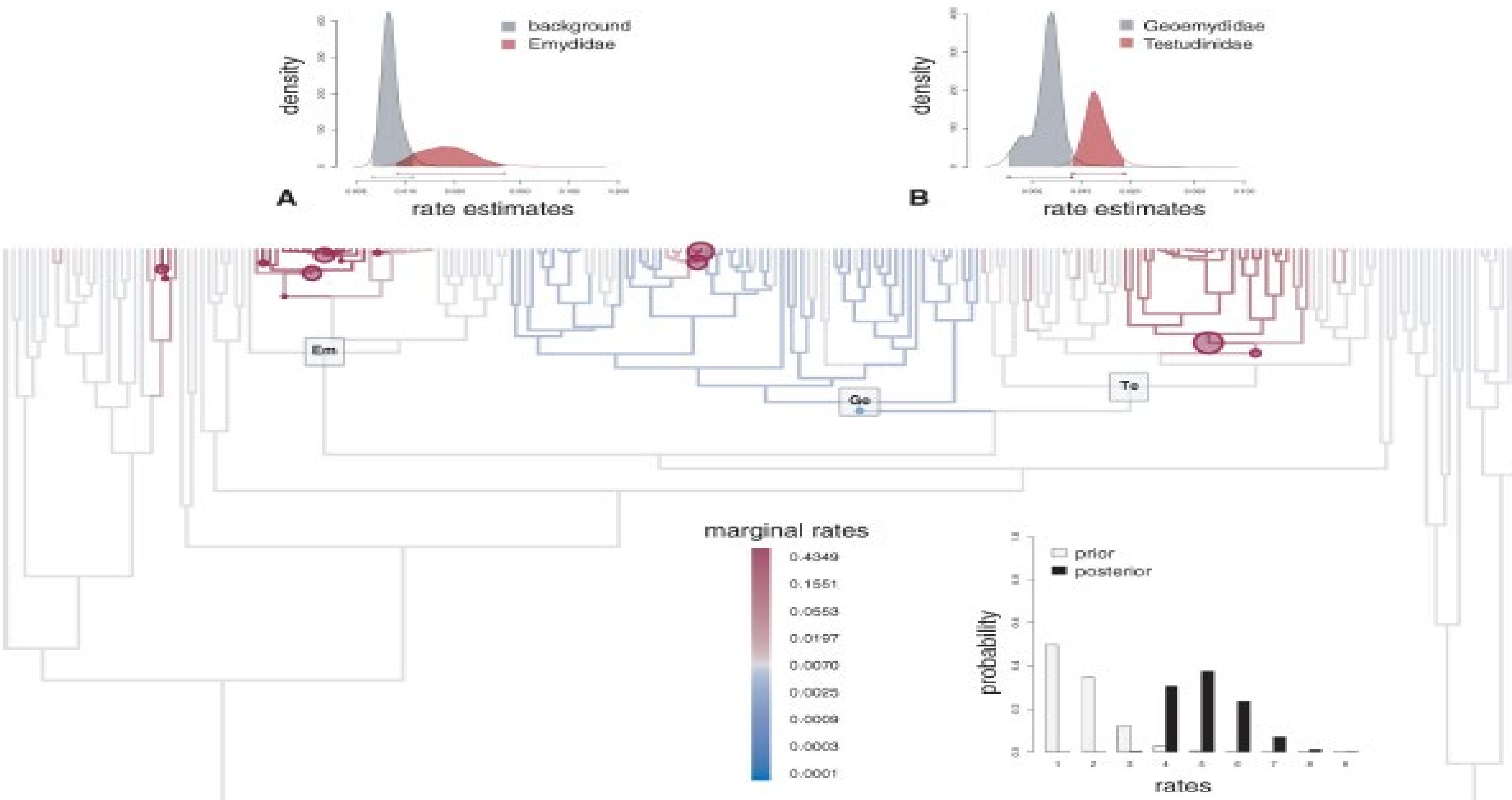


OUM strongly preferred

Data: Mahler et al. 2010

Multiple Rate Models

"How fast, as a matter of fact, do animals evolve in nature?" (Simpson, 1944)

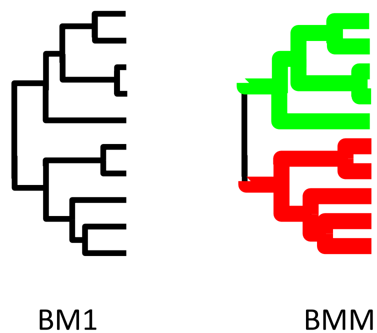


Eastman et al. 2011

Some scenarios with ‘groups’ are for rates, not optima

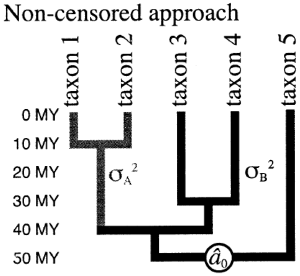
-“Does evolution occur faster on islands than on the mainland?”

Requires model with different σ^2 on different portions of the phylogeny



Procedurally, one ‘splits’ the phylogenetic covariance matrix **C** into components for each group, and multiplies by separate σ^2; then logL*

Non-censored approach



	taxon 1	taxon 2	taxon 3	taxon 4	taxon 5	E(**X**)
taxon 1	$40\sigma_A^2+10\sigma_B^2$	$30\sigma_A^2+10\sigma_B^2$	$10\sigma_B^2$	$10\sigma_B^2$	0	$\hat{a}_0$
taxon 2	$30\sigma_A^2+10\sigma_B^2$	$40\sigma_A^2+10\sigma_B^2$	$10\sigma_B^2$	$10\sigma_B^2$	0	$\hat{a}_0$
taxon 3	$10\sigma_B^2$	$10\sigma_B^2$	$50\sigma_B^2$	$20\sigma_B^2$	0	$\hat{a}_0$
taxon 4	$10\sigma_B^2$	$10\sigma_B^2$	$20\sigma_B^2$	$50\sigma_B^2$	0	$\hat{a}_0$
taxon 5	0	0	0	0	$50\sigma_B^2$	$\hat{a}_0$

*Again, a distinct **V** matrix in the logL

O’Meara et al. (2006)
Thomas et al. (2006)

Body size evolution in *Anolis* lizards

BM1:

logL = 5.256010

AIC = -6.512

$\sigma^2 = 0.01823$

BMM:

logL = 21.635

$P_{simul} = 0.001$

AIC = -29.307

$\sigma^2_{CG} = 0.0366$

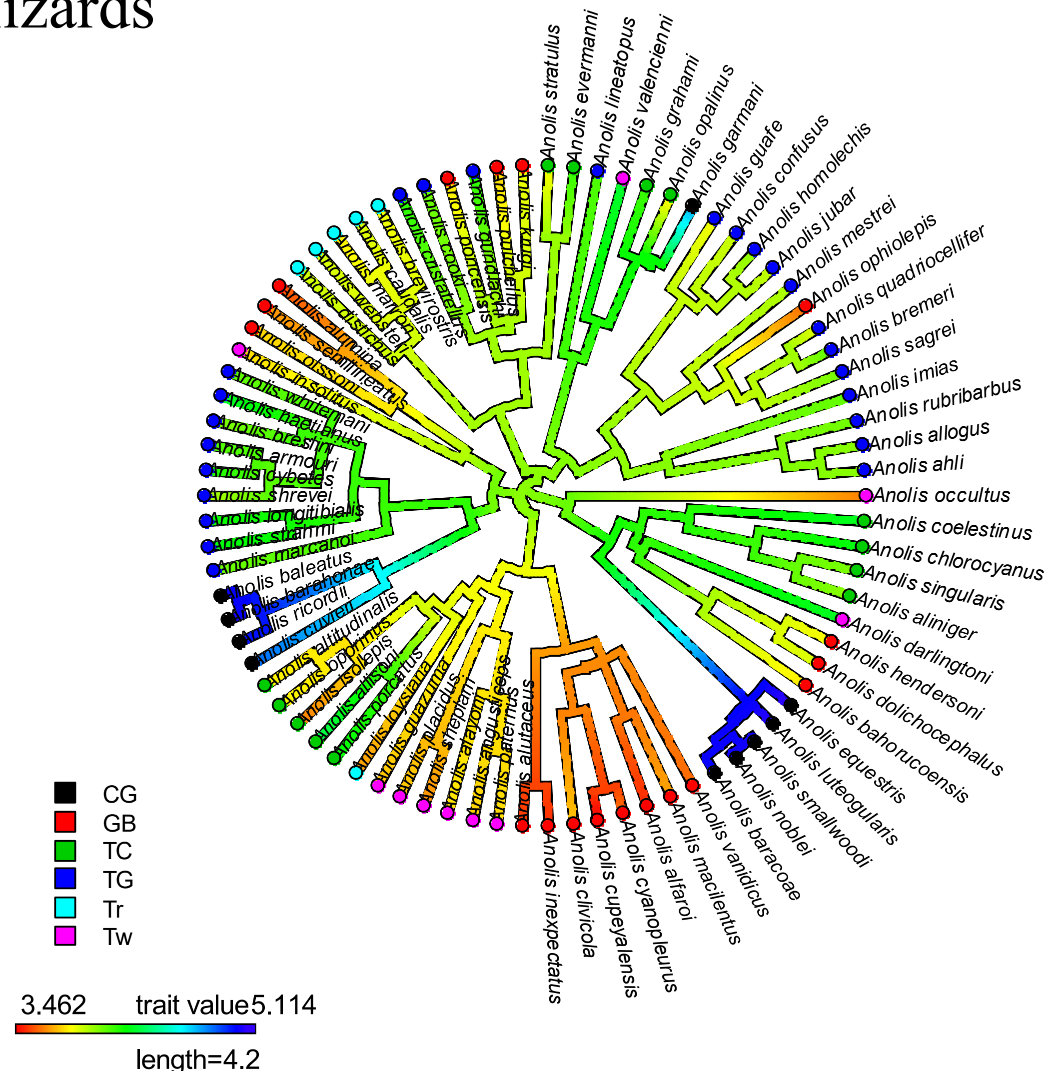
$\sigma^2_{GB} = 0.0259$

$\sigma^2_{TC} = 0.0242$

$\sigma^2_{TG} = 0.0058$

$\sigma^2_{TR} = 0.0014$

$\sigma^2_{TW} = 0.0021$



CG highest, TR,TW lowest

BMM strongly preferred

Data: Mahler et al. 2010

Extensions: Comparing Rates Among Traits

One can also compare evolutionary rates among traits

Does one trait evolve faster than another)?

Find *rate matrix* for set of traits:

$$\mathbf{R} = \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix} \qquad \mathbf{R} = \frac{\left(\mathbf{Y} - E\left(\mathbf{Y}\right)\right)^{\mathbf{t}} \mathbf{C}^{-1} \left(\mathbf{Y} - E\left(\mathbf{Y}\right)\right)}{N}$$

Obtain R_o and logL:

Estimate R_c & logL, where rates are constrained to be the same

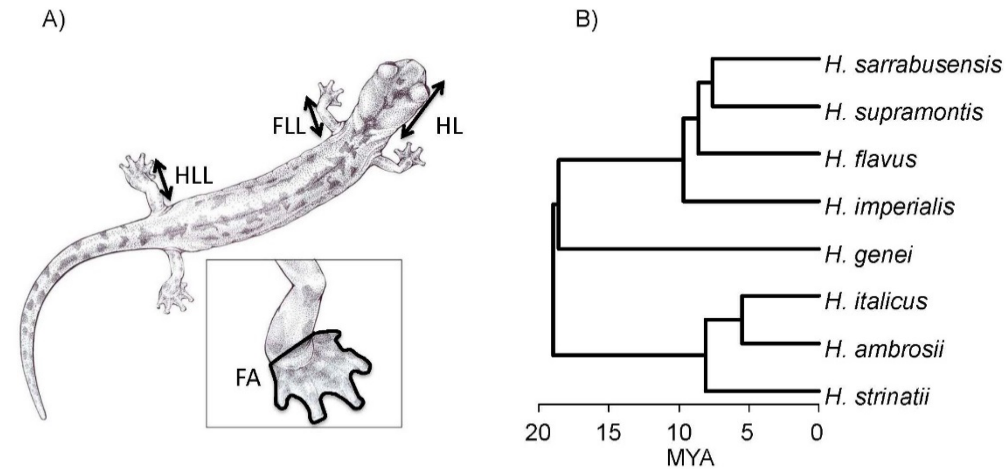
$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2 \qquad \mathbf{R}_C = \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{21} & \sigma_2^2 & \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{bmatrix}$$

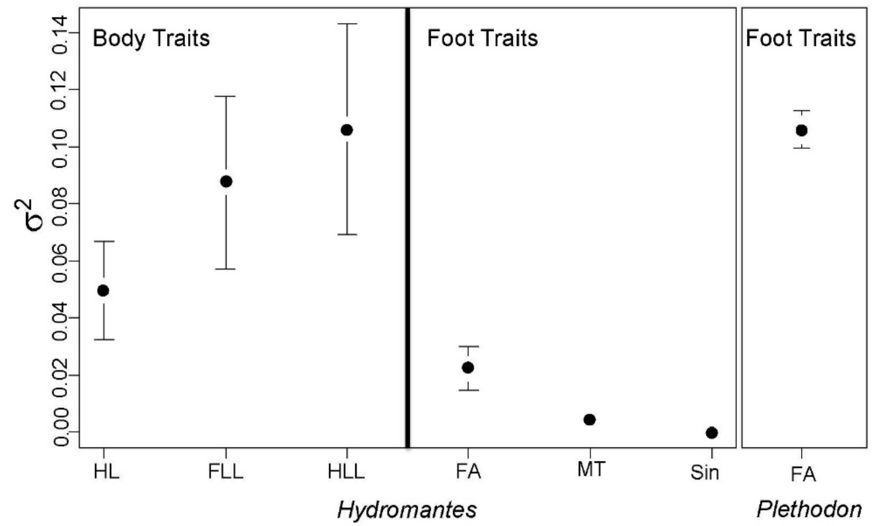
Compare the two models with LRT

Adams. (2013). Syst. Biol.

Extended to multivariate: Denton and Adams, 2015. Evol.

Compare morphological rates in cave-dwelling *Hydromantes*





Climbing traits evolve more slowly (consistent with evolutionary constraint)

Adams et al. 2017. *Am. Nat.*

Extensions: Comparing Rates Among Trees

One can also compare evolutionary rates for traits among trees

“Does body size evolve faster in clade X vs. clade Y?”

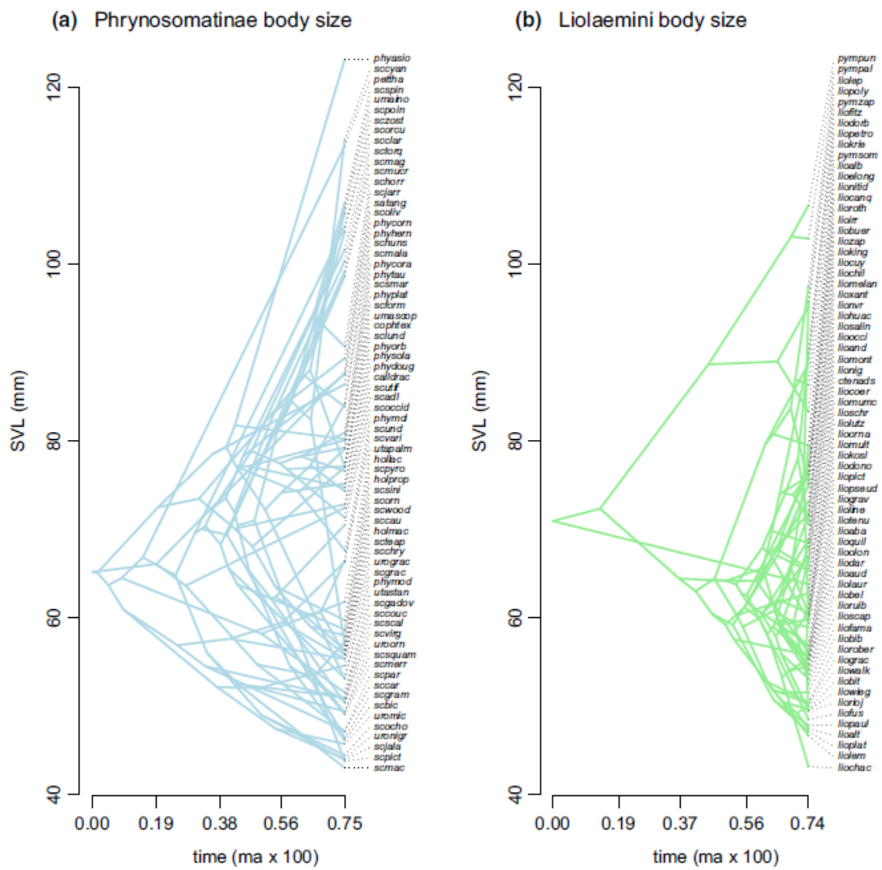


TABLE 2 Body size (SVL) and shape (common phylogenetic PC2) for two lizard clades: (1) the North American iguanian subfamily Phrynosomatinae; and (2) the South American lizard tribe Liolaemini

	σ_1^2	σ_2^2	a_1	a_2	k	$\log(L)$
Body size (SVL)						
ML common-rate model:						
Value	0.26	–	4.18	4.26	3	−4.85
SE	0.03	–	0.15	0.24		
ML multi-rate model:						
Value	0.19	0.33	4.18	4.26	4	−2.19
SE	0.03	0.06	0.13	0.27		
Likelihood ratio: 5.32; *p*-value (based on χ^2, $df = 1$): 0.021						

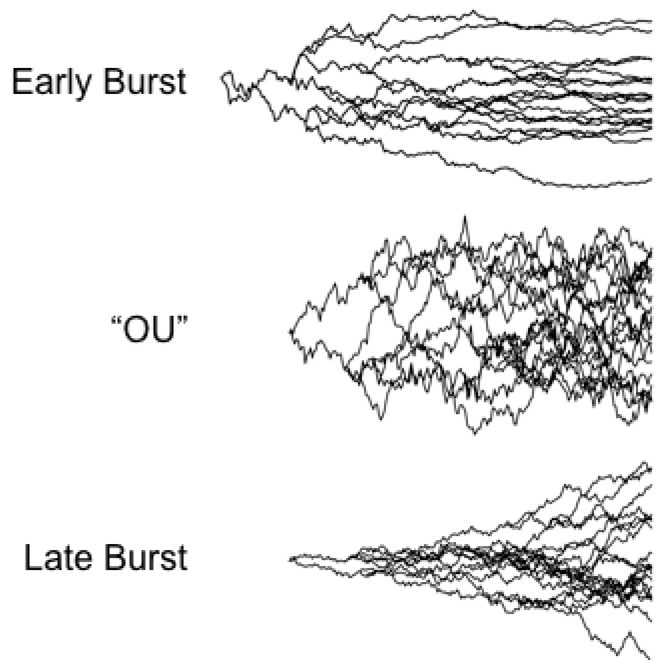
Method extends logic of O’Meara et al. (2006) & Adams (2013)

*Note: methods for identifying rate shifts on particular branches have also been developed (e.g., Castiglione et al. 2018)

Revell et al. 2018. MEE

Other Evolutionary Models: Early Burst

What if evolutionary rate is variable across phylogeny?
Many adaptive models predict a rapid early expansion of phenotypic diversity (a high initial rate of trait evolution that then slows down)

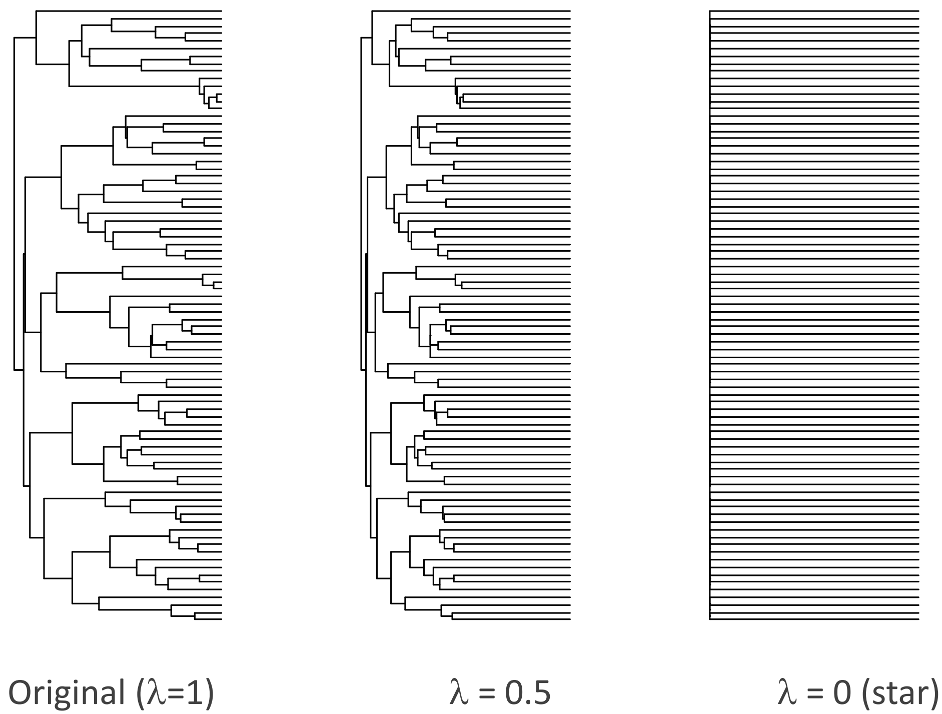


T. Ingram: www.anoleannals.org

Early Burst Model: Contains σ^2 and 'g' (which scales rate of trait change along branches).

Other Evolutionary Models: λ and K

Lambda model: The extent to which the phylogeny predicts covariance among trait values for species (effectively transforms branches by λ)



Kappa model: Punctuational/speciation model: the extent to which trait change corresponds to speciation events (also a branch-length transformation model)

*These models alter elements of **V** for the logL

Pagel 1999 (Nature)

Anolis Example: Multiple Models

BM1:
logL = 5.256010
AIC = -6.512

OU1:
logL = 5.256010
AIC = -4.512

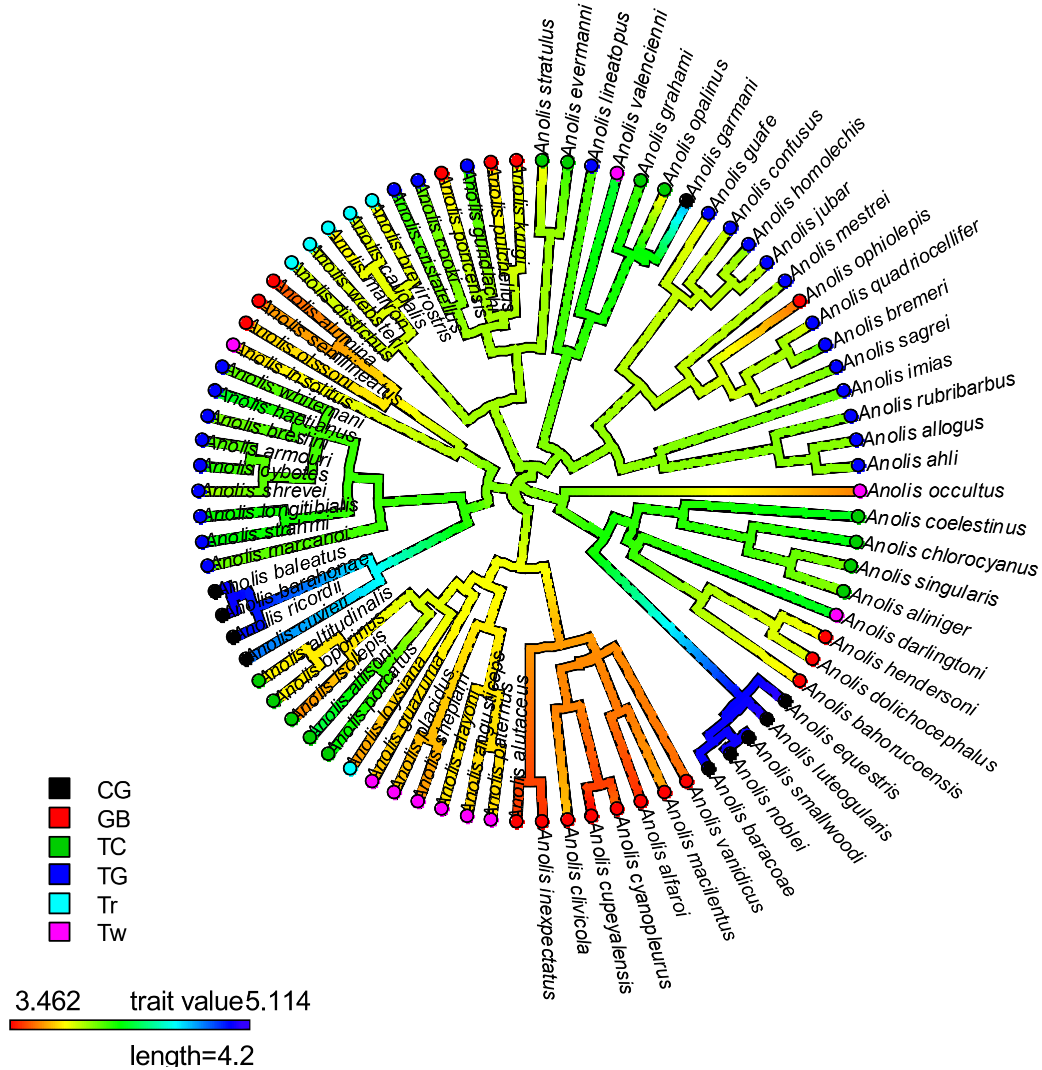
EB:
logL = 6.618
AIC = -7.235

λ:
logL = 5.758
AIC = -5.517

K:
logL = 5.256
AIC = -4.512

BMM:
logL = 21.635
AIC = -29.307

OUM:
logL = 39.4849
AIC = -62.969



OUM by far the best description of the data

Data: Mahler et al. 2010

Exploration: Identifying Evolutionary Models

Can we let the data tell us the best model?

-A HARD statistical problem, as it is 'unsupervised'

Several methods proposed for exploring rate-shifts on phylogeny

1: Bayesian MCMC (Revell et al. 2012)

-search for branches on tree for single largest rate shift

-compare single vs. two-rate model

2: Reversible-jump MCMC (Eastman et al. 2011)

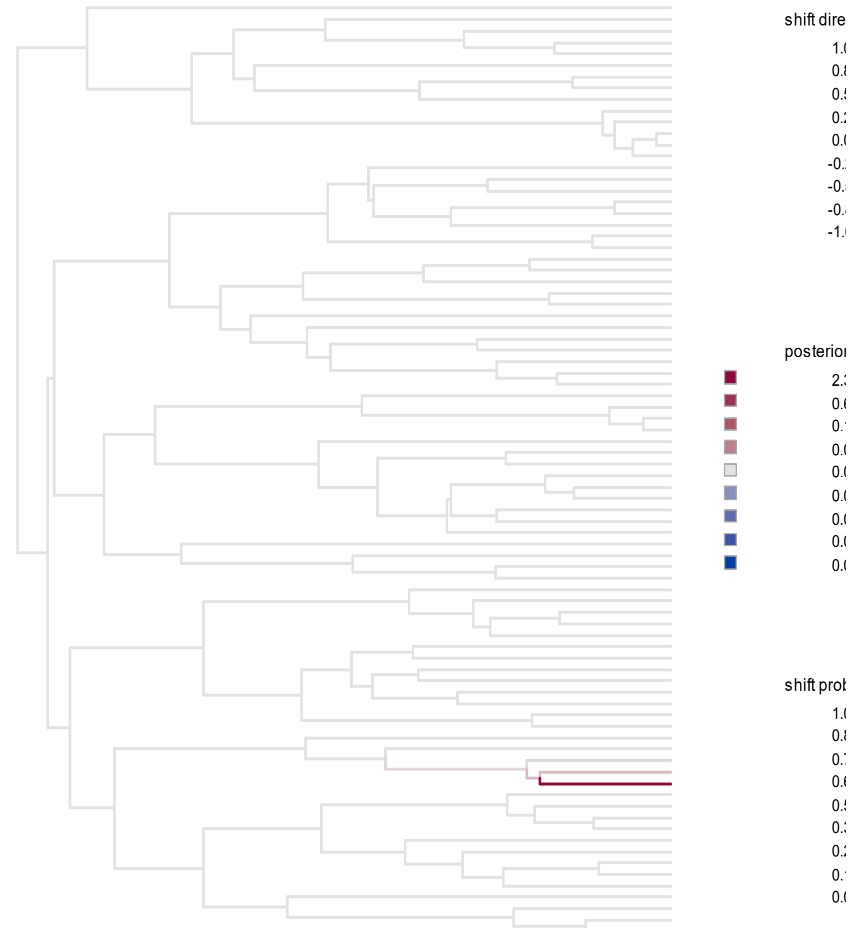
-Search for multiple rate shifts

Anolis Example: Rate Shifts

Bayesian MCMC

$\sigma^2_1 = 0.0255$

$\sigma^2_2 = 0.0182$



RJ-MCMC

Note: identified similar area of tree, but not identical clade/taxon

Data: Mahler et al. 2010

Exploration: Identifying Evolutionary Models

DCA: CAREFUL WITH THESE IMPLEMENTATIONS

- -Algorithms will try very hard to identify rate shifts*
- -Evaluating versus null (BM1) data underexplored
- -More work needed in this area

*NOTE: this is not unlike other unsupervised method in statistics such as for multivariate clustering. Approaches tend to over-identify groups when not present (high type II error) because they are maximizing a search statistic

Conclusions: Evolutionary Models

Evolutionary model comparison:
- -Fit data to phylogeny under alternative models
- -Compare fit using LRT, AIC, simulation, etc.

Very useful for evaluating macroevolutionary hypotheses

BM1, BMM, OU1, OUM, EB, λ and K common models

Evolutionary model comparison is fitting different **V** in the logL

DCA: Careful in interpretation! We tend to think of these as 'process-based' models, but they are phenomenological, pattern-based summaries only.

We don't have data on the branches and nodes to really get at process; all we can do is infer (take the inference with caution!)