Univariate comparative analysis using OUCH*!

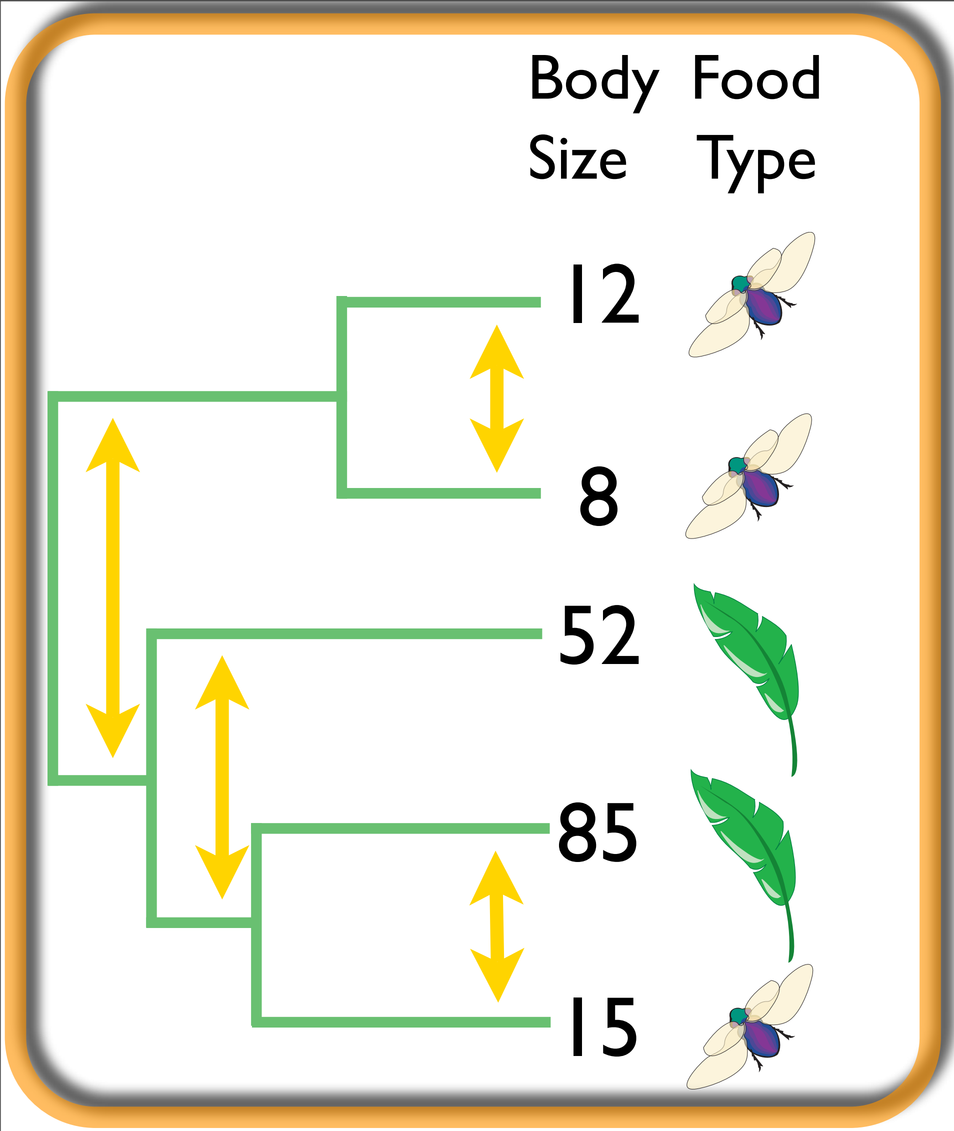
Marguerite Butler

University of Hawaii, Department of Zoology

Aaron King

University of Michigan, Ecology & Evolutionary Biology

(*Ornstein-Uhlenbeck for Comparative Hypotheses)



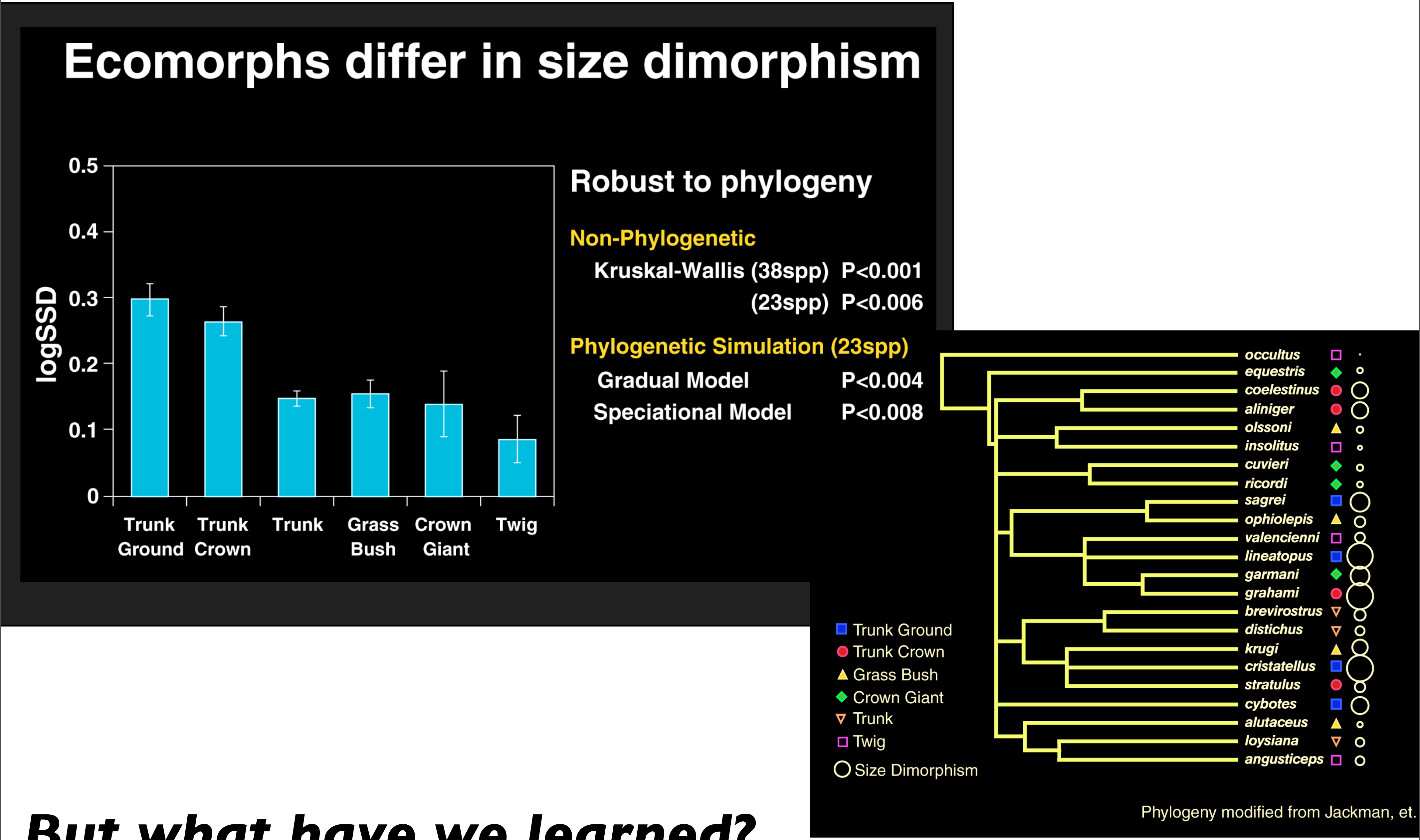
Quantiative character associated with a particular selective “regime”

We want to know the correlation between morphology and ecology

But we can’t use simple statistics because species are related

Statistically remove the effects of phylogeny (using Brownian Motion)

1. “Phylogenetic correction”



Ecomorphs differ in size dimorphism

Robust to phylogeny

Non-Phylogenetic

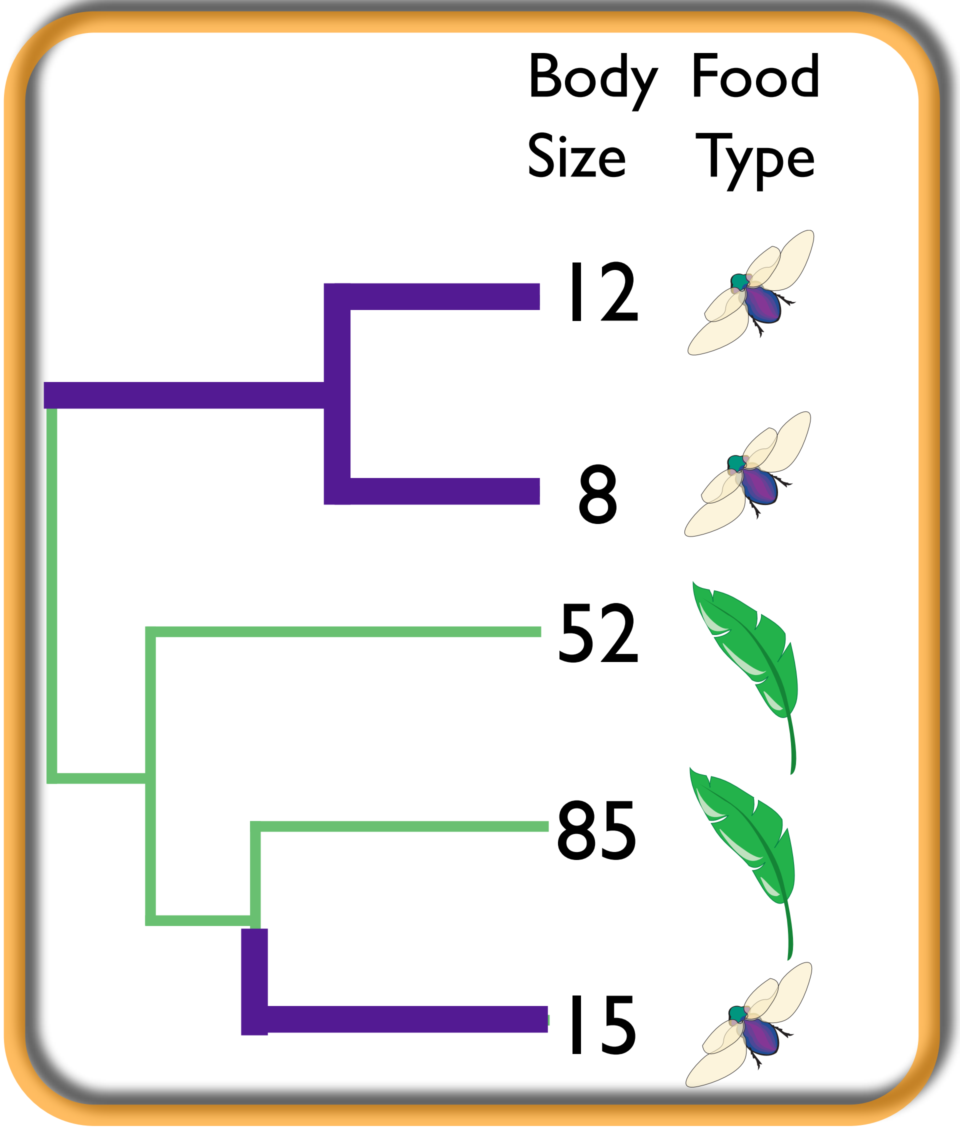
Kruskal-Wallis (38spp)	P<0.001
(23spp)	P<0.006

Phylogenetic Simulation (23spp)

Gradual Model	P<0.004
Speciational Model	P<0.008

Phylogeny modified from Jackman, et.

But what have we learned?



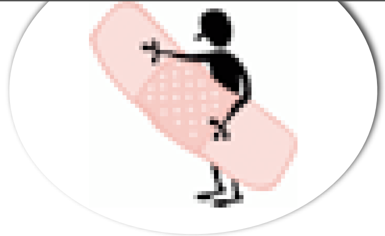
Recent theoretical developments have made it possible to perform comparative analyses using an explicit evolutionary model

But most often they take not account of selection

Even though many comparative studies focus on adaptive hypotheses

If we are to reflect variation in adaptive regimes, we must be able to model multiple evolutionary optima

2. “Model the Evolutionary Process”



The phylogeny (pattern and timing of evolutionary diversification) as well as the data contains important information

Model the evolutionary process along each branch of the phylogeny

Brownian Motion

$$dX_i(t) = \sigma\, dB_i(t), \quad t_i^{j-1} \le t \le t_i^j.$$

Orstein Uhlenbeck Process

$$dX_i(t) = \alpha\,(\beta_i^j - X_i(t))\, dt + \sigma\, dB_i(t)$$

Vary the models to reflect biology
Then compare to find the best model

Ornstein Uhlenbeck Process

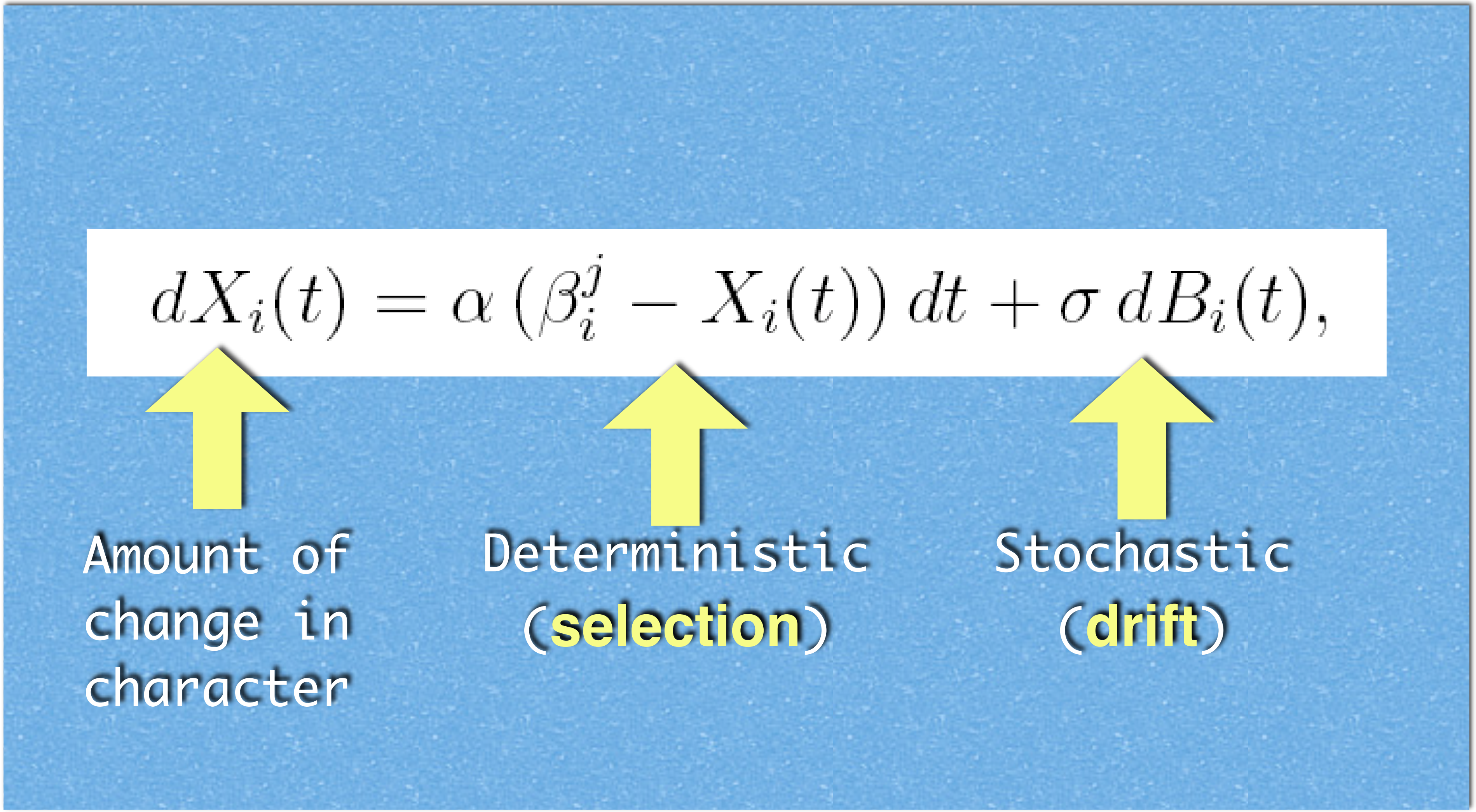
A model for evolution with selection

time interval t: $t_i^{j-1} \leq t \leq t_i^j.$

$$dX_i(t) = \alpha \left(\beta_i^j - X_i(t)\right) dt + \sigma\, dB_i(t),$$

Hansen (1997)

Ornstein Uhlenbeck Process



$$dX_i(t) = \alpha\,(\beta_i^j - X_i(t))\,dt + \sigma\,dB_i(t),$$

Amount of change in character

Deterministic (**selection**)

Stochastic (**drift**)

OU becomes Brownian Motion in limit as α → 0

$$dX_i(t) = \overset{0}{\cancel{\alpha}}(\beta_i^j - X_i(t))\,dt + \sigma\,dB_i(t),$$

OU becomes Brownian Motion in limit as α → 0

$$dX_i(t) = \xcancel{\cancelto{}{\alpha}(\beta_i^j - X_i(t))}\,dt + \sigma\,dB_i(t),$$

Brownian Motion

$$dX_i(t) = \qquad \sigma\, dB_i(t),$$

Amount of change in character

Brownian Motion

$$dX_i(t) = \qquad\qquad \sigma\, dB_i(t),$$

proportional
to std dev

Amount of
change in
character

Brownian Motion

$$dX_i(t) = \sigma\, dB_i(t),$$

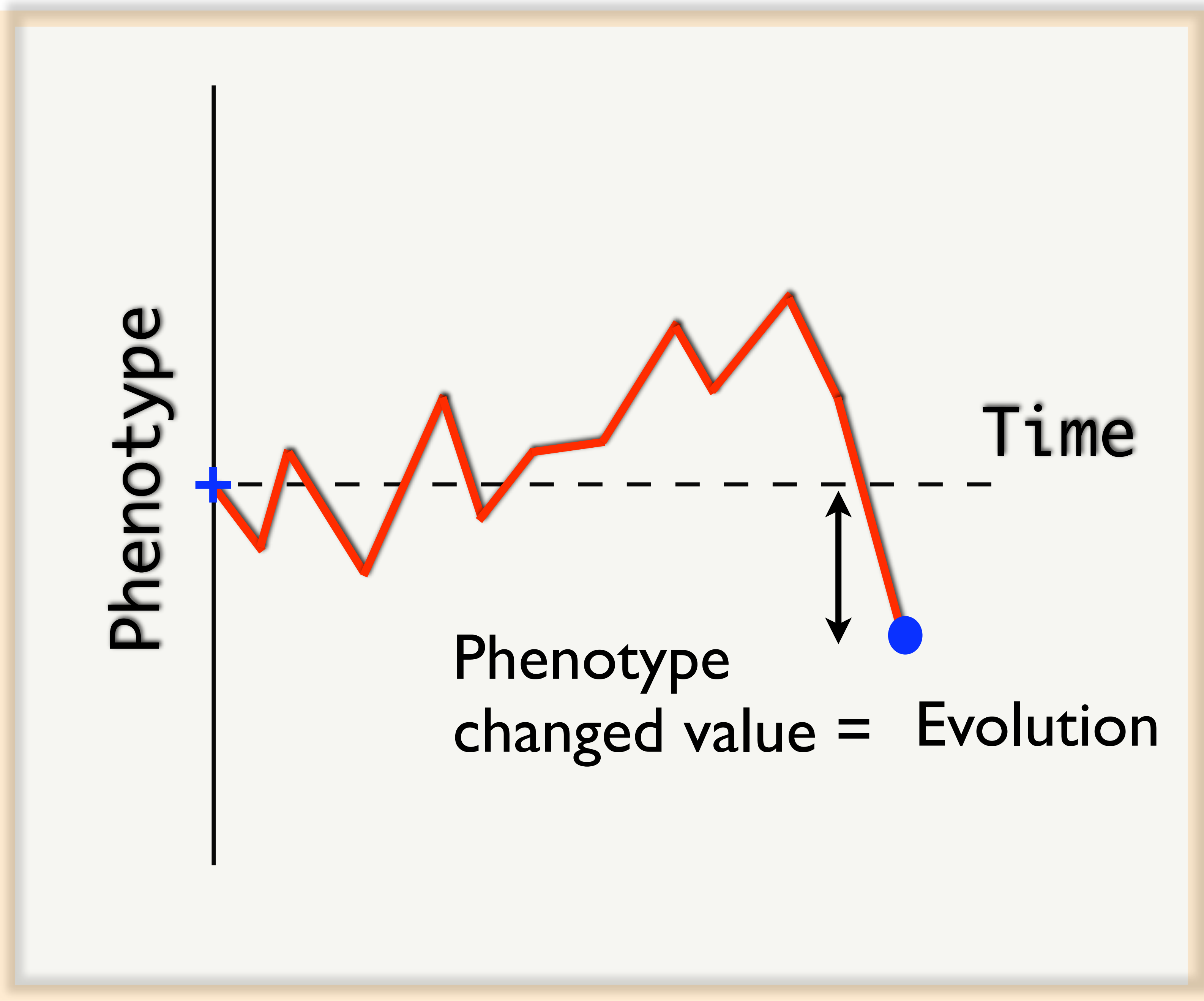
Amount of change in character

proportional to std dev

and a random amount of change

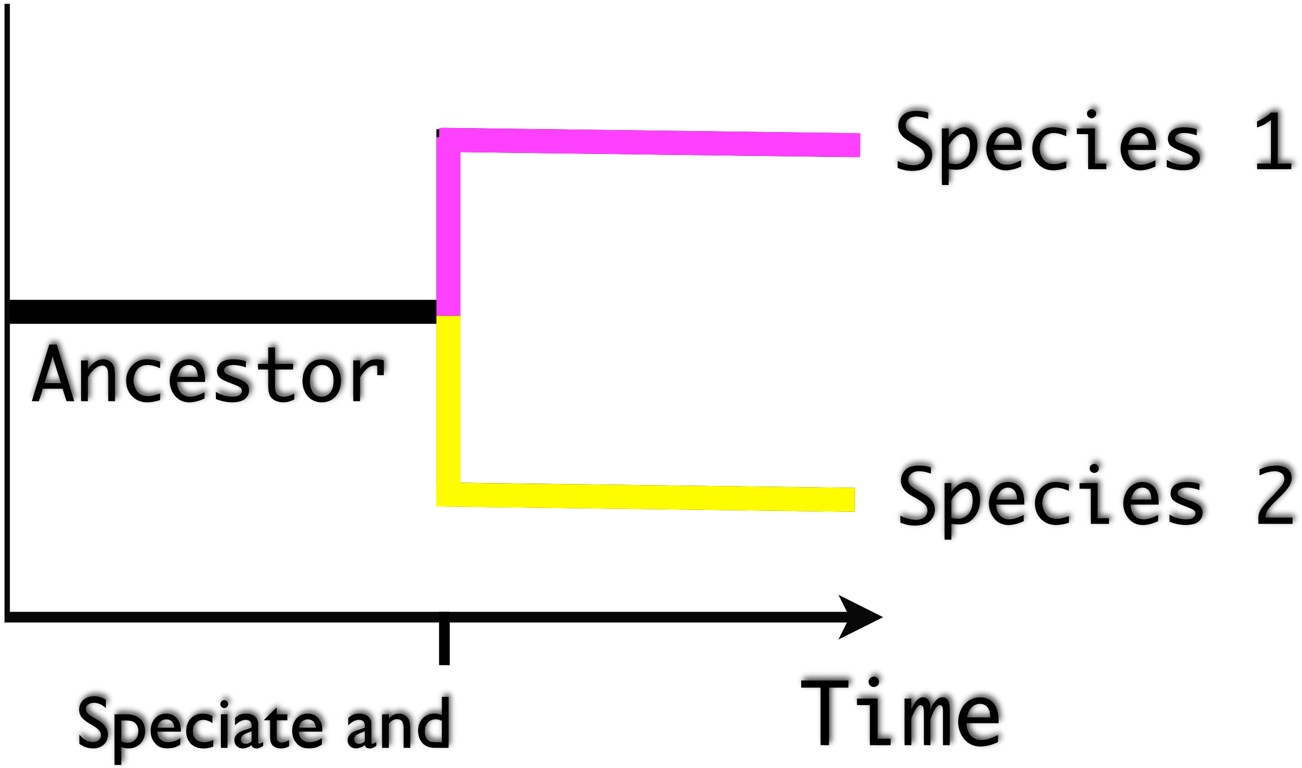
Illustrating the Model

We can build up simulations of evolution



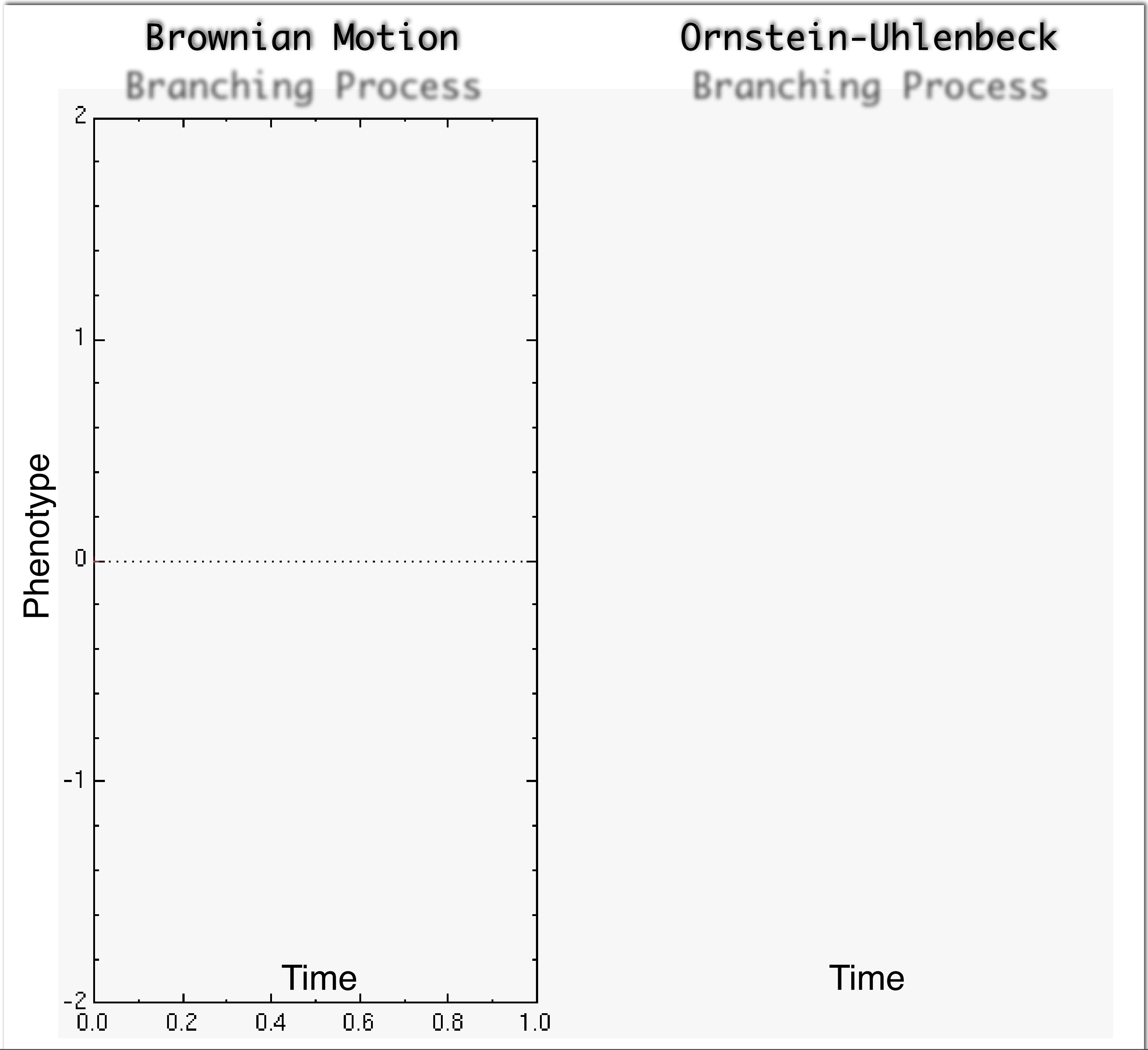
Phenotype changed value = Evolution

BM and OU models make different predictions



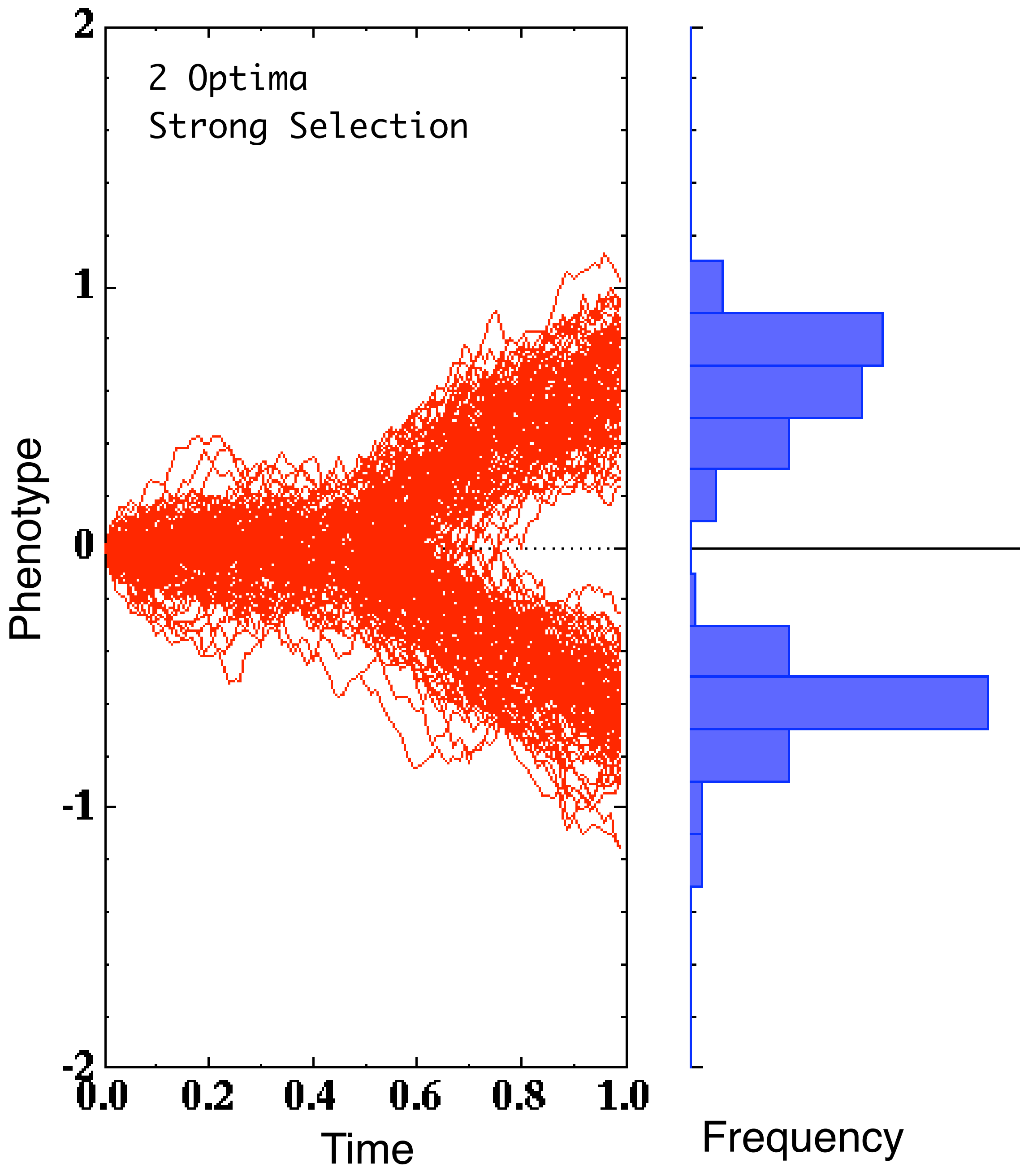
Speciate and
enter different
adaptive regimes
at time t=0.5

BM and OU models make different predictions



OU Branching Process

Phenotypic Distribution



Thus, with:

a set of interspecific data,
a phylogeny, and
a little biological insight,

we can explore alternative evolutionary scenarios

and potentially make a statement about how characters evolved!

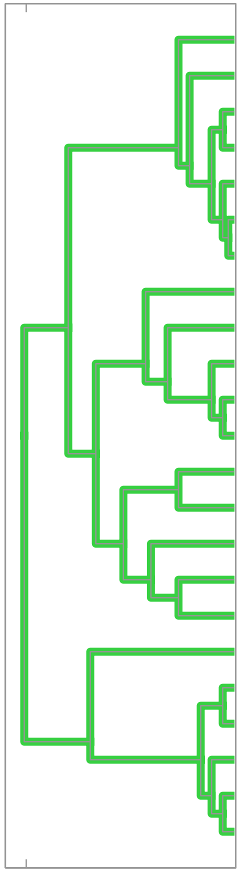
Islands in the Lesser Antilles have only 1 or 2 species of anoles.

On the 2 species islands, anoles differ substantially in size.

On the single species islands, anoles are “intermediate” in size.

Alternative Adaptive Regime Models

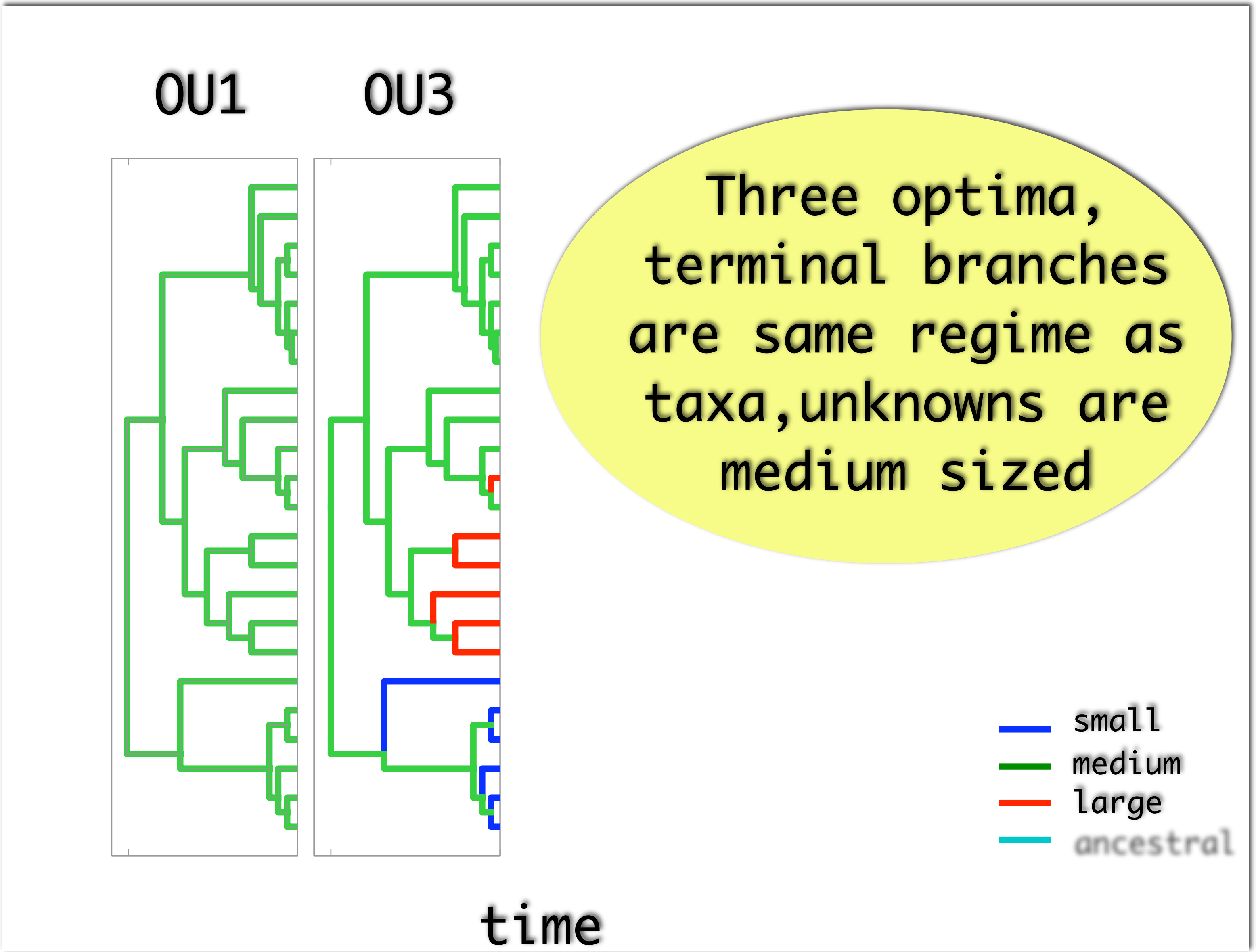
OU1



Ornstein Ulhenbeck Process with one (global) optimum

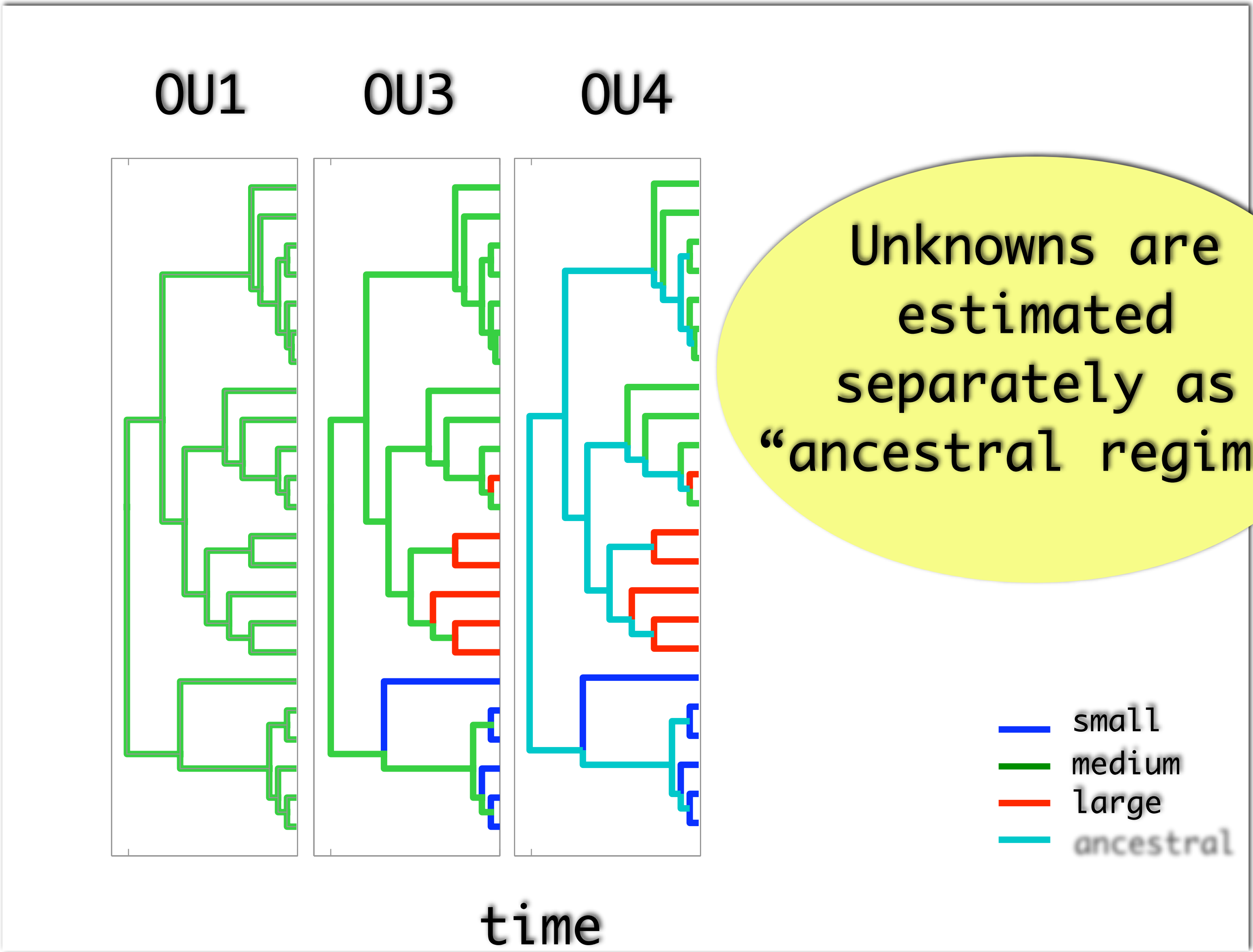
time

Alternative Adaptive Regime Models



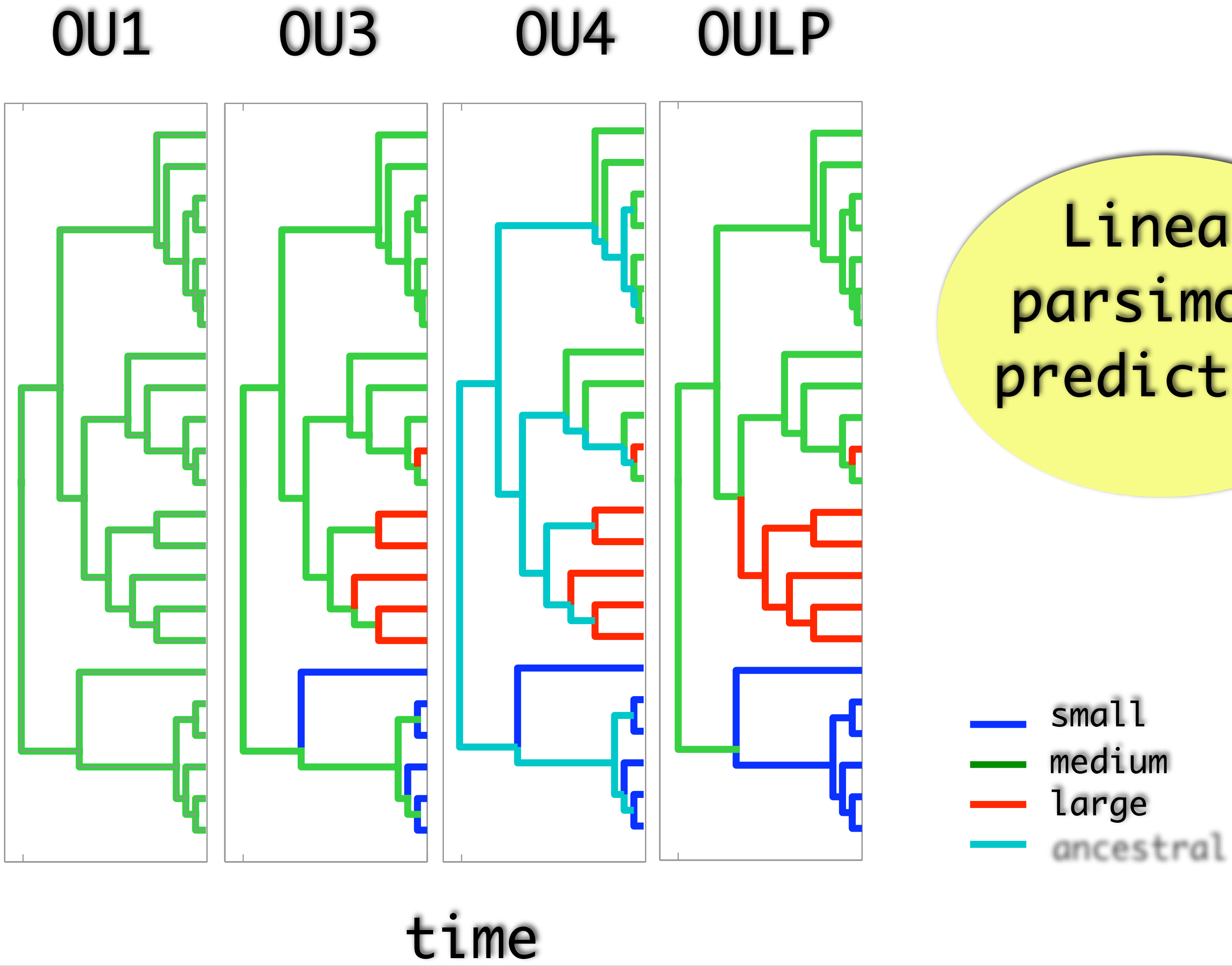
Three optima, terminal branches are same regime as taxa,unknowns are medium sized

Alternative Adaptive Regime Models



Unknowns are estimated separately as “ancestral regime”

Alternative Adaptive Regime Models



Linear
parsimony
prediction

Model Comparison

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
-2x Log Likelihood	-36.31	-34.66	-40.06	-47.22	-49.69
AIC	-32.31	-26.66	-28.06	-33.22	-37.69
LR		0.00	0.56	0.95	0.99
Pvalue		1.00	0.44	0.05318	0.00955

Model Comparison

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
-2x Log Likelihood	-36.31	-34.66	-40.06	-47.22	-49.69
AIC	-32.31	-26.66	-28.06	-33.22	-37.69
LR		0.00	0.56	0.95	0.99
Pvalue		1.00	0.44	0.05318	0.00955

Bad Models

Strange Parameter Estimates

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
alpha		0	0.32	14.67	2.49
sigma	0.2	0.21	0.2	0.47	0.22
ancestor	2.83	2.95	3.99	--	0.86
optima 1		--	-1.4	2.58	2.75
optima 2			0.18	3.11	3.24
optima 3			2.71	3.3	3.56
optima 4				2.83	

Best Model

	BM	OU(1)	OU(3)	OU(4)	OU(LP)
alpha		0	0.32	14.67	2.49
sigma	0.2	0.21	0.2	0.47	0.22
ancestor	2.83	2.95	3.99	--	0.86
optima 1		--	-1.4	2.58	2.75
optima 2			0.18	3.11	3.24
optima 3			2.71	3.3	3.56
optima 4				2.83	

Best Model

$$\mathbb{E}\left[X_i(T)\,\middle|\,X_i(0)=\theta_0\right]$$

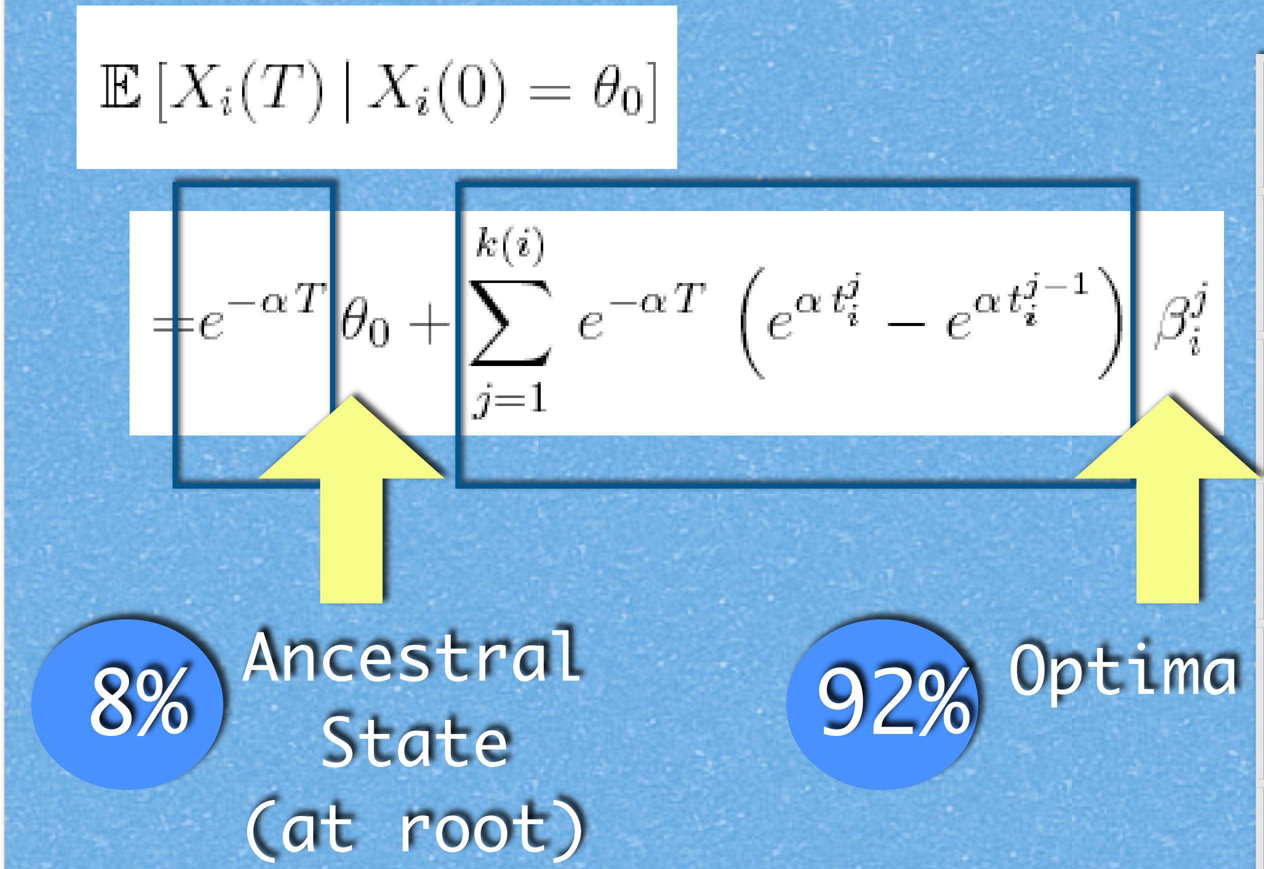
$$= e^{-\alpha T}\theta_0 + \sum_{j=1}^{k(i)} e^{-\alpha T}\left(e^{\alpha t_i^j} - e^{\alpha t_i^{j-1}}\right)\beta_i^j$$

Ancestral State (at root)

Optima

	OU(LP)
alpha	2.49
sigma	0.22
ancestor	0.86
optima 1	2.75
optima 2	3.24
optima 3	3.56

Best Model



$$\mathbb{E}\left[X_i(T)\,|\,X_i(0)=\theta_0\right]$$

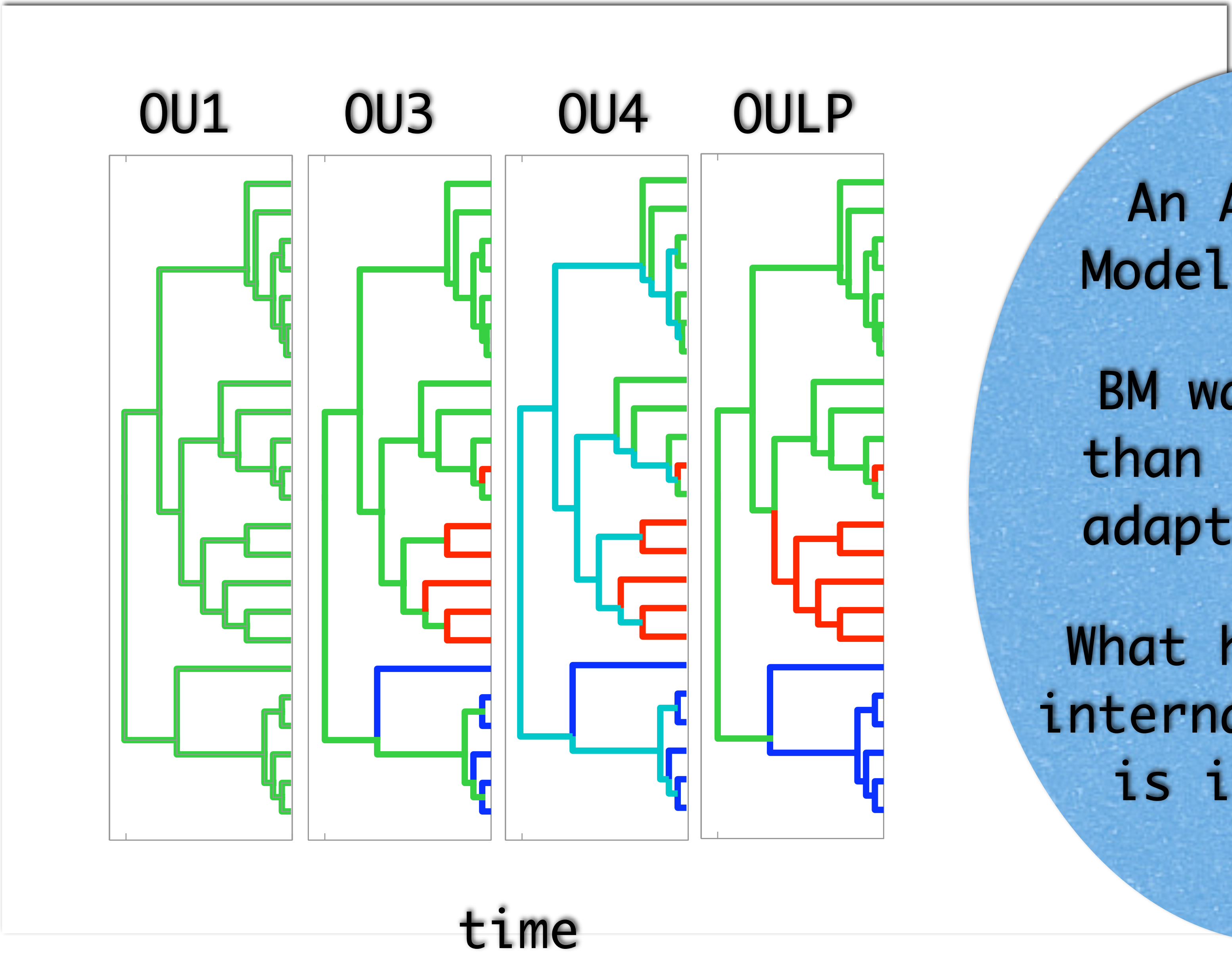
$$= \boxed{e^{-\alpha T}}\,\theta_0 + \boxed{\sum_{j=1}^{k(i)} e^{-\alpha T}\left(e^{\alpha t_i^j} - e^{\alpha t_i^{j-1}}\right)}\,\beta_i^j$$

8% Ancestral State (at root)

92% Optima

	OU(LP)
alpha	2.49
sigma	0.22
ancestor	0.86
optima 1	2.75
optima 2	3.24
optima 3	3.56

Univariate Conclusions



An Adaptive Model was best

BM was better than the wrong adaptive model

What happens on internal branches is important

Parametric Bootstrap indicates “**high**” and “**low**” SSD optima associated with habitat type

	OU(7)	95% CI
alpha	1.44	(0.24, 4.5)
sigma	0.09	(0.04, 0.11)
ancestor	0.23	(0.03, 0.27)
optima tg	0.327	(0.26, 0.97)
optima tc	0.134	(0.00, 0.27)
optima gb	-0.013	(-0.28, 0.13)
optima cg	-0.060	(-0.39, 0.09)
optima tw	-0.074	(-0.42, 0.06)
optima tr	-0.051	(-0.39, 0.12)
optima uk	0.468	(0.34, 1.25)

N=1000

Final Thoughts (Rantings):

Biologists think about adaptive evolution
tion, so WHY use a
tral?

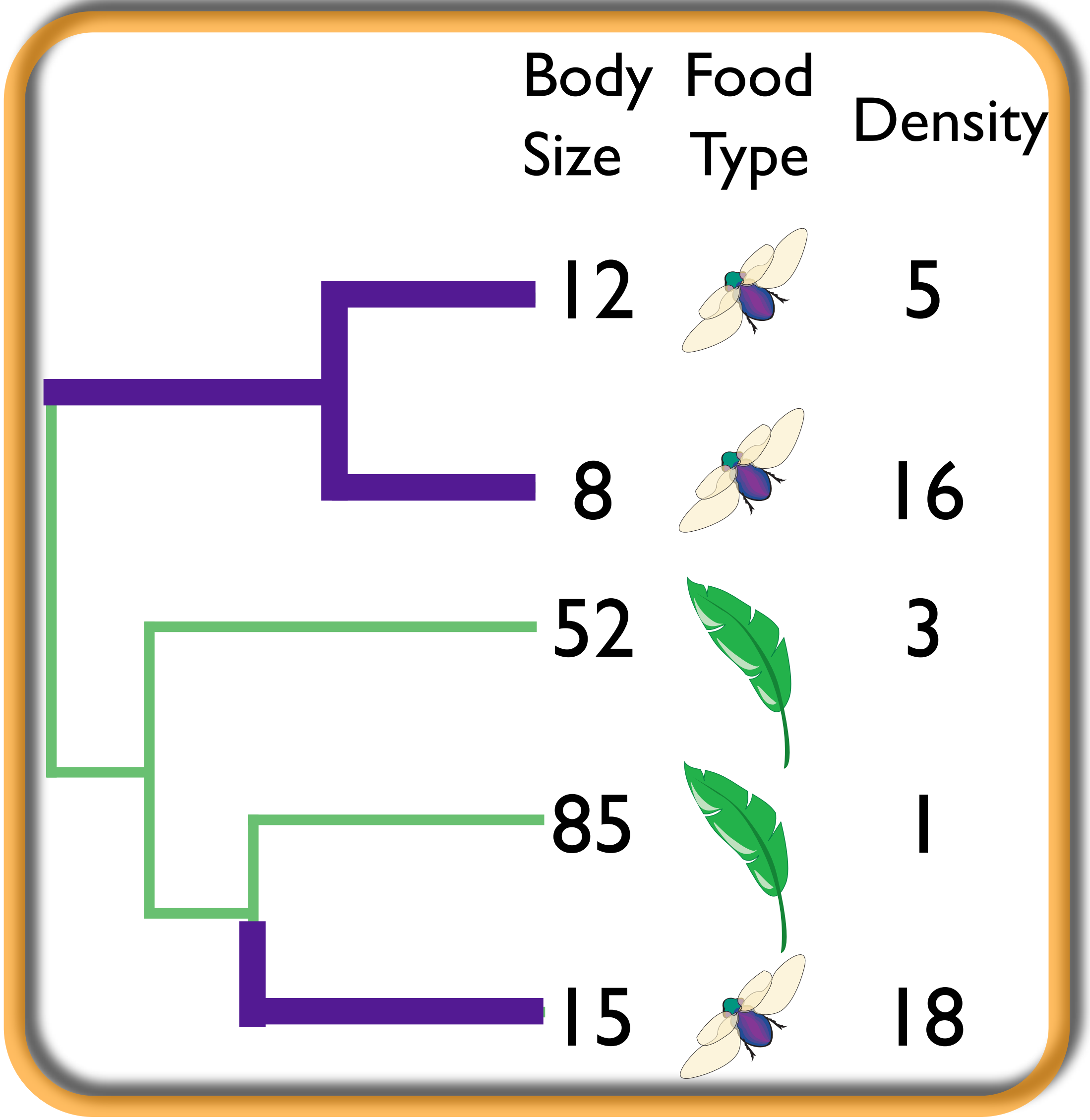
This should be a conclusion, not an assumption.

BE EXPLICIT!! It’s important.

We gain nothing by using a null model that is **really bad** and then rejecting it!

We gain much more insight by using reasonable models and comparing them.

“Correcting for phylogeny” is not a useful way to think about comparative analysis.



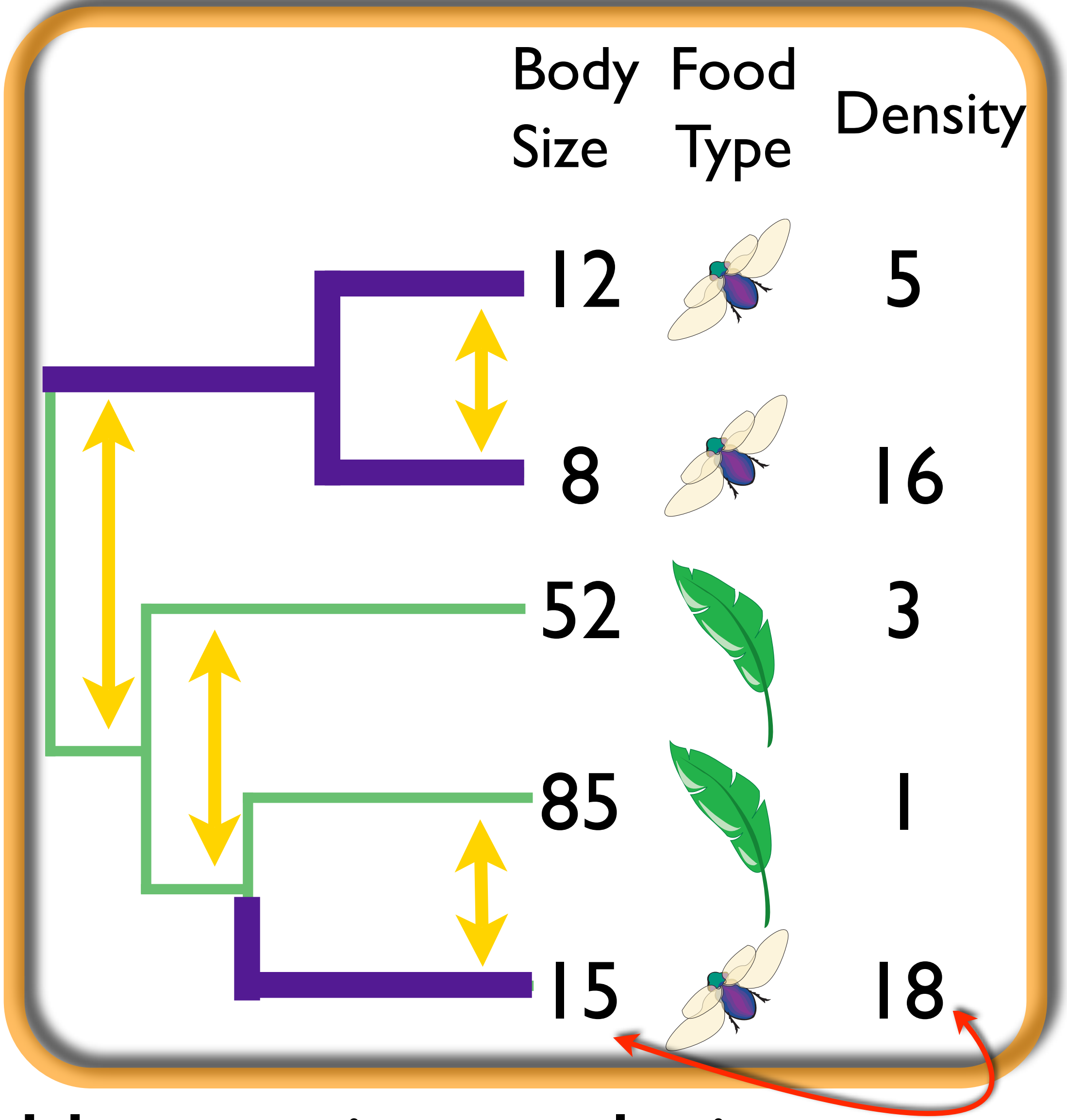
Why study correlated evolution?

Helps us to understand patterns in diversity

Reveals important functional relationships

Correlated selection can also prevent traits from reaching their individual optima

Here we develop a general modeling approach for ***multiple characters***

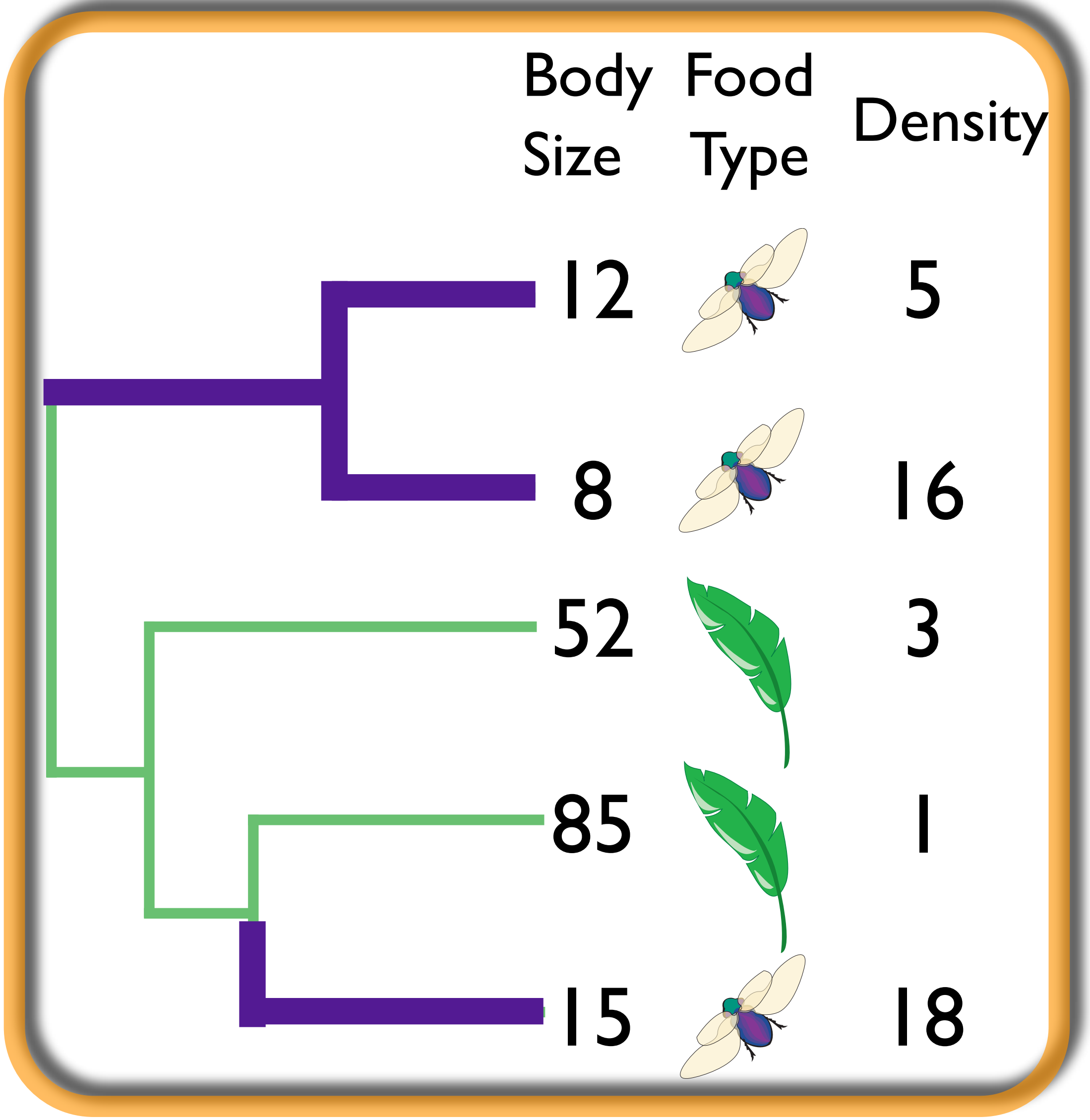


Previously, we developed an approach for comparative analysis which allows testing for alternative adaptive regimes

But...
only for one character

Uncovering evolutionary correlations is the most popular use of comparative methods

Here we develop a general modeling approach for ***multiple characters***



Body Size	Food Type	Density
12	🪰	5
8	🪰	16
52	🍃	3
85	🍃	1
15	🪰	18

Why study correlated evolution?

Helps us to understand patterns in diversity

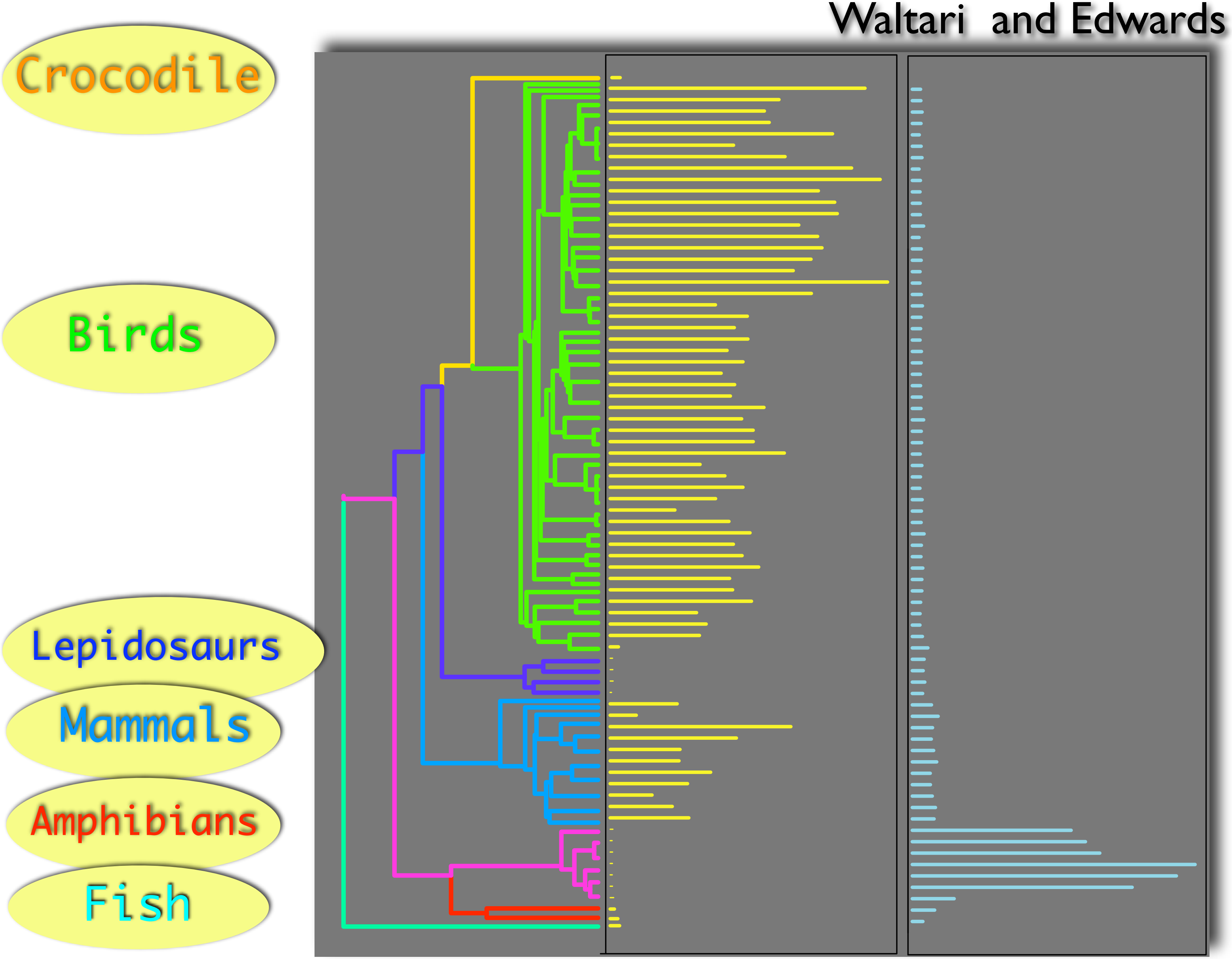
Reveals important functional relationships

Correlated selection can also prevent traits from reaching their individual optima

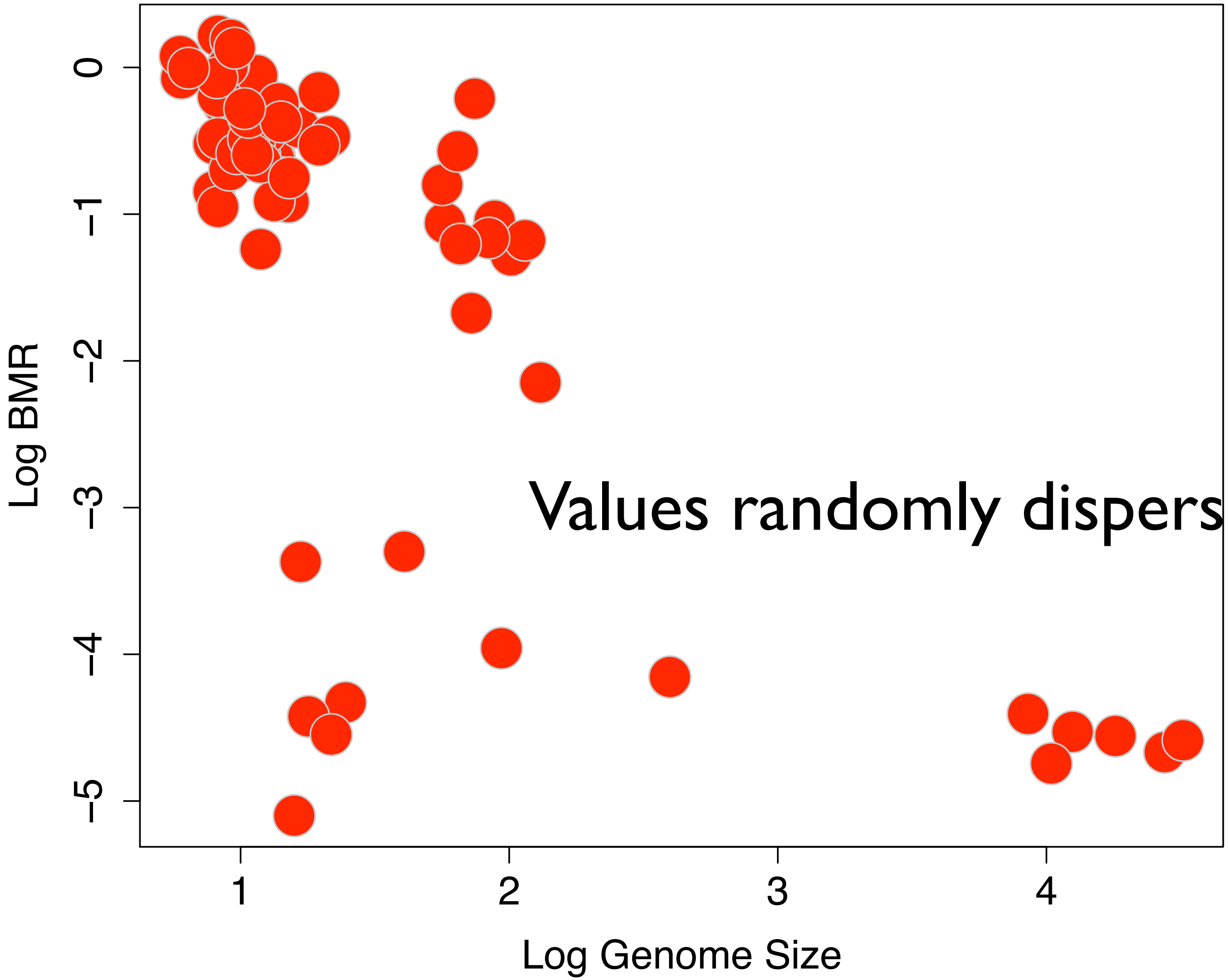
Here we develop a general modeling approach for ***multiple characters***

Basal Metabolic Rate and Genome Size

Waltari and Edwards 2002

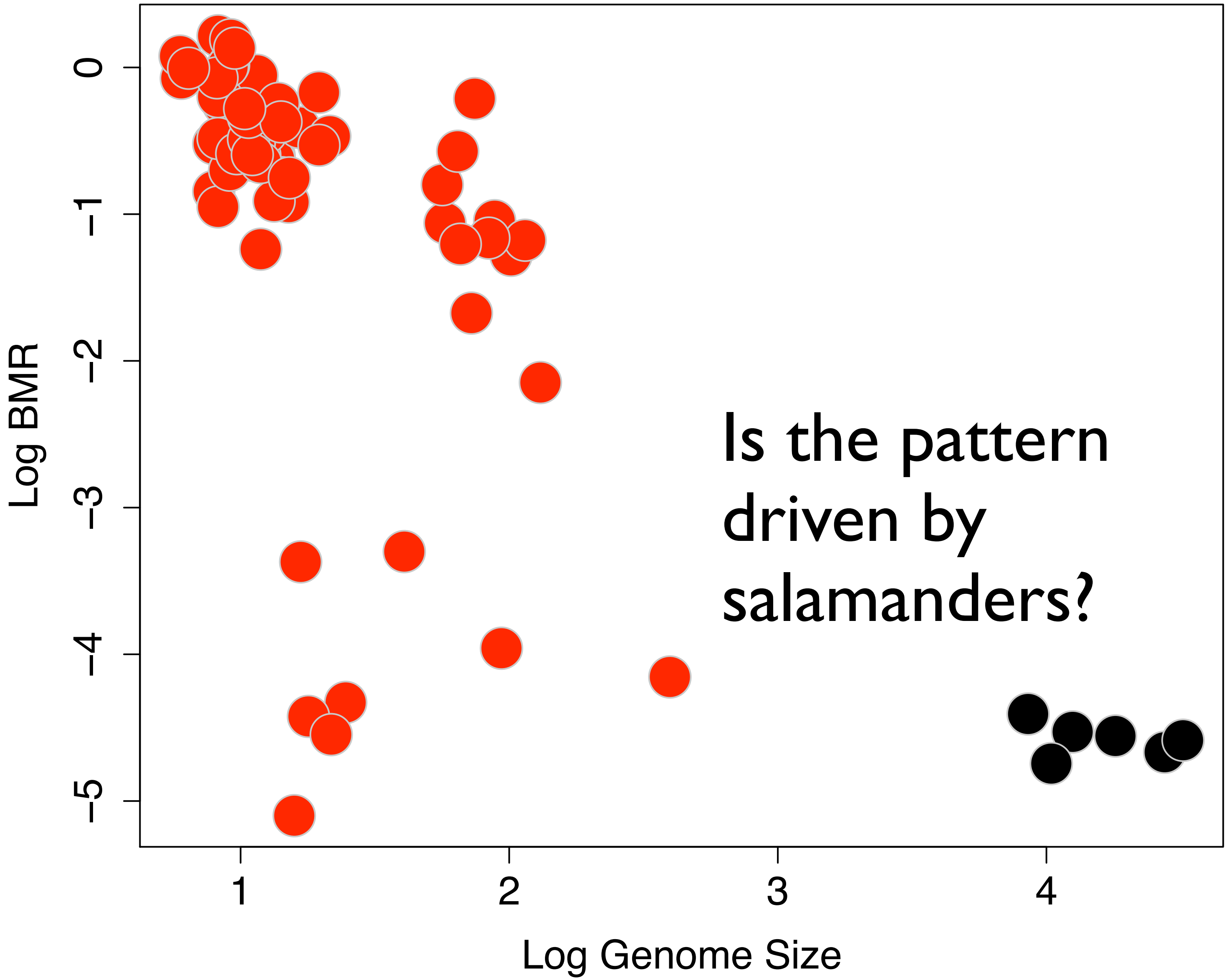


How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?



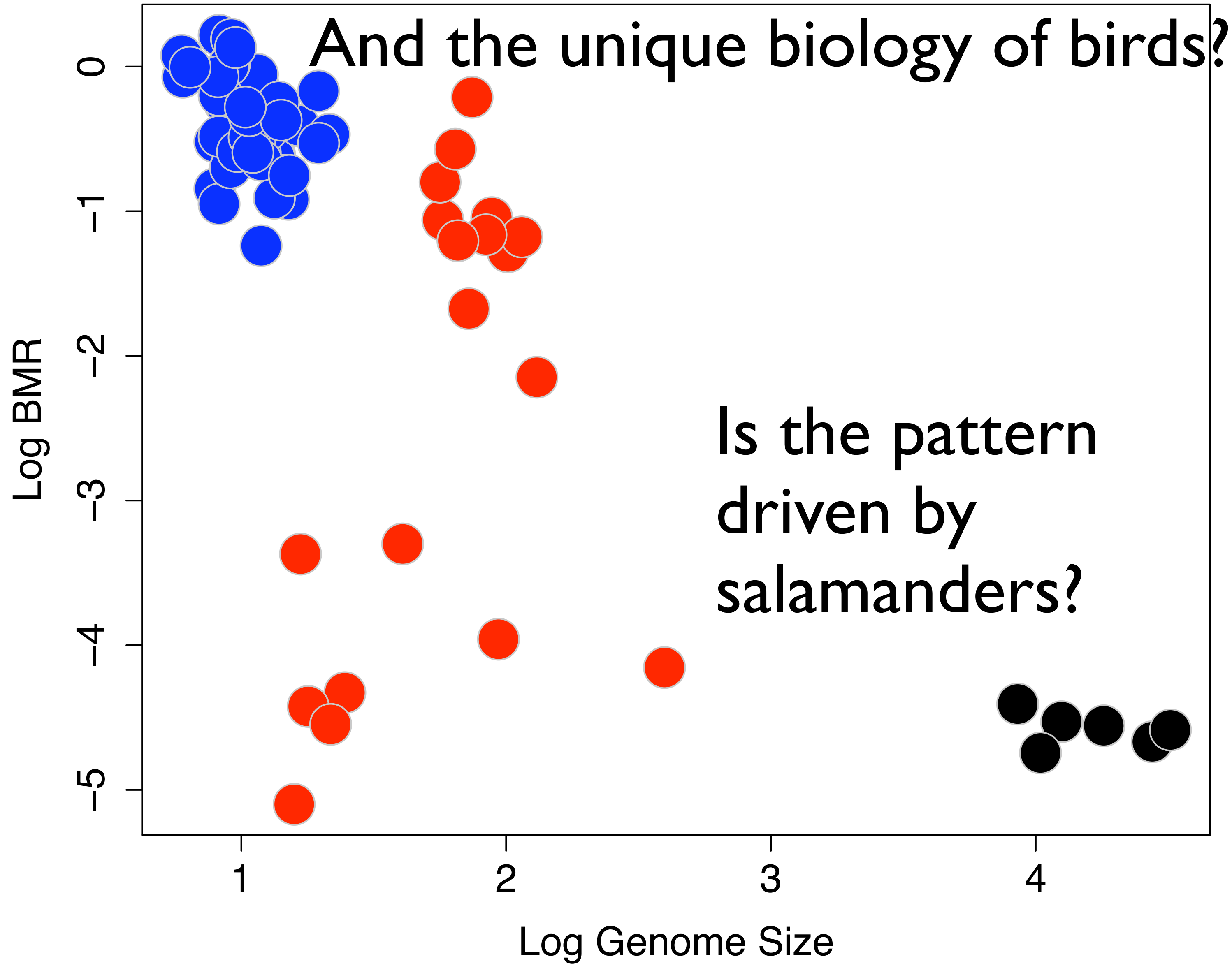
Values randomly dispersed?

How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?



Is the pattern driven by salamanders?

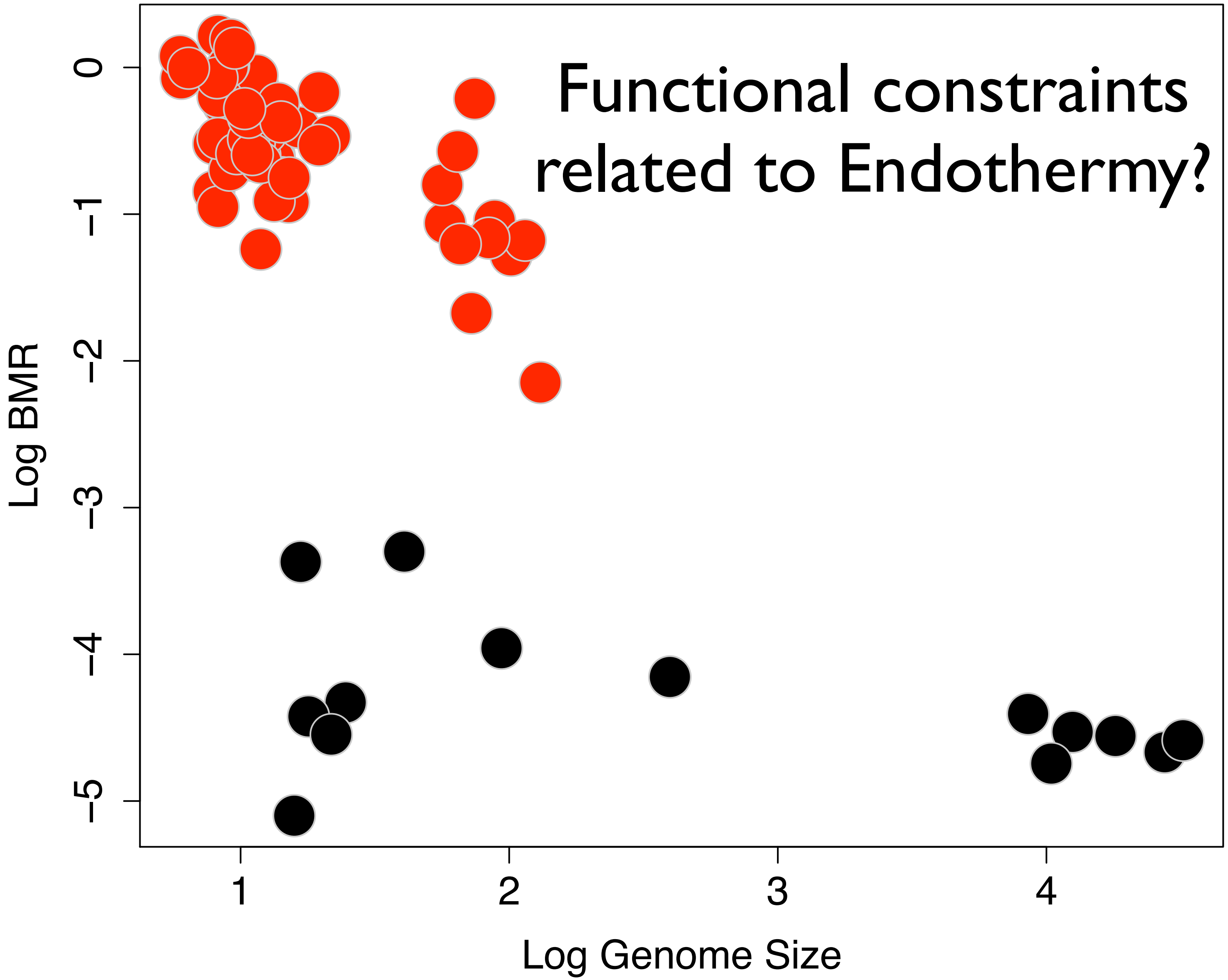
How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?



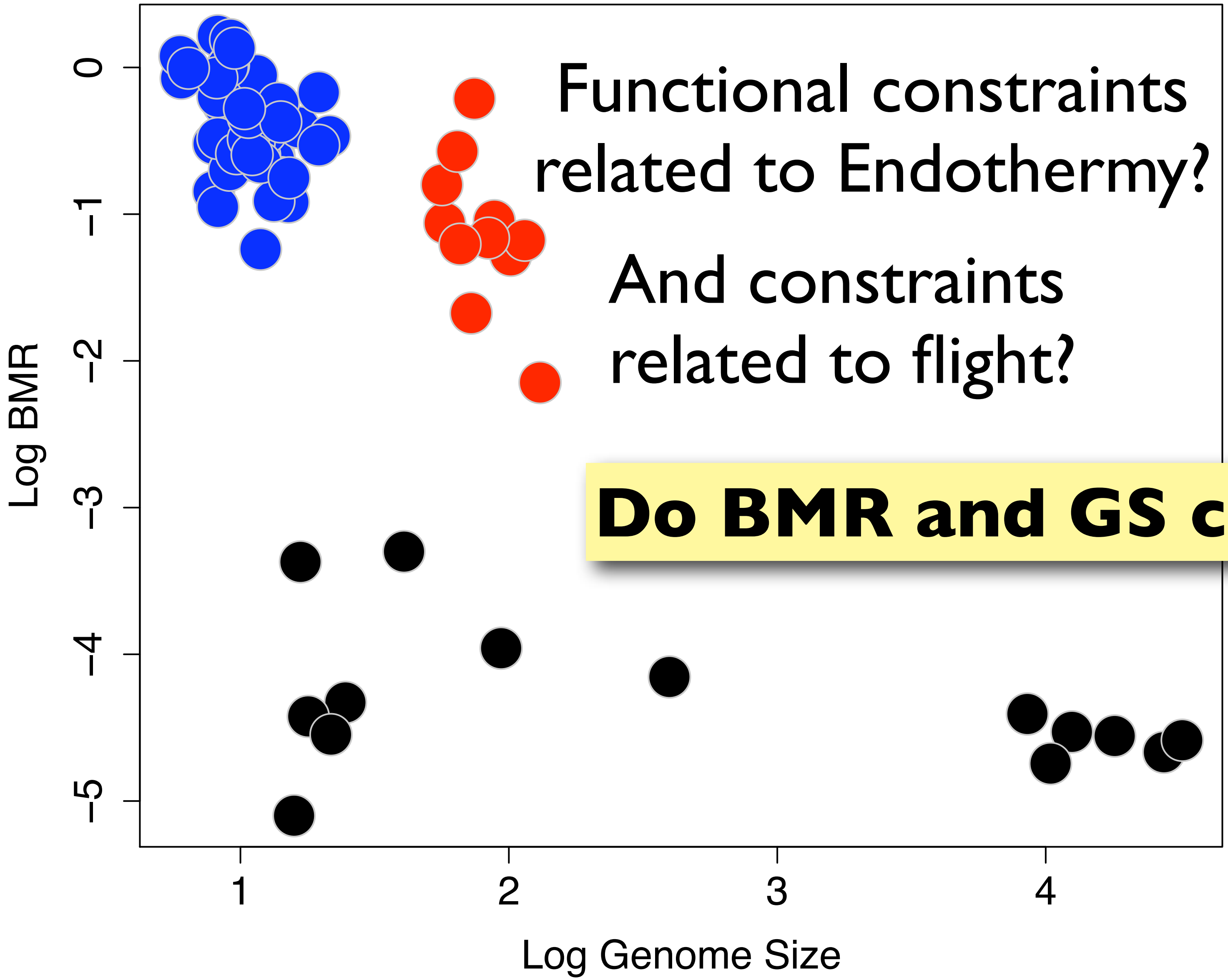
And the unique biology of birds?

Is the pattern driven by salamanders?

How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?



How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?



Functional constraints related to Endothermy?

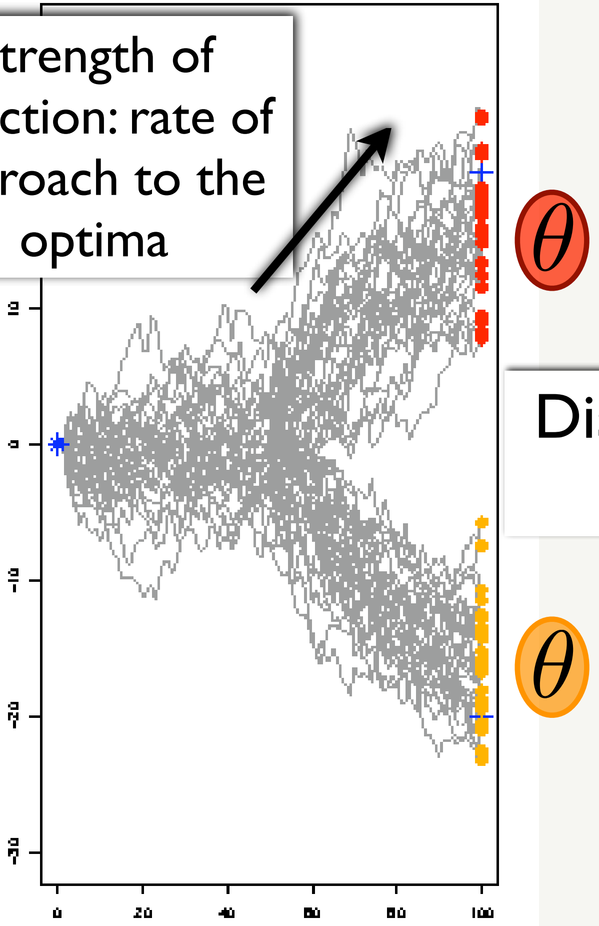
And constraints related to flight?

Do BMR and GS covary?

Two well-known players

$$dX(t) = \boxed{\sigma}\, dB(t)$$

Ornstein-Uhlenbeck

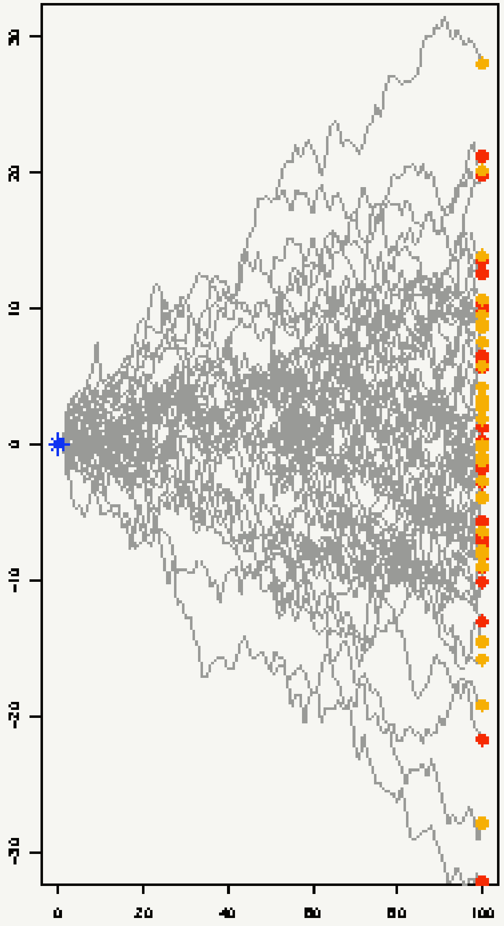


Strength of selection: rate of approach to the optima

Distance from optima

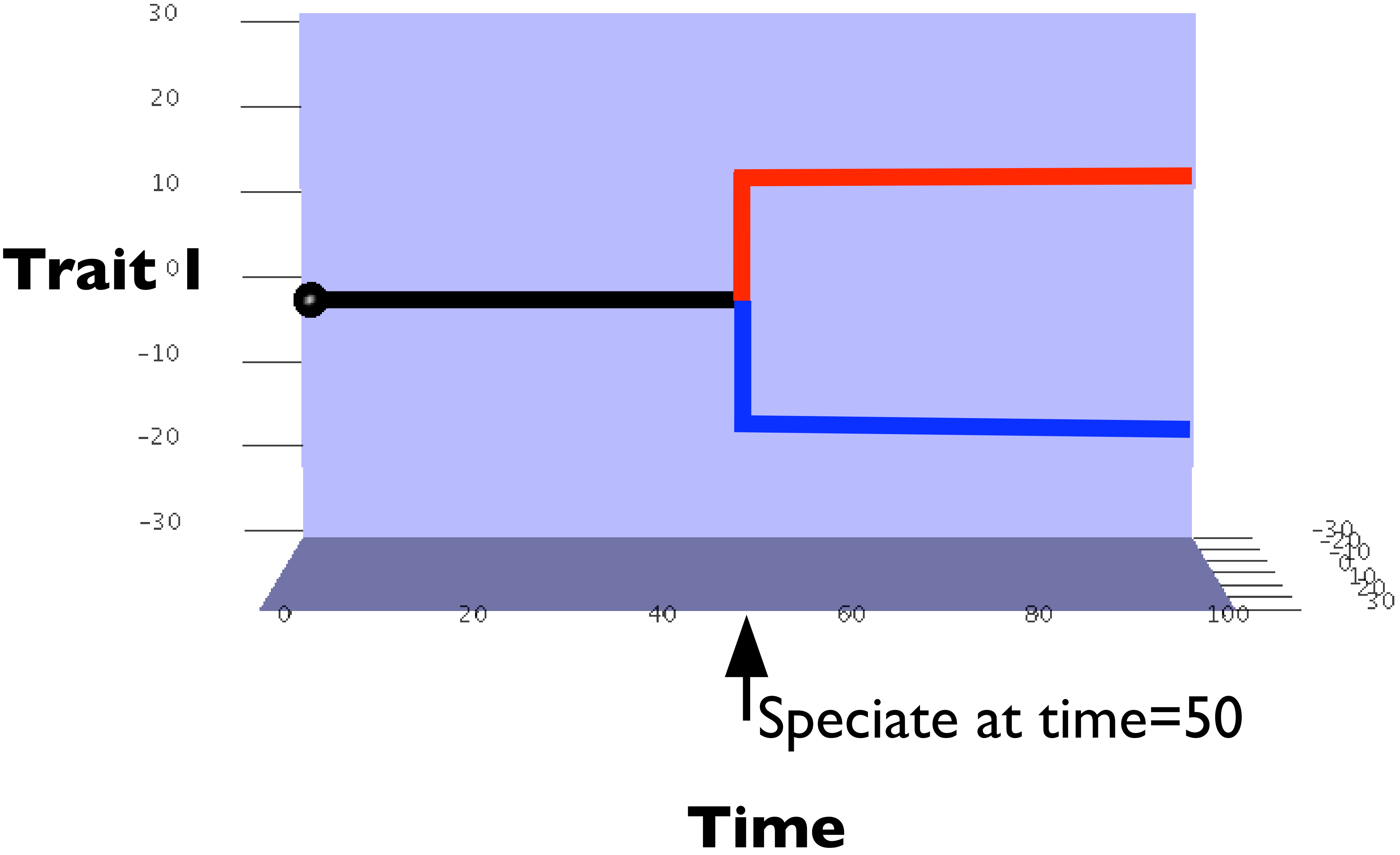
"Selection"

Brownian motion

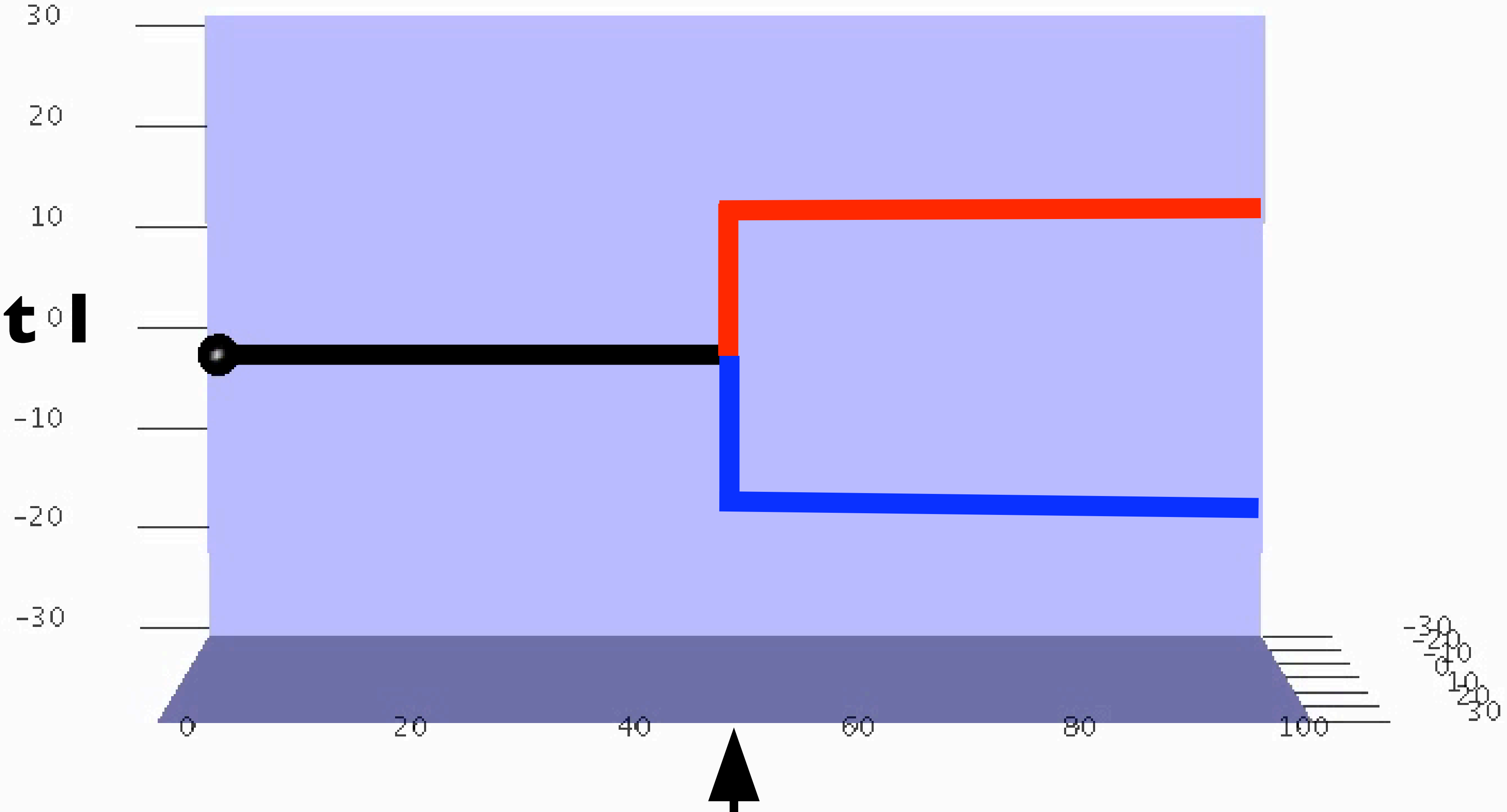


"Drift"

Brownian motion in 2D



Brownian motion in 2D

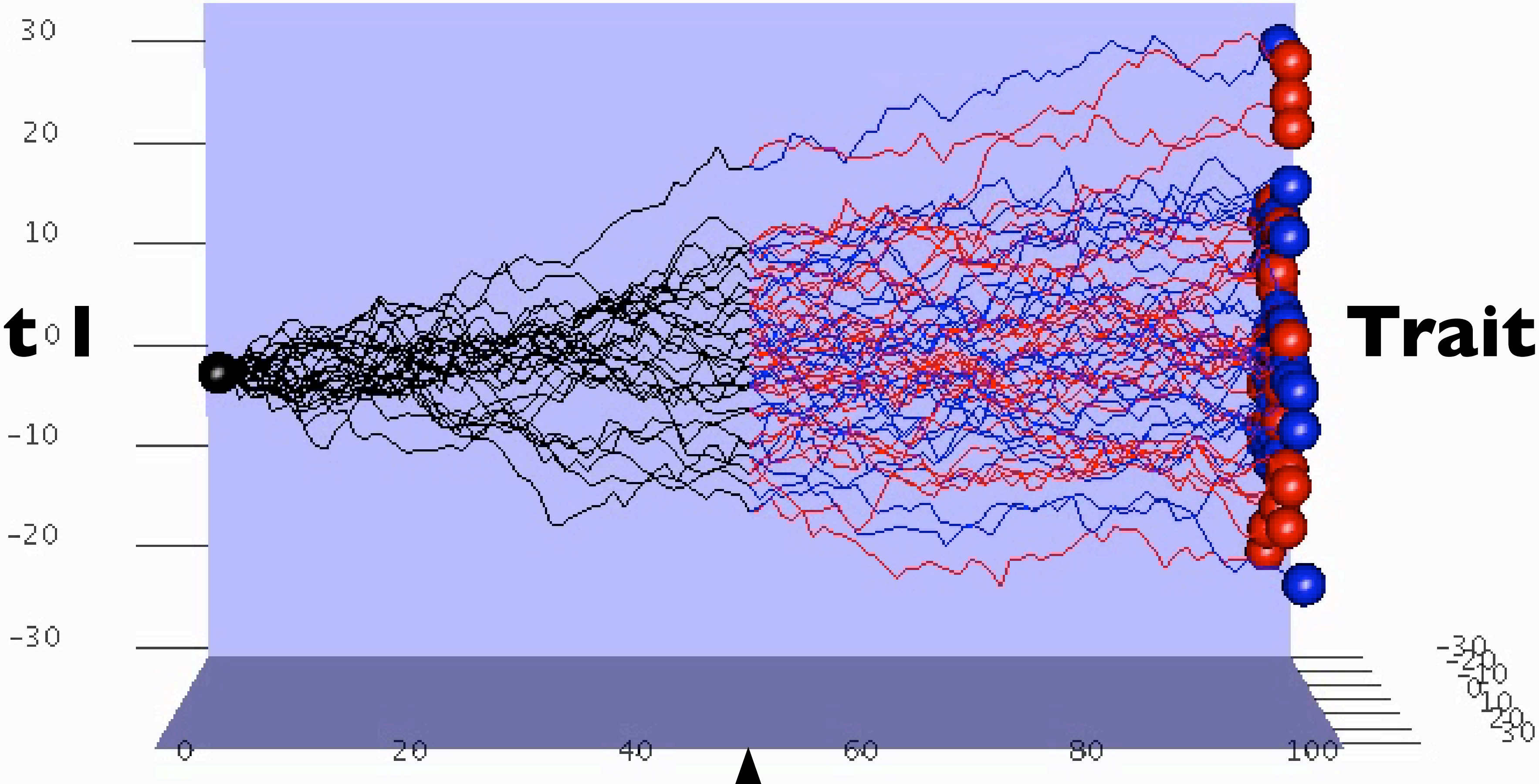


Trait 1

Speciate at time=50

Time

Brownian motion in 2D



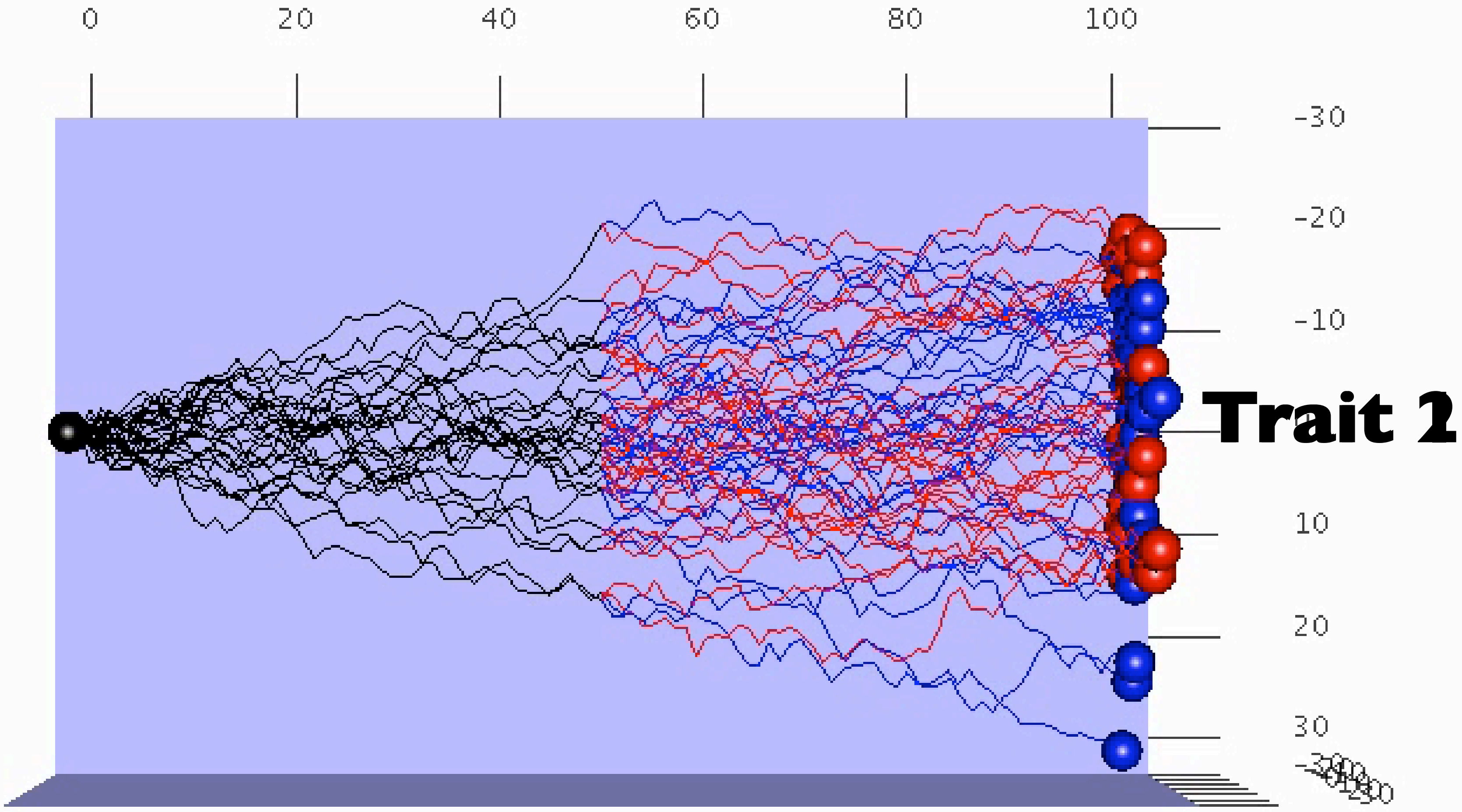
Trait 1

Trait 2

Speciate at time=50

Time

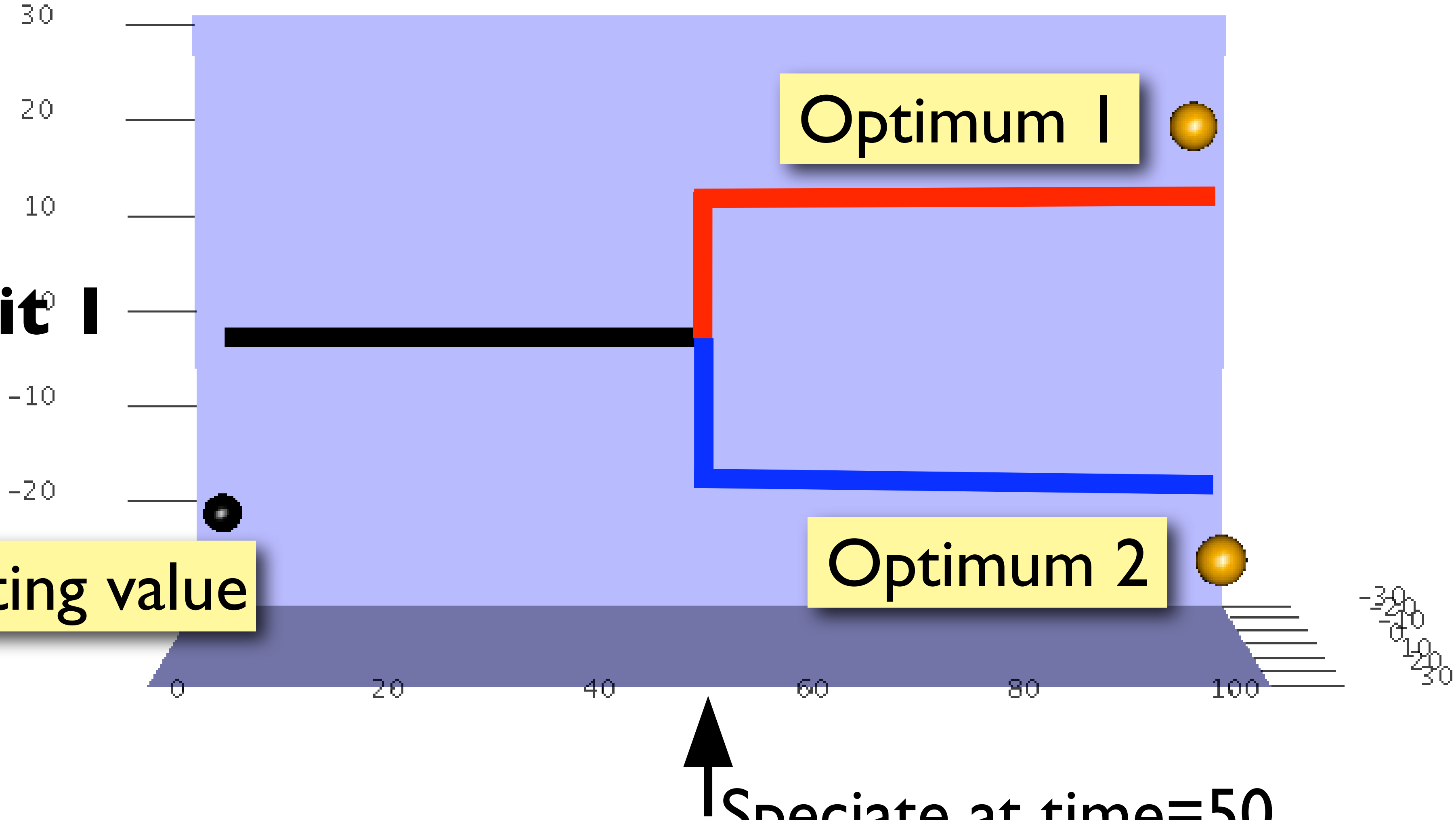
Brownian motion in 2D



Trait 2

Time

Orstein-Uhlenbeck with strong selection



Optimum 1

Optimum 2

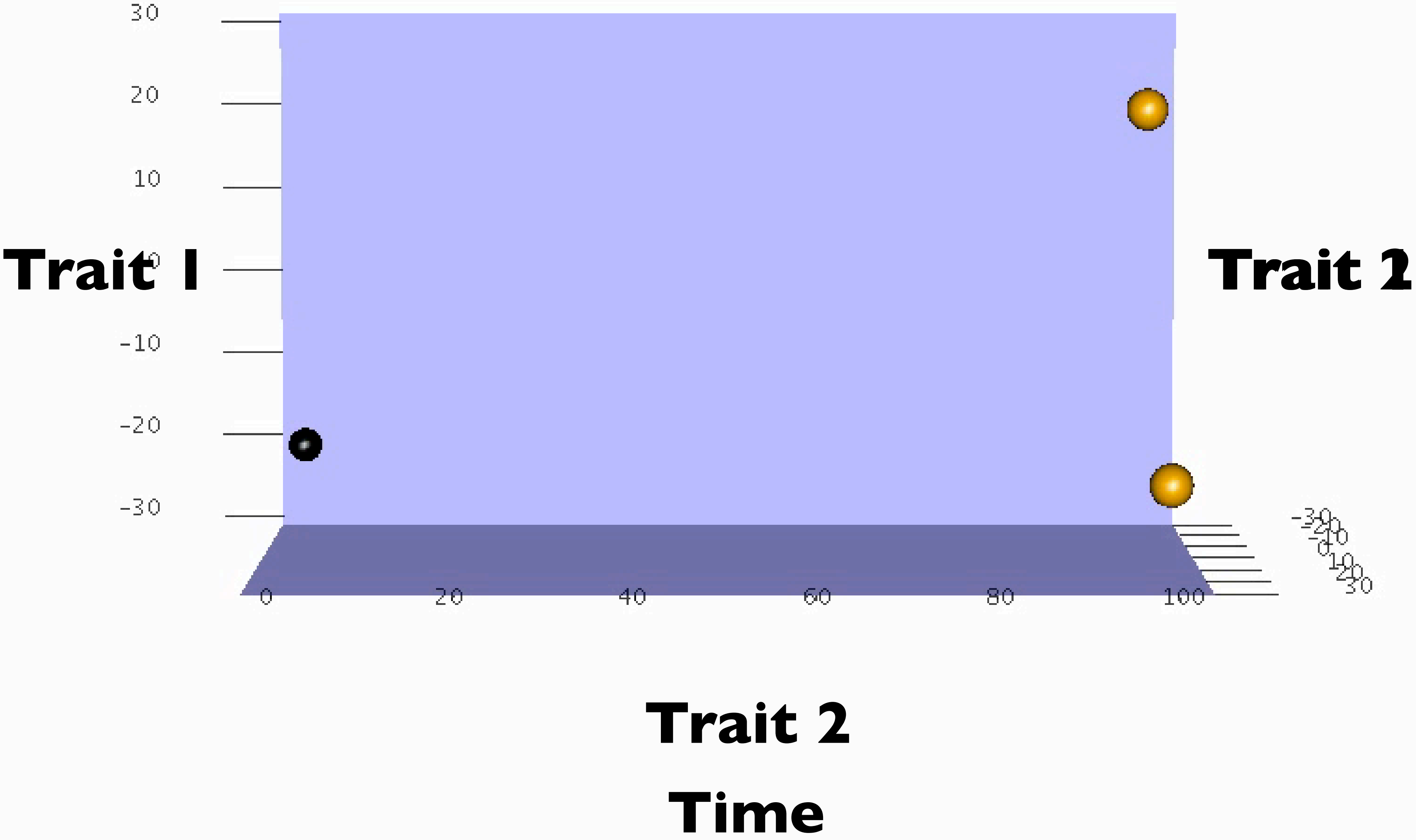
Trait 1

Starting value

Speciate at time=50

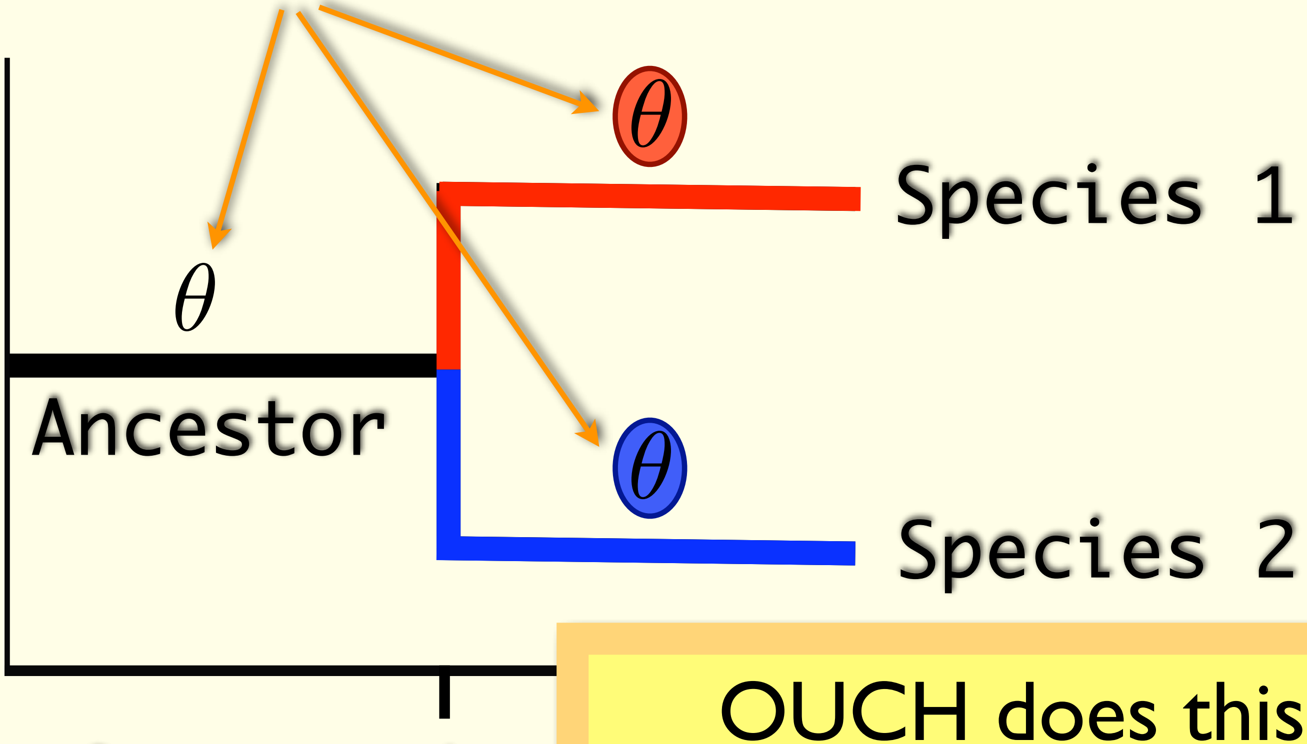
Time

Orstein-Uhlenbeck with strong selection



Entering new adaptive regimes: as easy as a new optimum!

$$dX(t) = \alpha\,(\theta - X(t))\,dt + \sigma\,dB(t)$$

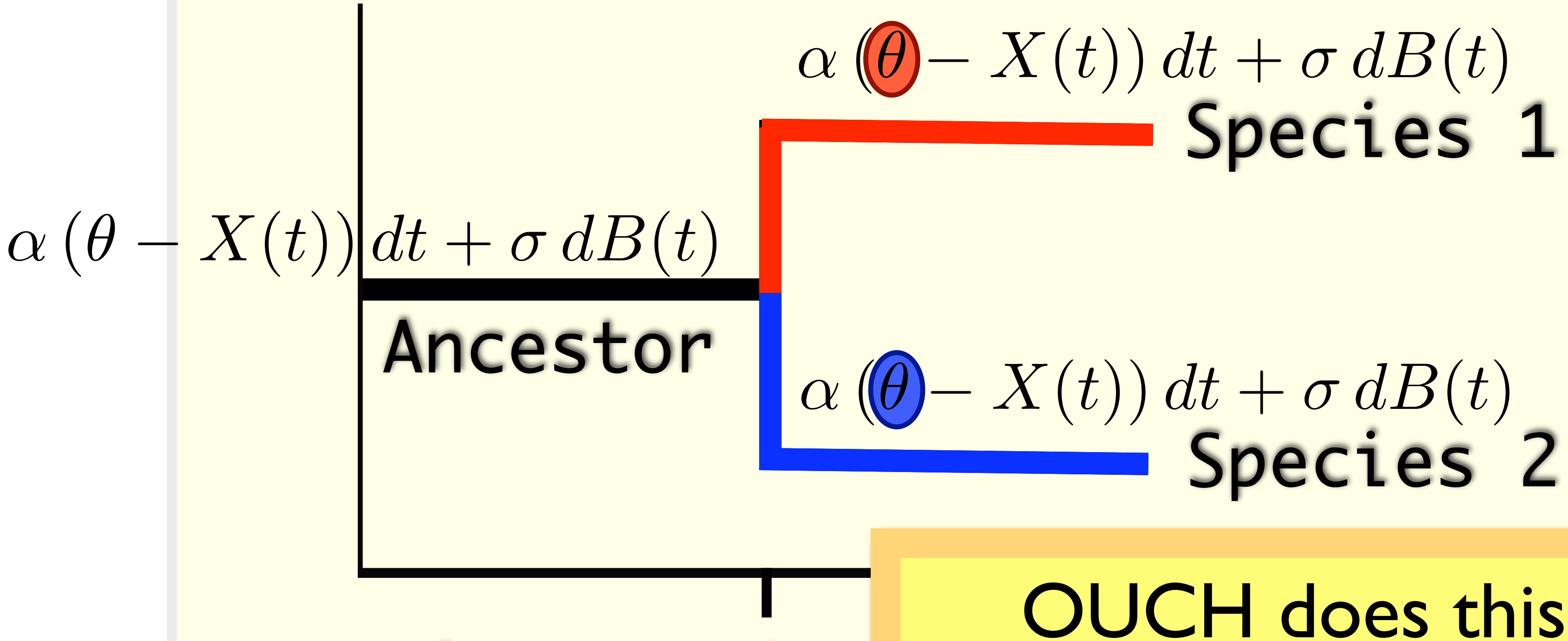


Speciate and enter different adaptive regimes at time t=0.5

OUCH does this by writing a model for each branch of the phylogeny, according to the selective regimes we hypothesize

Entering new adaptive regimes: as easy as a new optimum!

$$dX(t) = \alpha\,(\theta - X(t))\,dt + \sigma\,dB(t)$$



$\alpha\,(\theta - X(t))\,dt + \sigma\,dB(t)$

Species 1

$\alpha\,(\theta - X(t))\,dt + \sigma\,dB(t)$

Ancestor

$\alpha\,(\theta - X(t))\,dt + \sigma\,dB(t)$

Species 2

Speciate and enter different adaptive regimes at time t=0.5

OUCH does this by writing a model for each branch of the phylogeny, according to the selective regimes we hypothesize

With two traits,

$$dX(t) = \alpha\left(\theta - X(t)\right)dt + \sigma\, dB(t)$$

The alpha and sigma parameters become matrices

Superlinear increase in parameters with more traits

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix} \qquad \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

With diagonal terms being selection and drift as usual on trait 1 and trait 2

And off-diagonal terms for correlational effects (the effect of trait 1 on trait 2, and vice versa)

How complex a bivariate model?

“Double Univariate”

$$\alpha = \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \quad \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$$

Correlated “drift”

$$\alpha = \begin{pmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \end{pmatrix} \quad \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

Correlated “selection” + “drift”

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix} \quad \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$

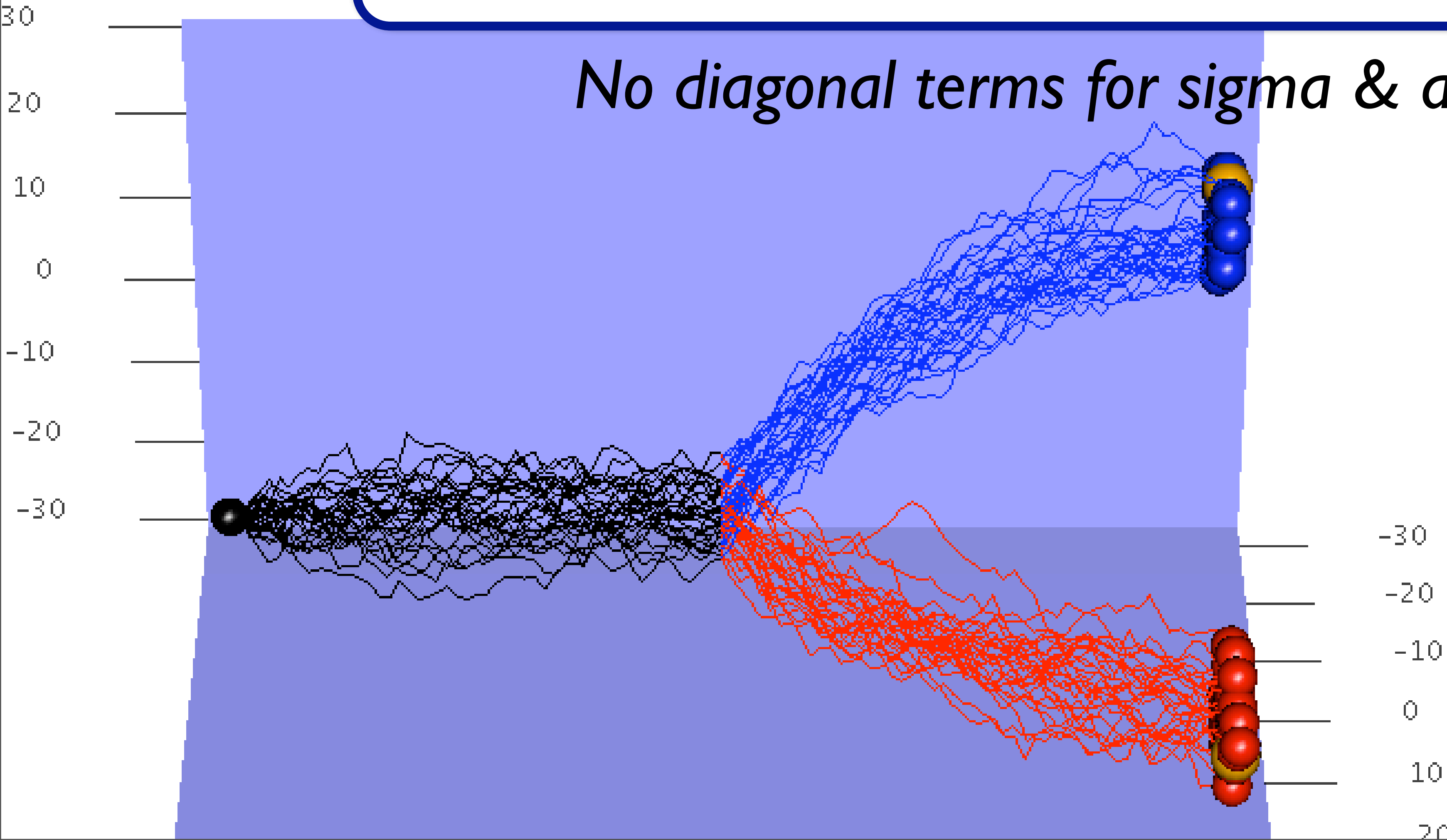
add off-diagonal sigma and alpha terms

How complex a bivariate model?

“Double Univariate”

$$\alpha = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

No diagonal terms for sigma & alpha

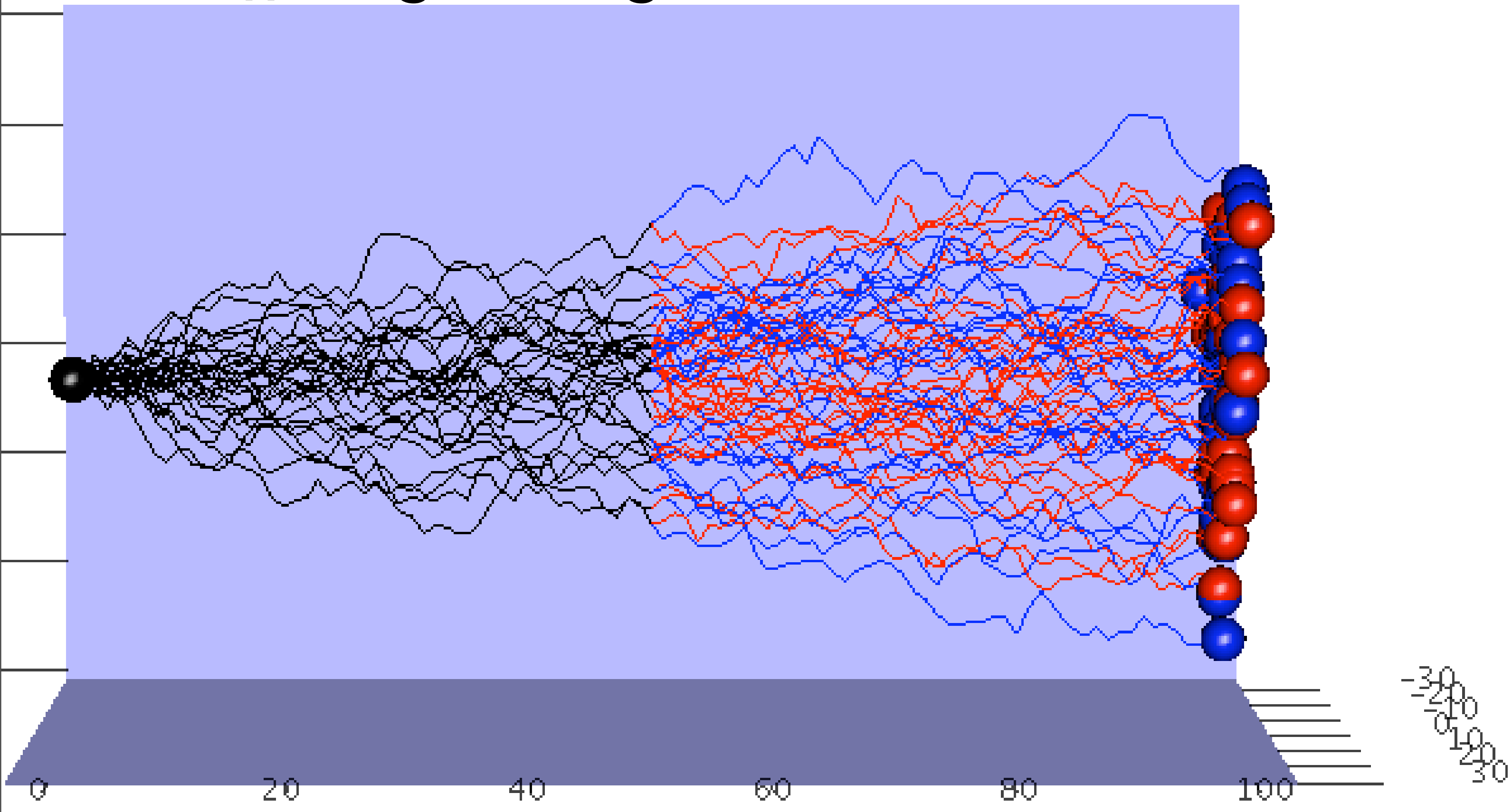


How complex a bivariate model?

Correlated
“drift”

$$\alpha = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 5 & 10 \end{pmatrix}$$

add off-diagonal sigma term

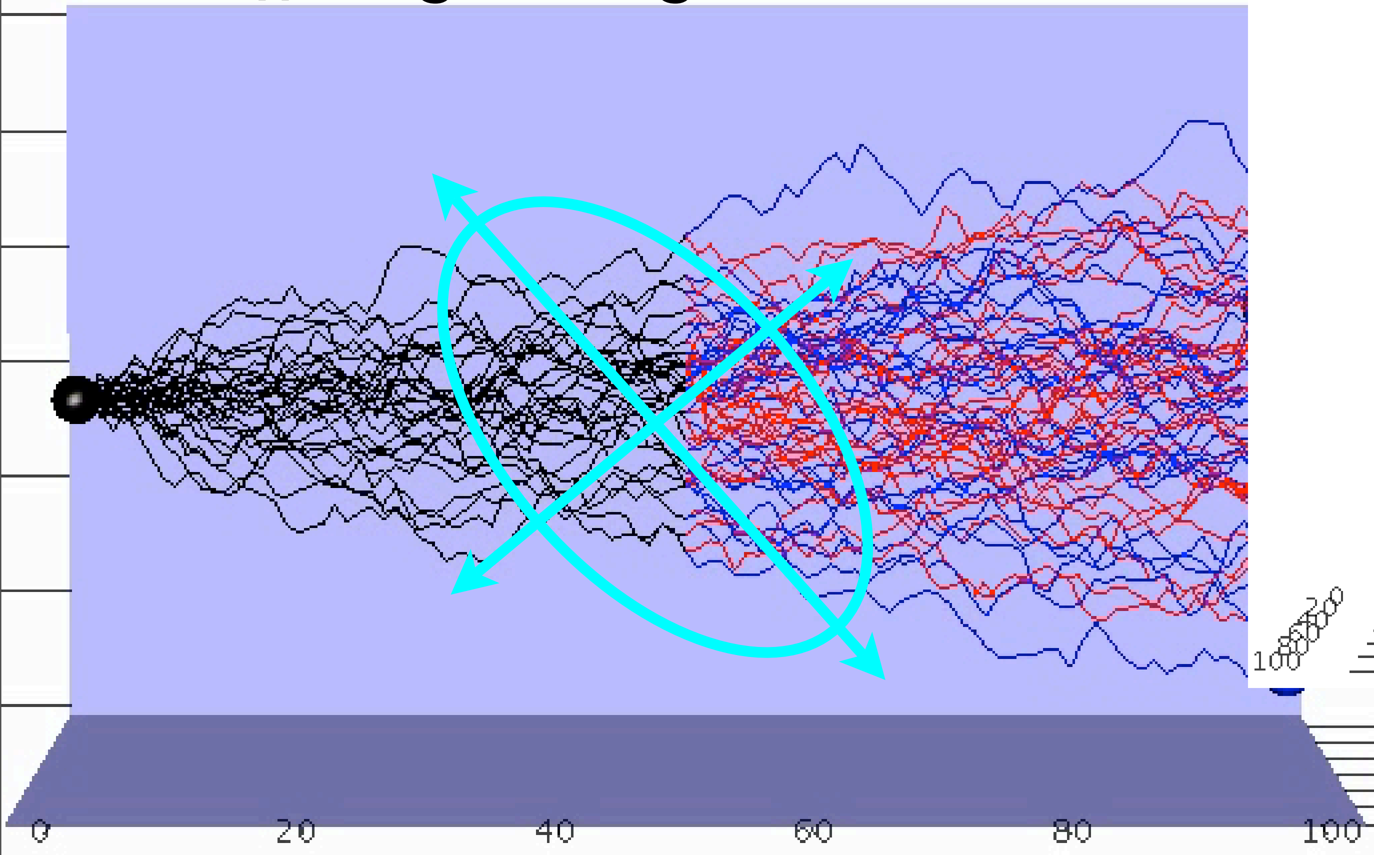


How complex a bivariate model?

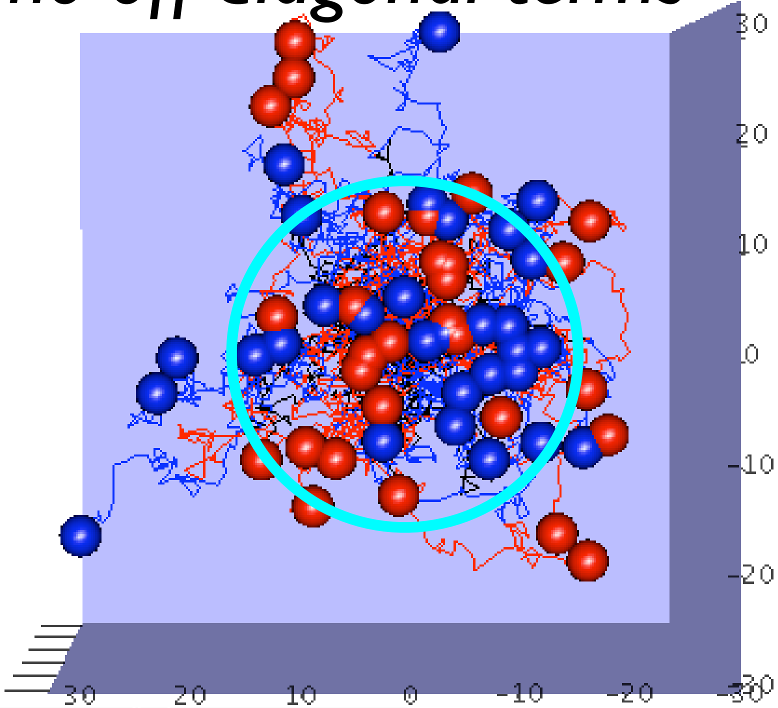
Correlated
“drift”

$$\alpha = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 5 & 10 \end{pmatrix}$$

add off-diagonal sigma term



vs. no off-diagonal terms



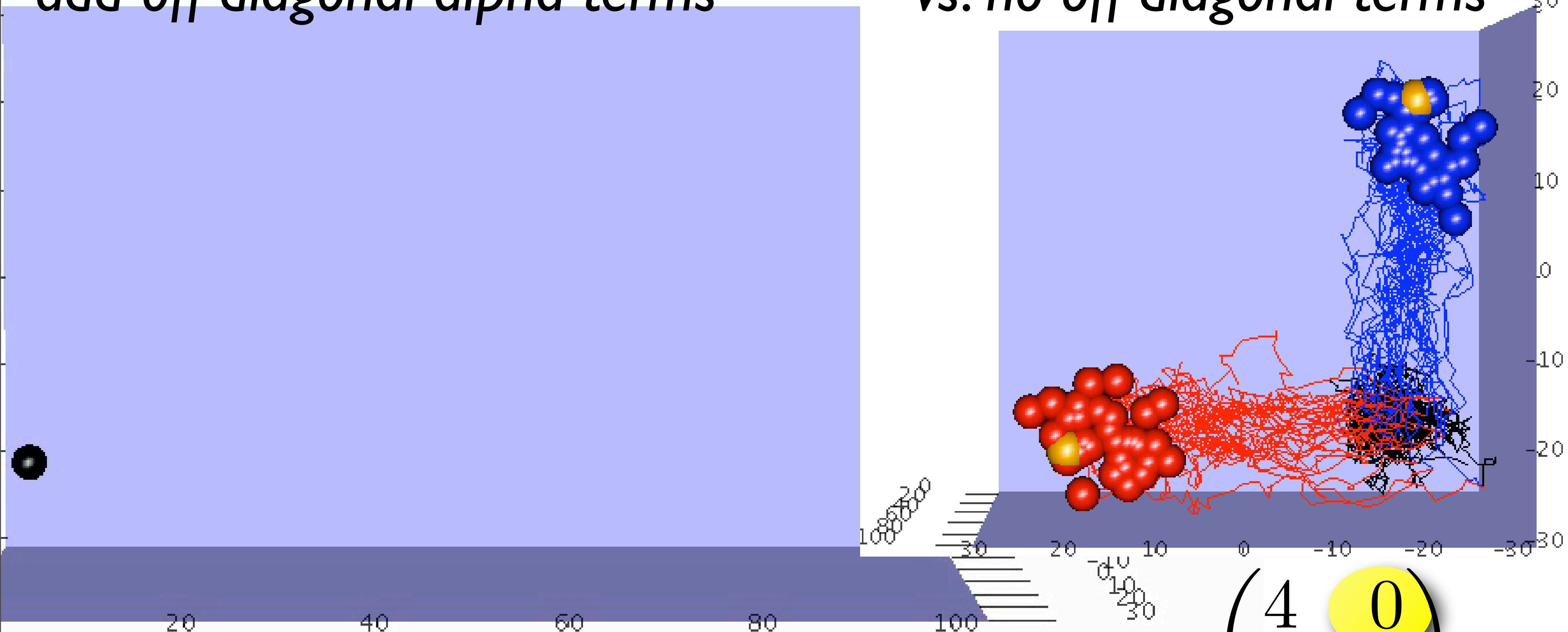
$$\sigma = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

How complex a bivariate model?

Correlated
“selection”

$$\alpha = \begin{pmatrix} 4 & 3 \\ 3 & 4 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

add off-diagonal alpha terms *vs. no off-diagonal terms*



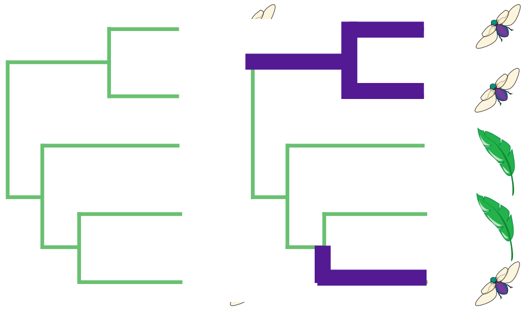
$$\alpha = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

How to apply OUCH?

☐ Data:

- ✱ Quantitative characters
- ☐ Phylogeny and branch lengths (from independent data)

☐ Construct alternative hypotheses (**paint** selective regimes on the tree, selection, drift)



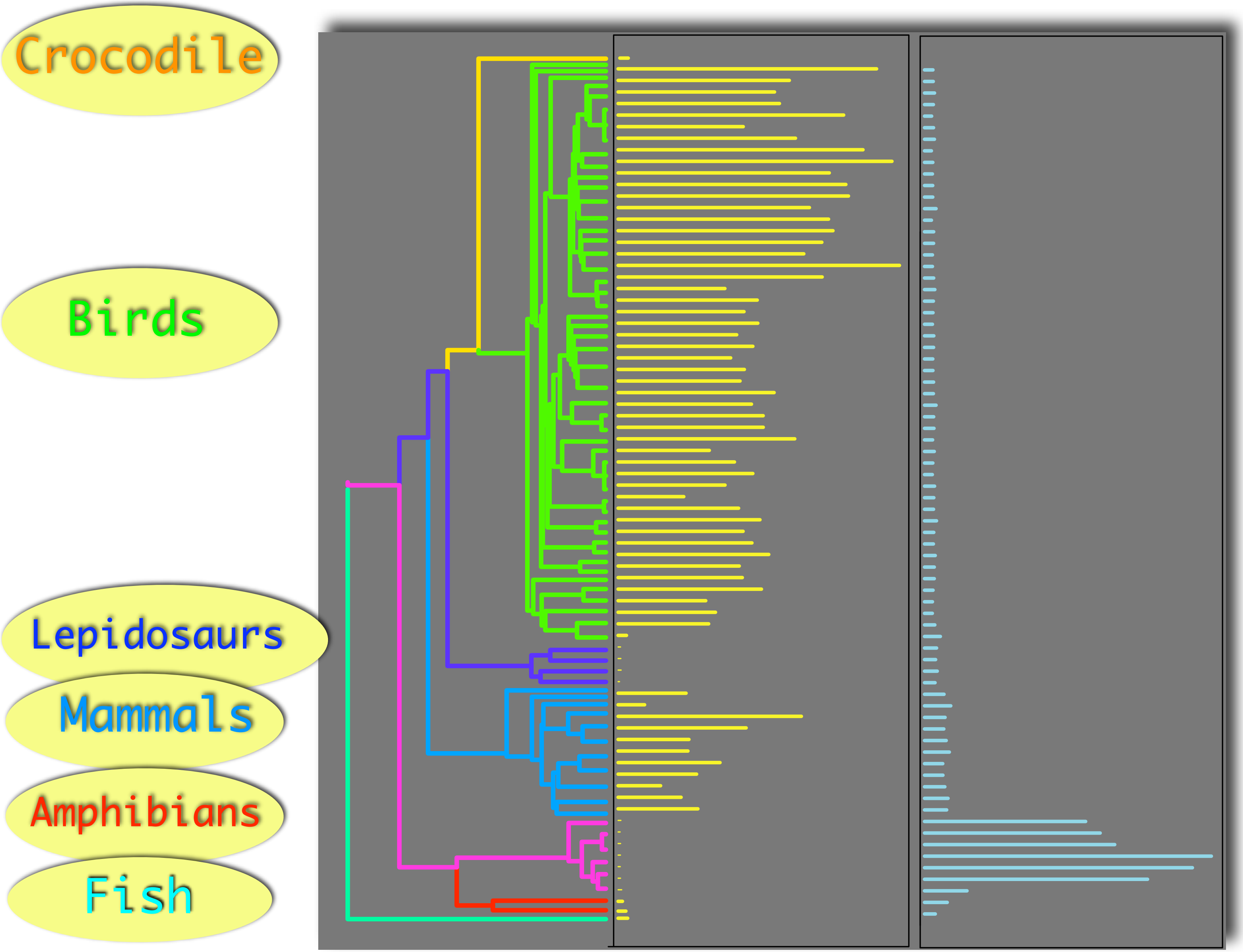
☐ Run OUCH to fit data to alternative models (BM, OU)

☐ Compare models using Likelihood and Information Criteria

Back to our example

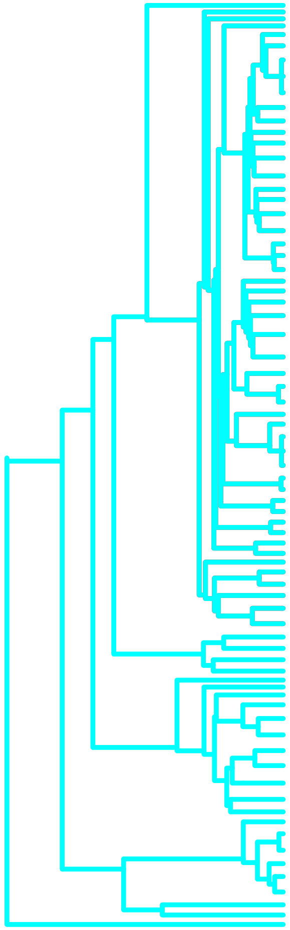
Basal Metabolic Rate and Genome Size

Waltari & Edwards 2002



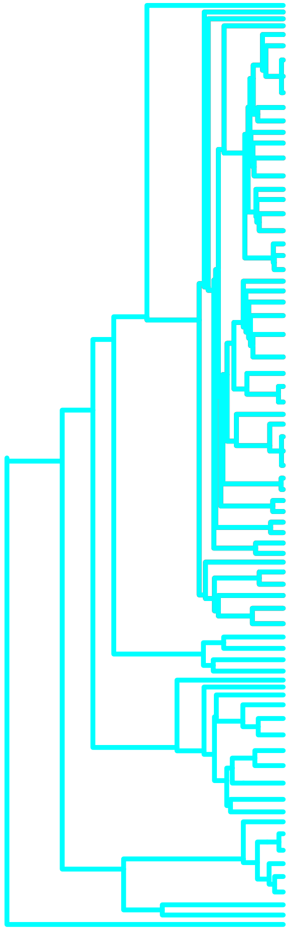
Alternative Adaptive Regime Models

BM

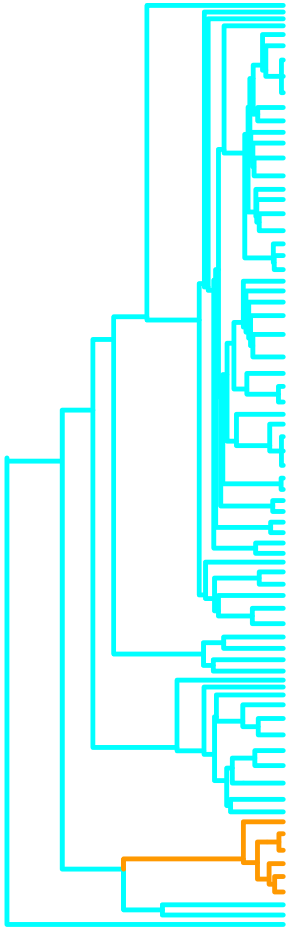


Alternative Adaptive Regime Models

BM

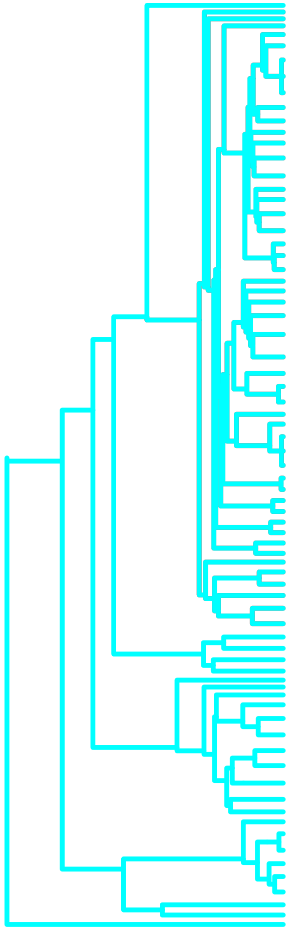


Salamanders
Vertebrates

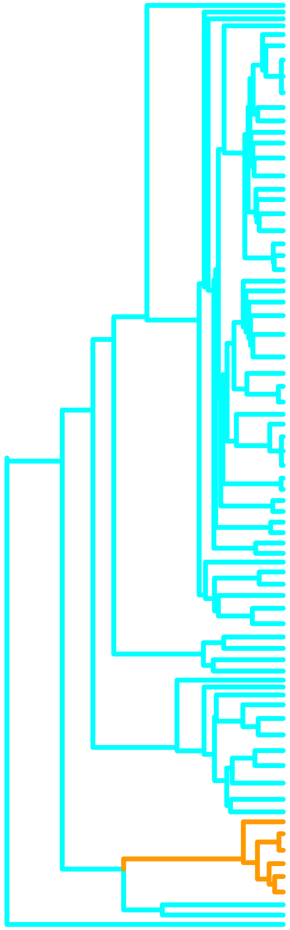


Alternative Adaptive Regime Models

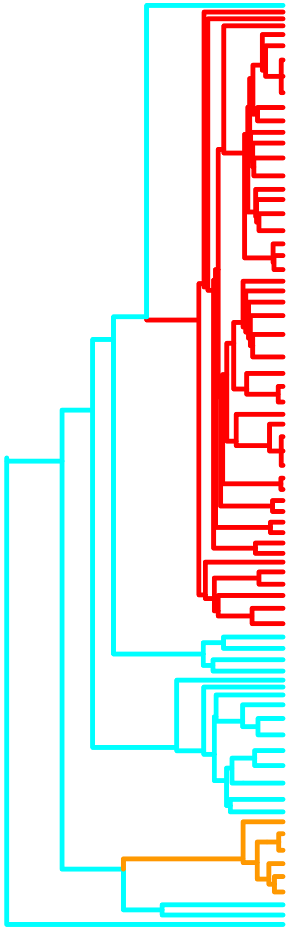
BM



Salamanders
Vertebrates



Birds
Salamander
Vertebrate

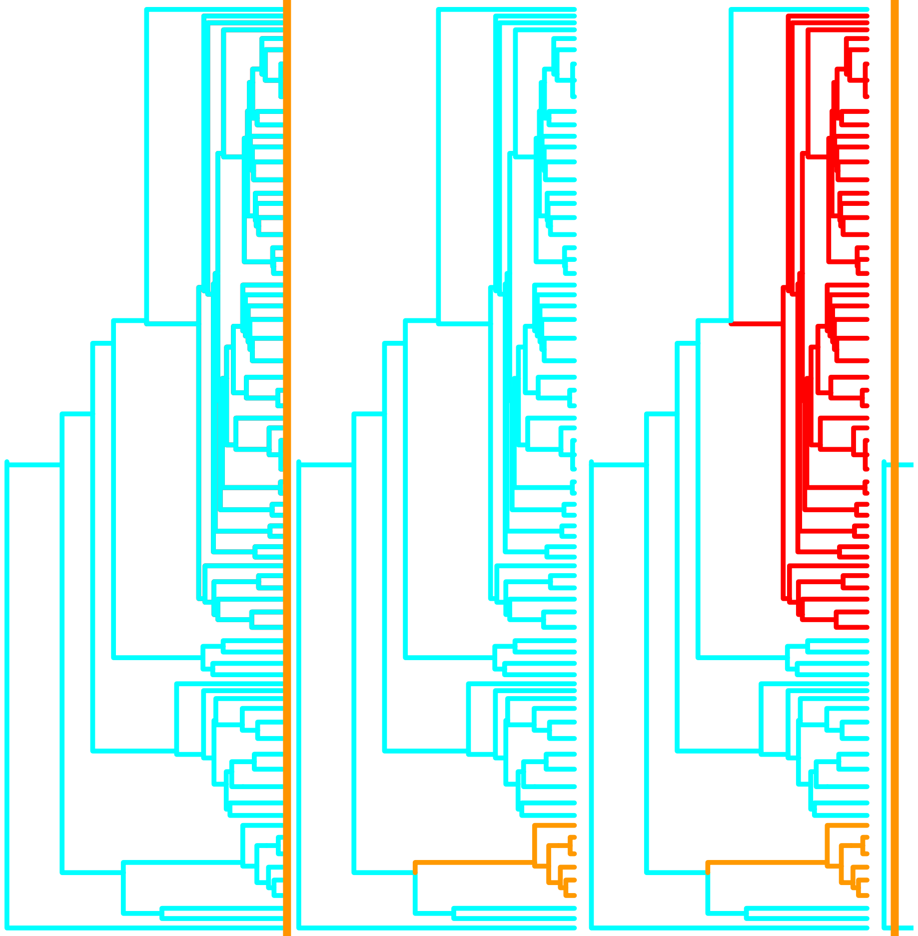


Alternative Adaptive Regime Models

BM

Salamanders
Vertebrates

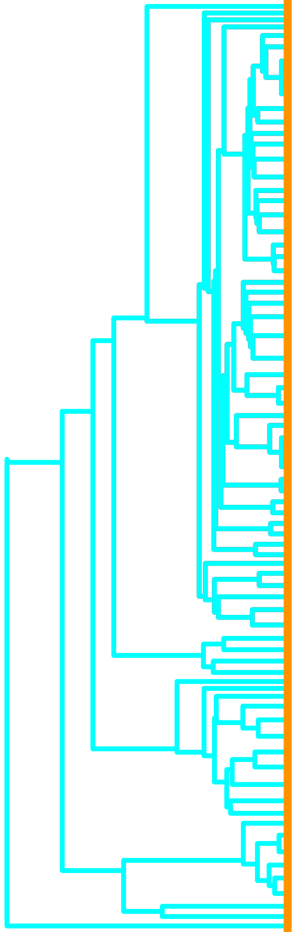
Birds
Salamanders
Vertebrates



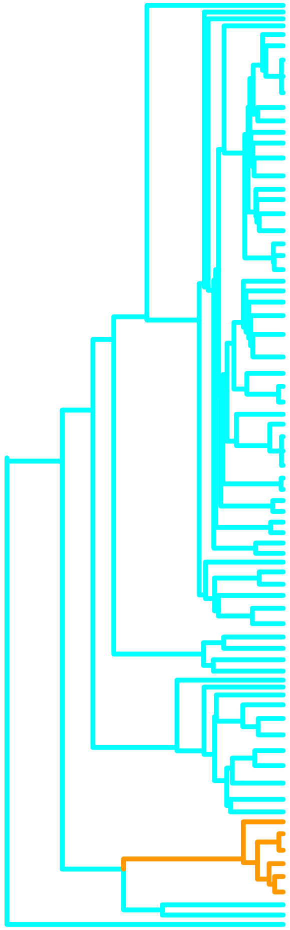
Genome Size Hypotheses

Alternative Adaptive Regime Models

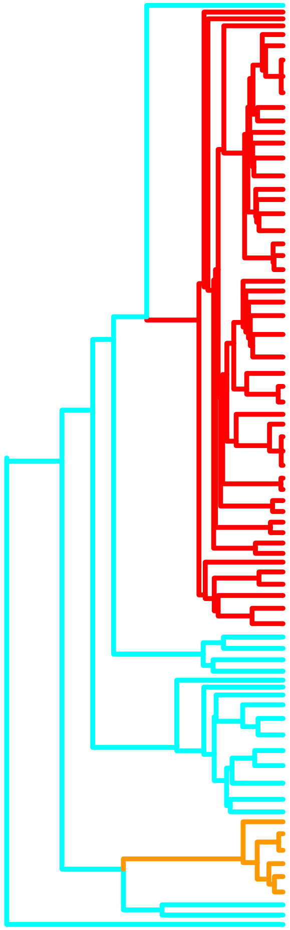
BM



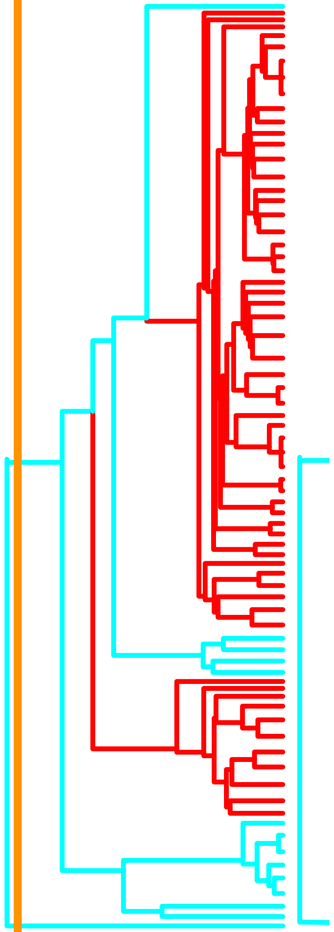
Salamanders
Vertebrates



Birds
Salamanders
Vertebrates

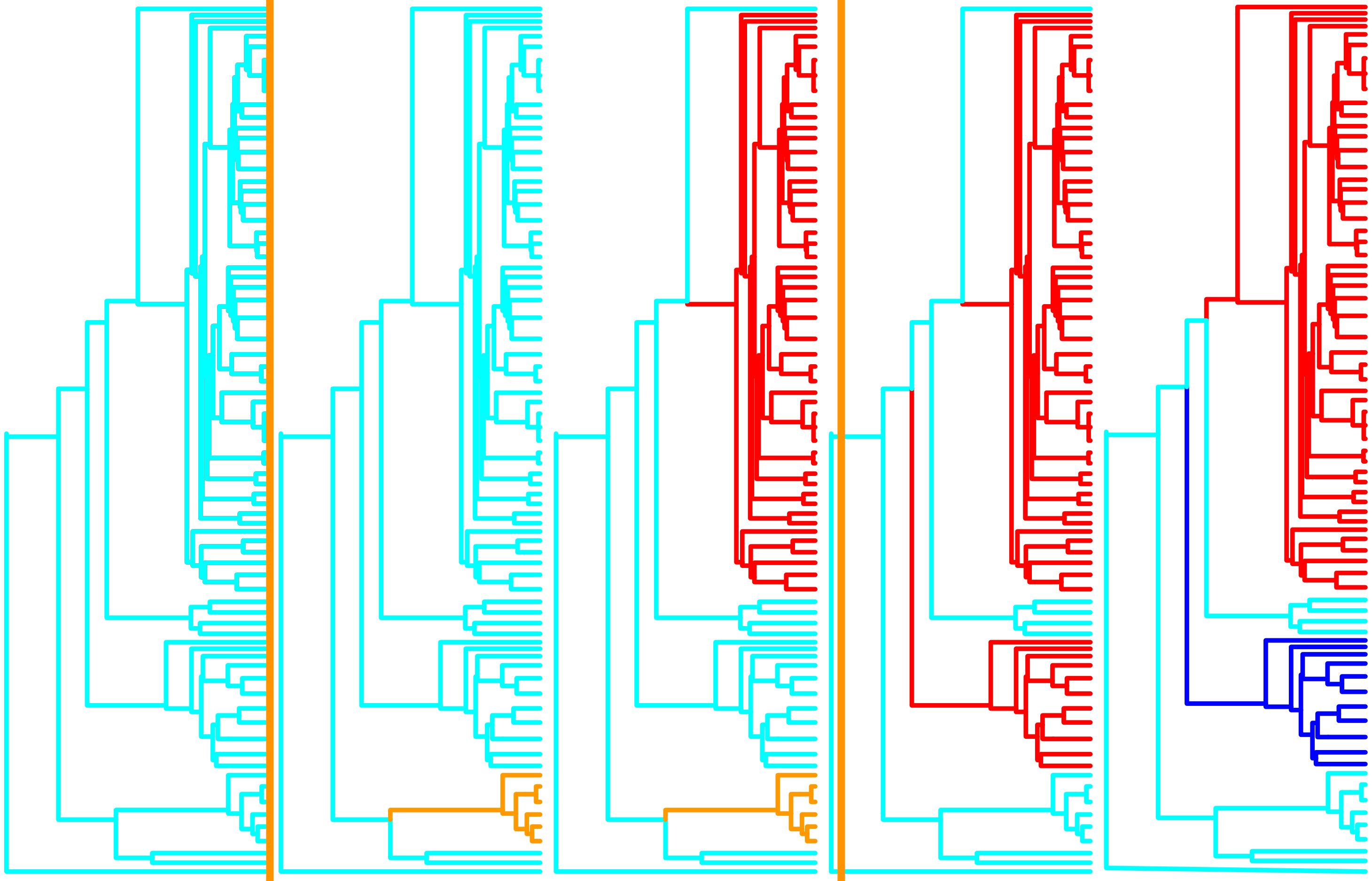


Endotherms
Ectotherms



Genome Size Hypotheses

Alternative Adaptive Regime Models



BM

Salamanders
Vertebrates

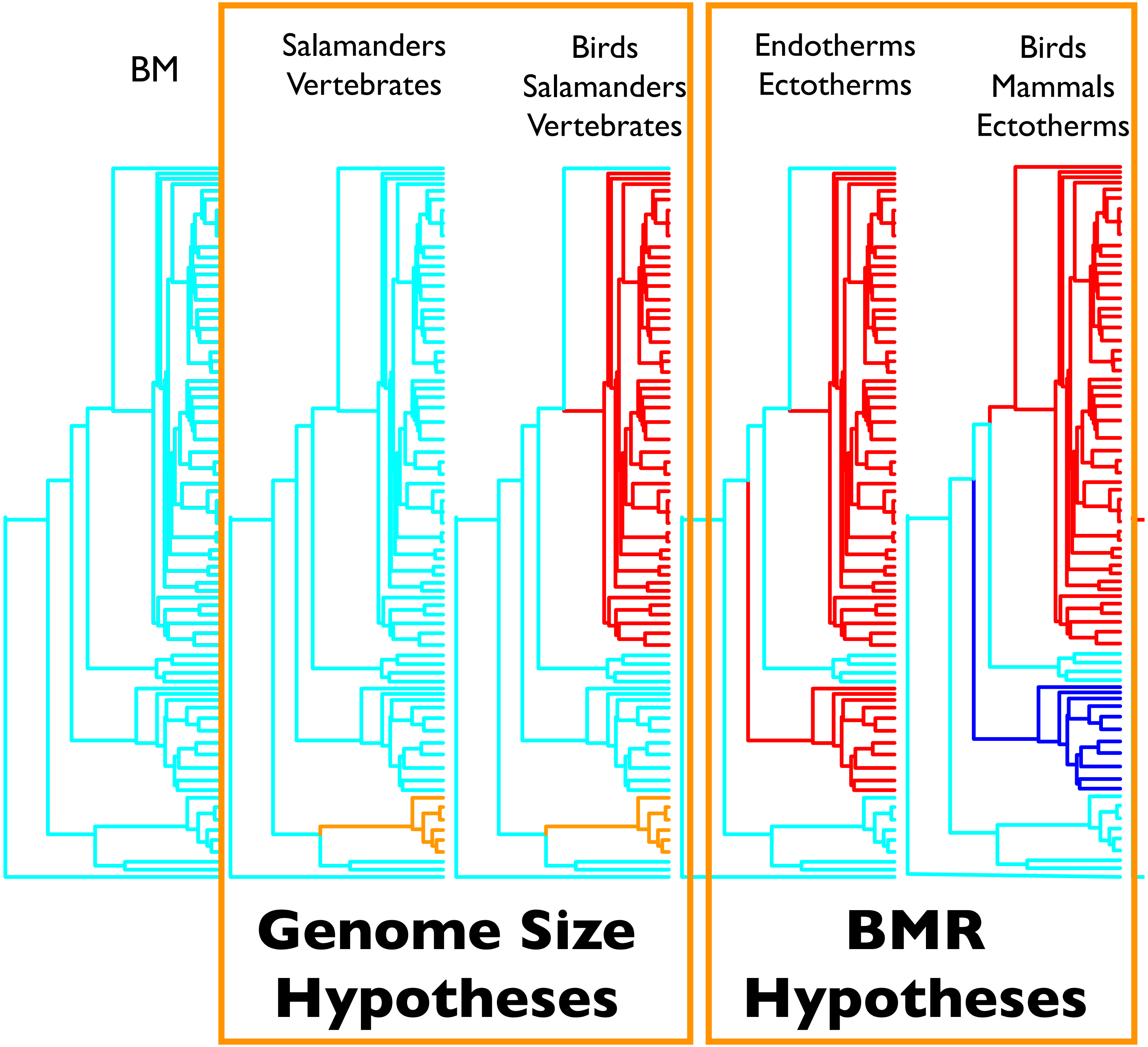
Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

Birds
Mammals
Ectotherms

Genome Size Hypotheses

Alternative Adaptive Regime Models



BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

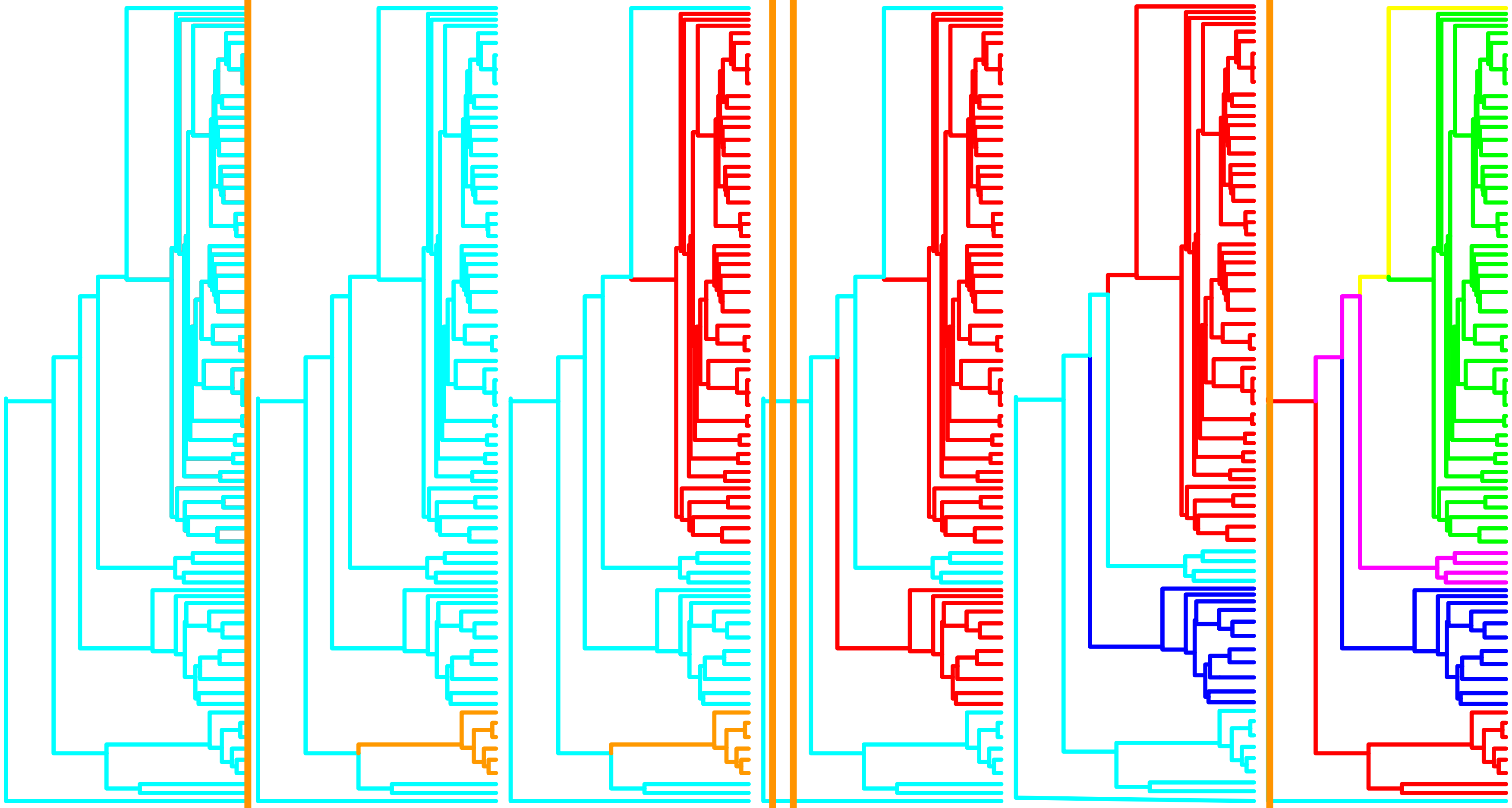
Endotherms
Ectotherms

Birds
Mammals
Ectotherms

Genome Size Hypotheses

BMR Hypotheses

Alternative Adaptive Regime Models



BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

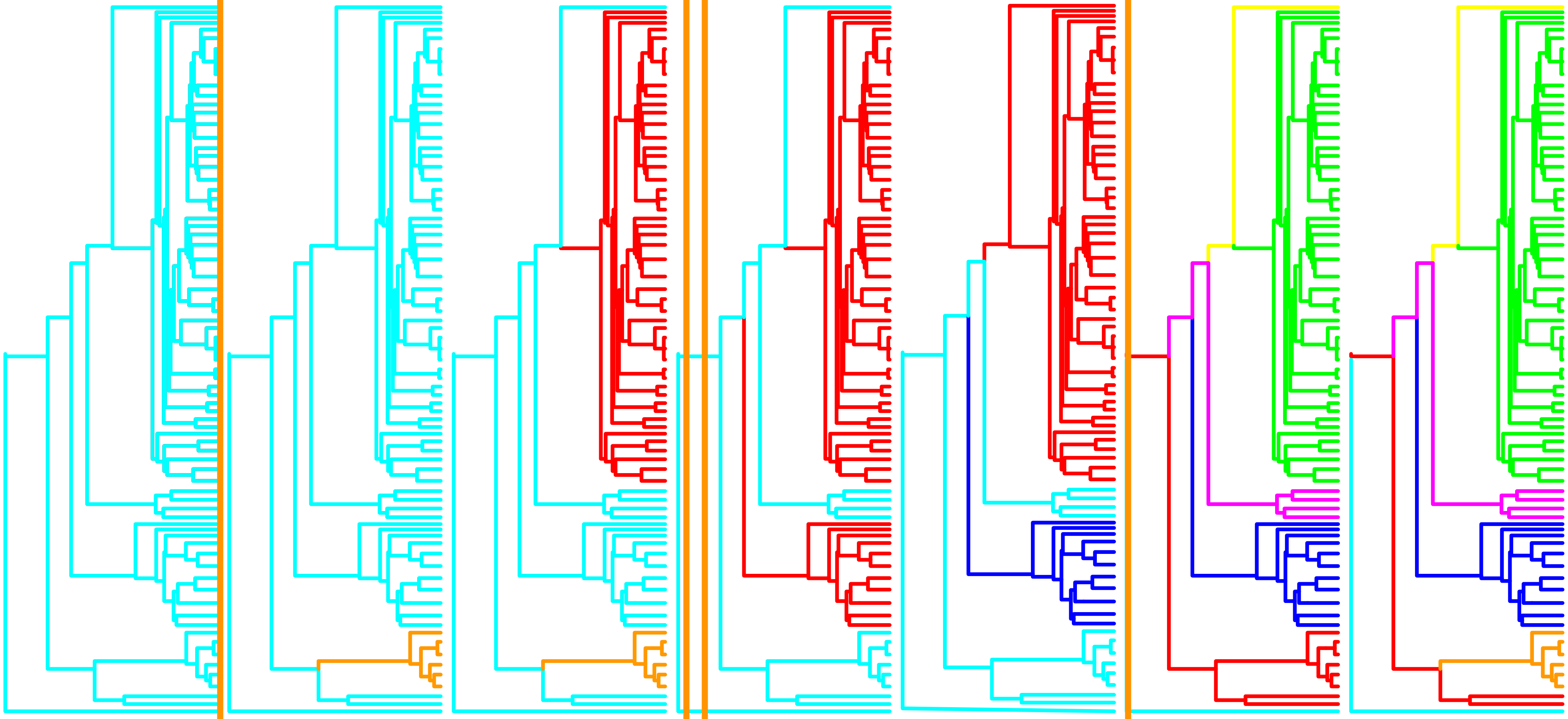
Birds
Mammals
Ectotherms

Major
Taxa

**Genome Size
Hypotheses**

**BMR
Hypotheses**

Alternative Adaptive Regime Models



BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

Birds
Mammals
Ectotherms

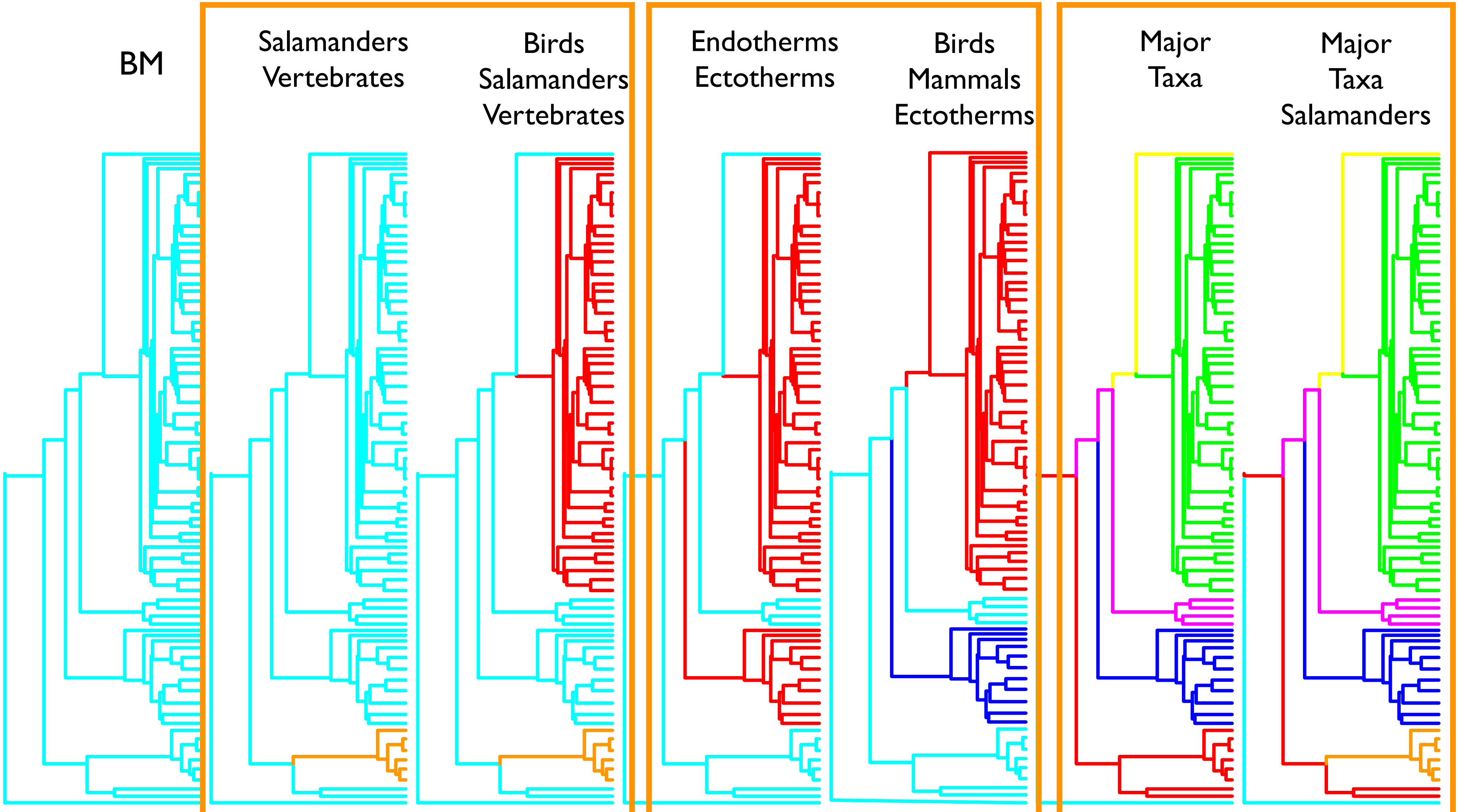
Major
Taxa

Major
Taxa
Salamanders

Genome Size Hypotheses

BMR Hypotheses

Alternative Adaptive Regime Models



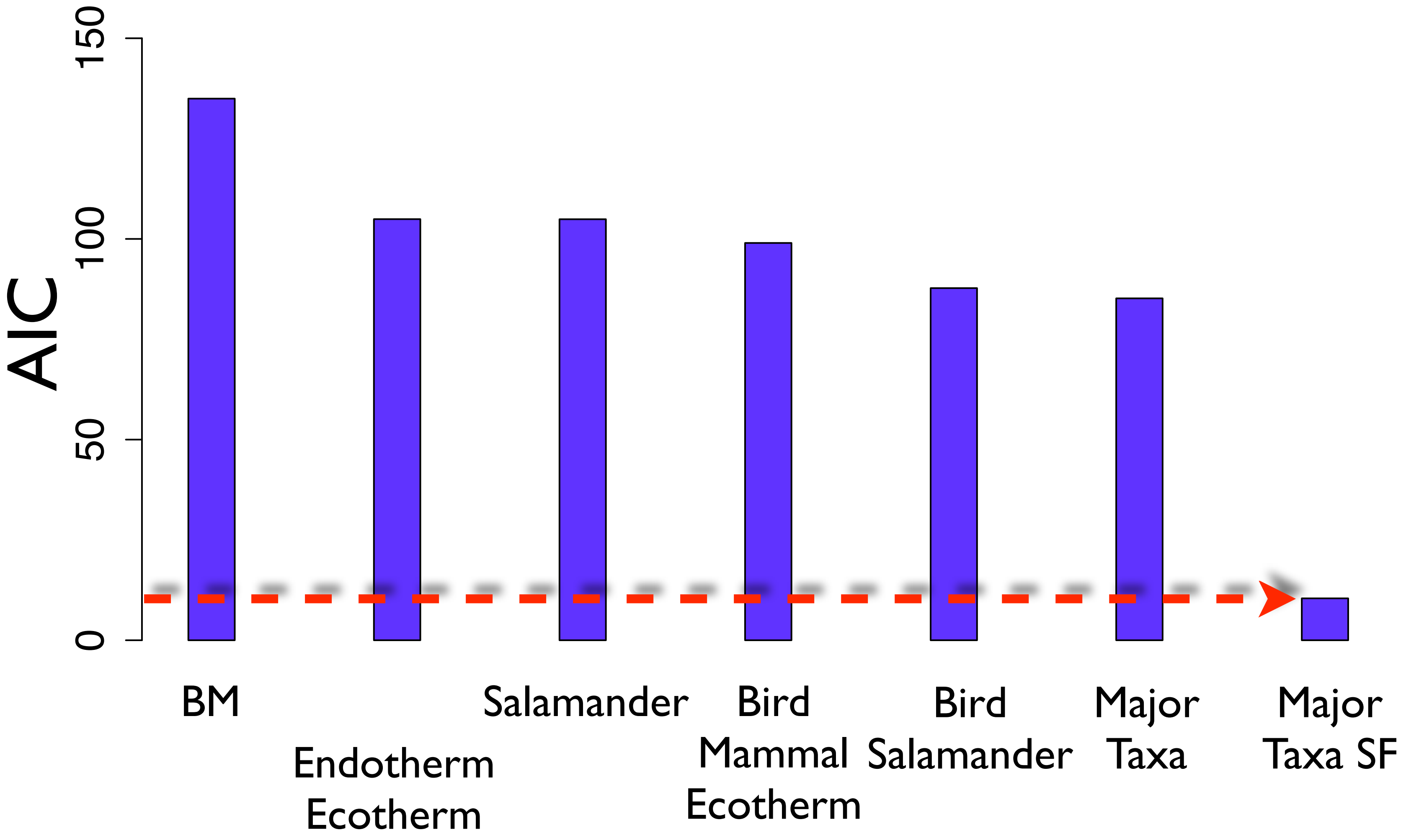
Genome Size Hypotheses

BMR Hypotheses

Major Taxa Hypotheses

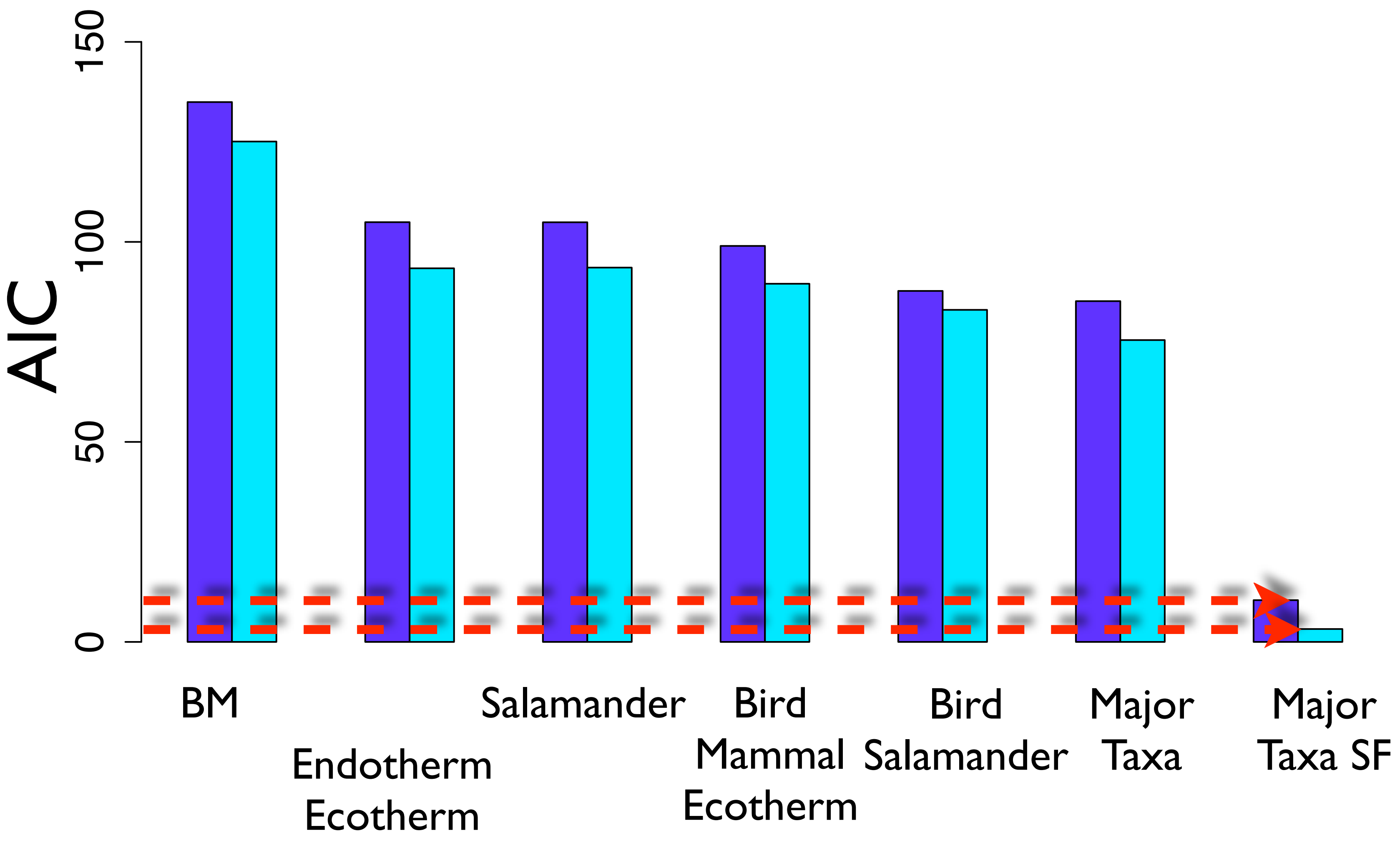
Separate adaptive regimes for each Major Taxon (splitting salamanders & frogs) was the Best Model

simplest bivariate model: no off-diagonal terms



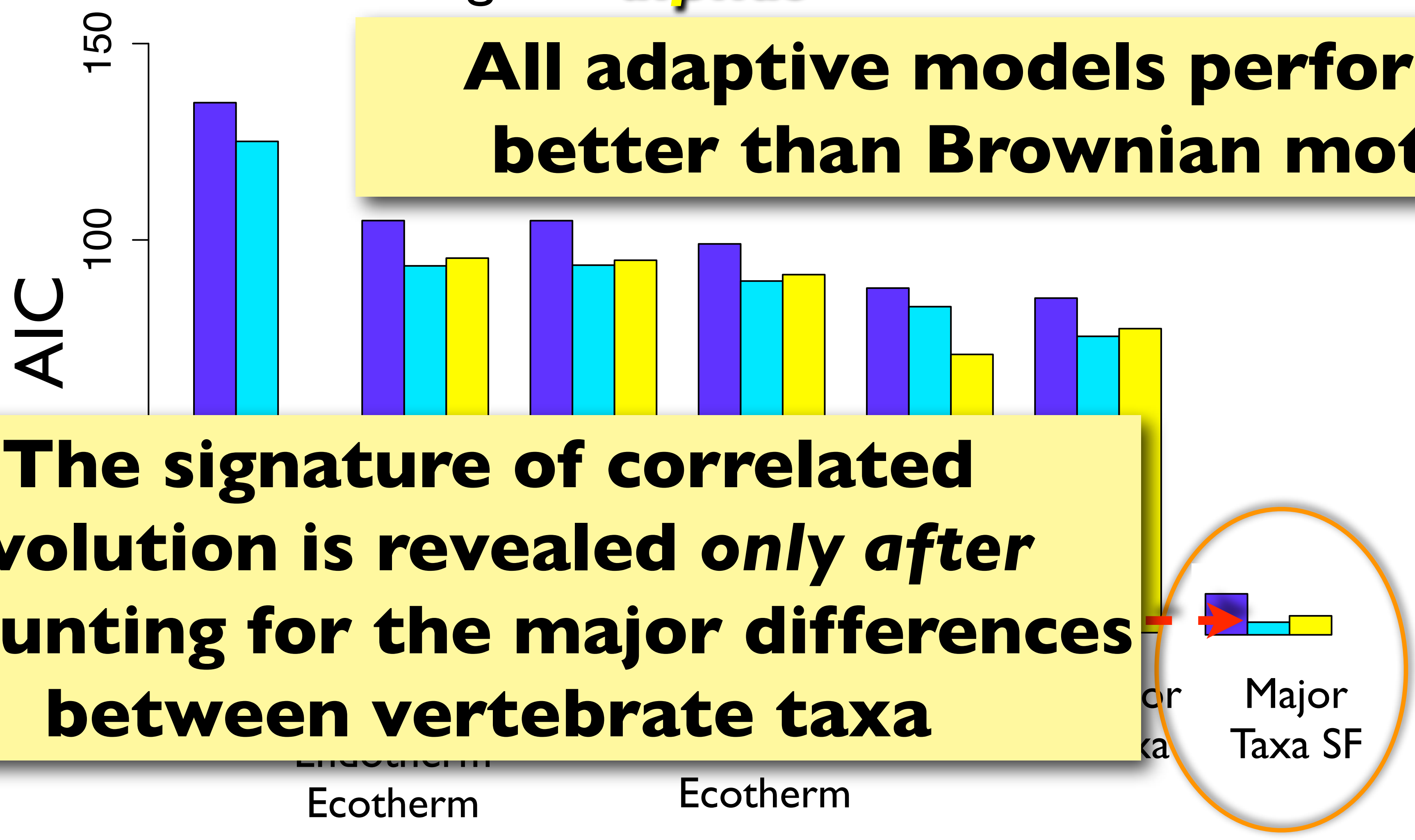
Separate adaptive regimes for each Major Taxon (splitting salamanders & frogs) was the Best Model

correlated drift: off-diagonal ***sigmas*** improve fit



Separate adaptive regimes for each Major Taxon (splitting salamanders & frogs) was the Best Model

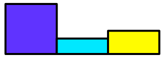
correlated drift: off-diagonal **sigmas** improve fit
but off-diagonal **alphas** do not



All adaptive models performed better than Brownian motion

The signature of correlated evolution is revealed *only after* accounting for the major differences between vertebrate taxa

Results



Major
Taxa SF

Double
Univariate

$$\alpha = \begin{pmatrix} 38.1 & 0 \\ 0 & 4.7 \end{pmatrix} \quad \sigma = \begin{pmatrix} 1.2 & 0 \\ 0 & 1.0 \end{pmatrix}$$

Very strong selection, especially on Genome Size. Moderate drift.

Results

Major
Taxa SF

Off-diagonal sigmas improved the model

Double Univariate

$$\alpha = \begin{pmatrix} 38.1 & 0 \\ 0 & 4.7 \end{pmatrix} \quad \sigma = \begin{pmatrix} 1.2 & 0 \\ 0 & 1.0 \end{pmatrix}$$

Correlated "drift"

$$\alpha = \begin{pmatrix} 30.8 & 0 \\ 0 & 4.8 \end{pmatrix} \quad \sigma = \begin{pmatrix} 1.1 & 0 \\ -0.44 & 0.91 \end{pmatrix}$$

Correlated "selection" + "drift"

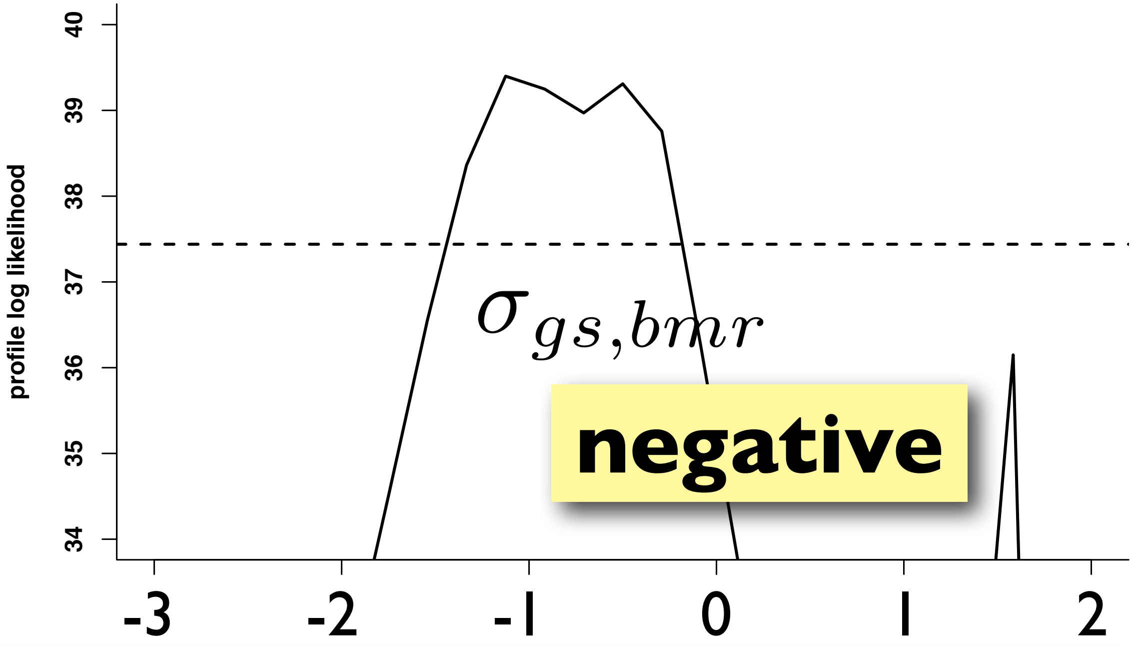
$$\alpha = \begin{pmatrix} 33.5 & 1.5 \\ 1.5 & 4.8 \end{pmatrix} \quad \sigma = \begin{pmatrix} 1.1 & 0 \\ -0.35 & 0.93 \end{pmatrix}$$

BUT off-diagonal alphas did not

Major
Taxa SF

Results

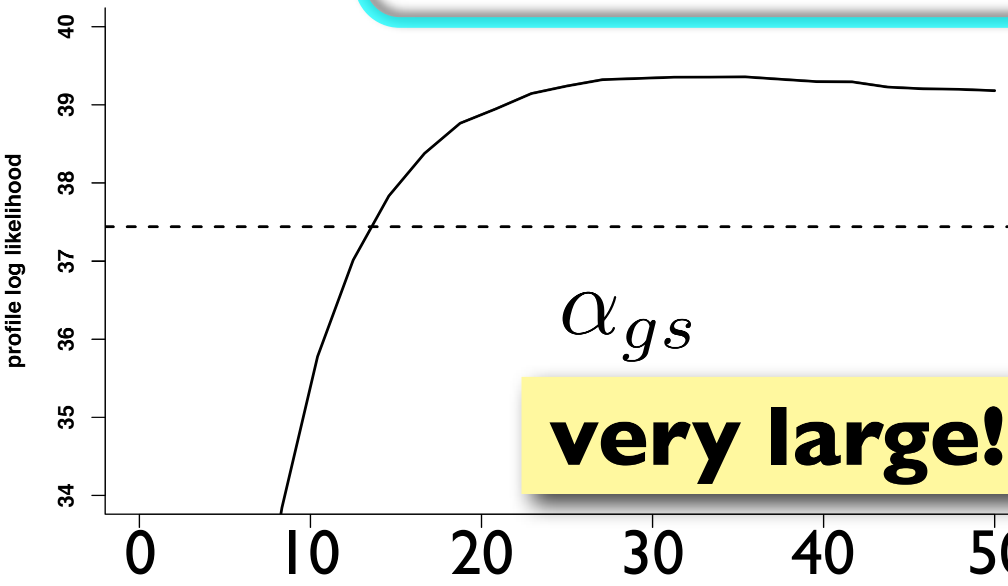
Profile Likelihoods for parameter estimates



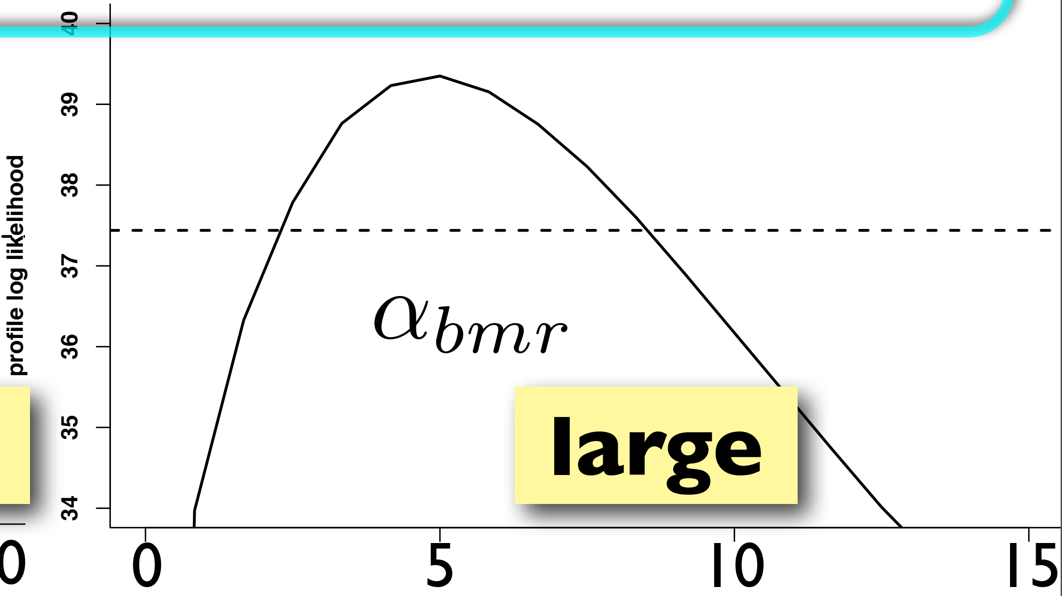
negative

Correlated
"drift"

$$\alpha = \begin{pmatrix}30.8 & 0\\ 0 & 4.8\end{pmatrix} \quad \sigma = \begin{pmatrix}1.1 & 0\\ -0.44 & 0.91\end{pmatrix}$$

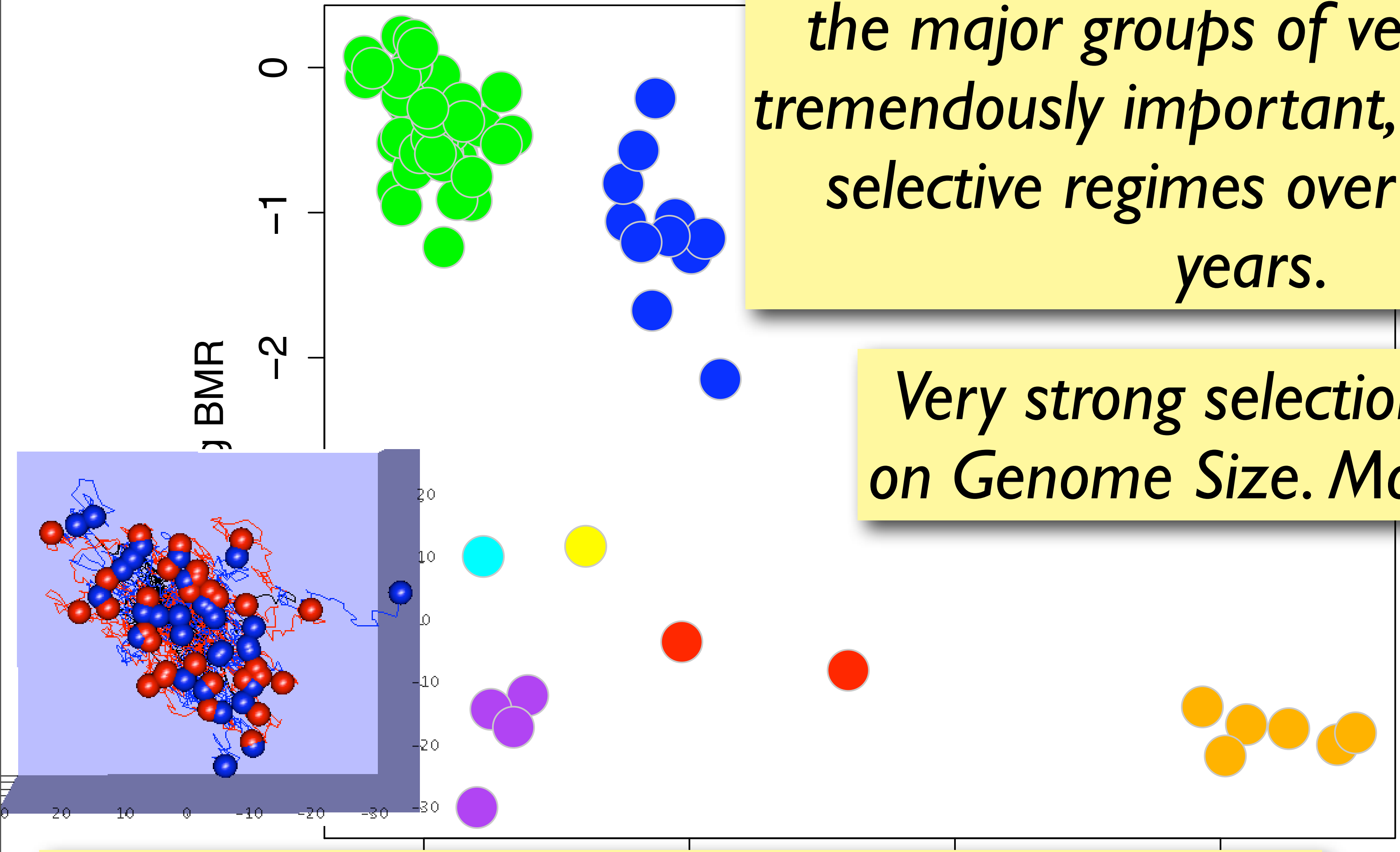


very large!



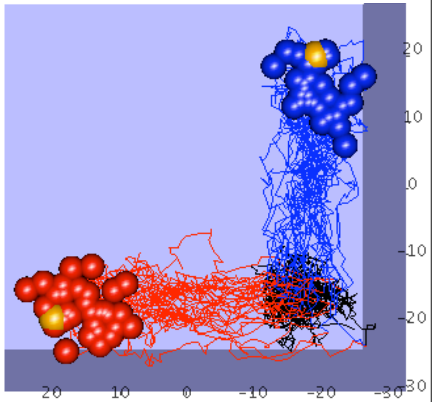
large

So… how to explain these data?



The stamp of history. The biology of the major groups of vertebrates is tremendously important, shaping their selective regimes over millions of years.

Very strong selection, especially on Genome Size. Moderate drift.



After accounting for major differences among taxa, strong signature of correlated evolution

But this looks like a lot of math.

I can already calculate statistically valid correlations.

What can modeling really do for me?

Only by being explicit can we see which elements of our evolutionary scenario are most important

selective regimes

selection

drift

evolutionary correlation

Remember: Modeling is your friend.

What we are advocating is the method of multiple working hypotheses.

~~e “best” model ?~~

it’s about learning the most you can about your data

Please see Jeff Scales' talk tomorrow:

9:00am Wednesday, Breakout Room 4,5

Adaptation in muscle fiber type composition in locomotor muscles of lizards

Many Thanks...

Eric Waltari and Scott Edwards
Jeff Scales and Butler lab group

NSF

National Science Foundation
Systematic Biology Panel (DEB 0515390 and DEB 0542360)

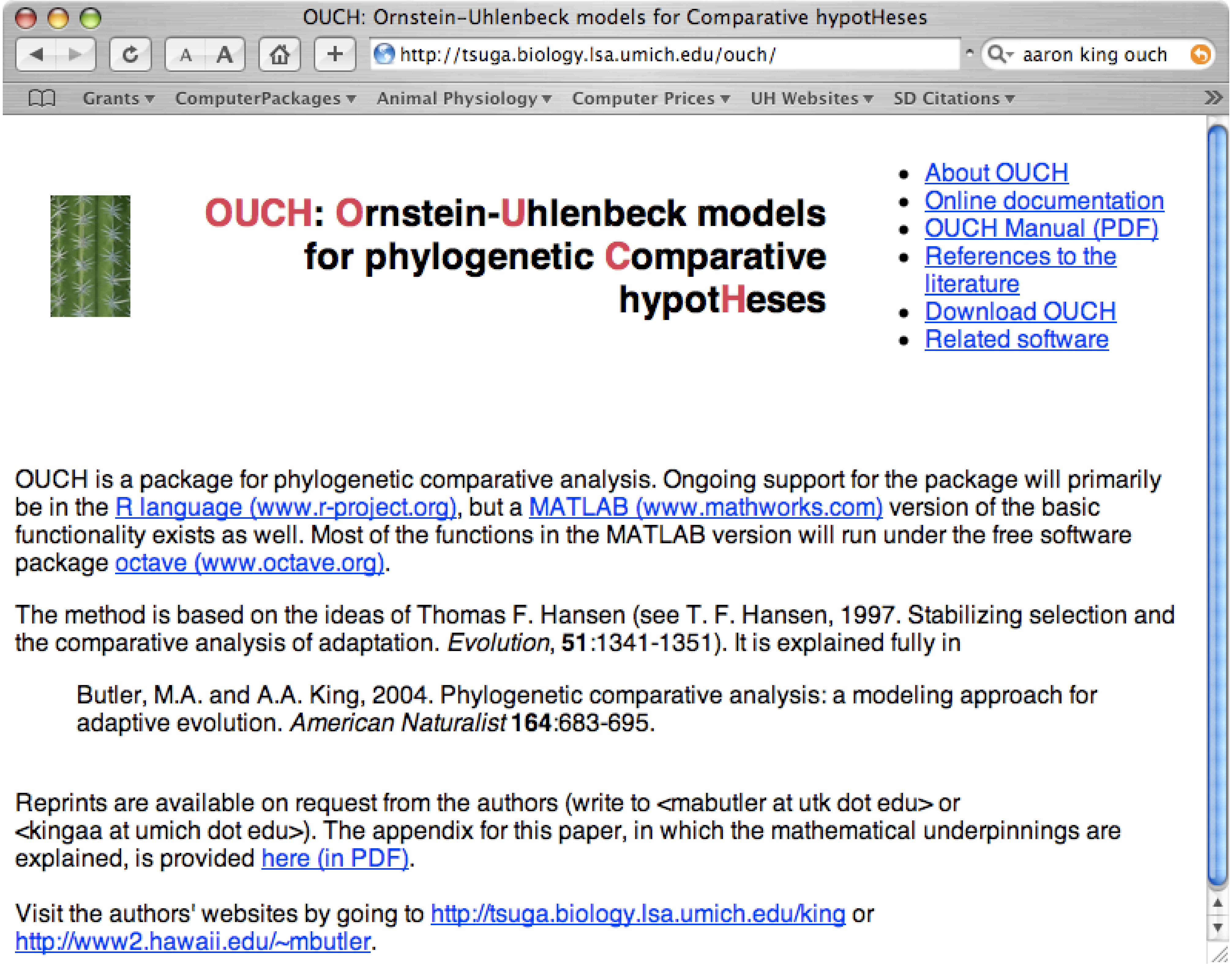
EEB
UNIVERSITY OF MICHIGAN
ECOLOGY AND EVOLUTIONARY BIOLOGY

Department of
ZOOLOGY
UNIVERSITY OF HAWAI'I at MANOA

Seeking postdoc and/or graduate students

Try OUCH (our software) - OUCH 2 coming soon!

http://tsuga.biology.lsa.umich.edu/ouch/



OUCH: Ornstein-Uhlenbeck models for phylogenetic Comparative hypotHeses

- About OUCH
- Online documentation
- OUCH Manual (PDF)
- References to the literature
- Download OUCH
- Related software

OUCH is a package for phylogenetic comparative analysis. Ongoing support for the package will primarily be in the R language (www.r-project.org), but a MATLAB (www.mathworks.com) version of the basic functionality exists as well. Most of the functions in the MATLAB version will run under the free software package octave (www.octave.org).

The method is based on the ideas of Thomas F. Hansen (see T. F. Hansen, 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, **51**:1341-1351). It is explained fully in

> Butler, M.A. and A.A. King, 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *American Naturalist* **164**:683-695.

Reprints are available on request from the authors (write to <mabutler at utk dot edu> or <kingaa at umich dot edu>). The appendix for this paper, in which the mathematical underpinnings are explained, is provided here (in PDF).

Visit the authors' websites by going to http://tsuga.biology.lsa.umich.edu/king or http://www2.hawaii.edu/~mbutler.