Multivariate comparative analysis using OUCH*!
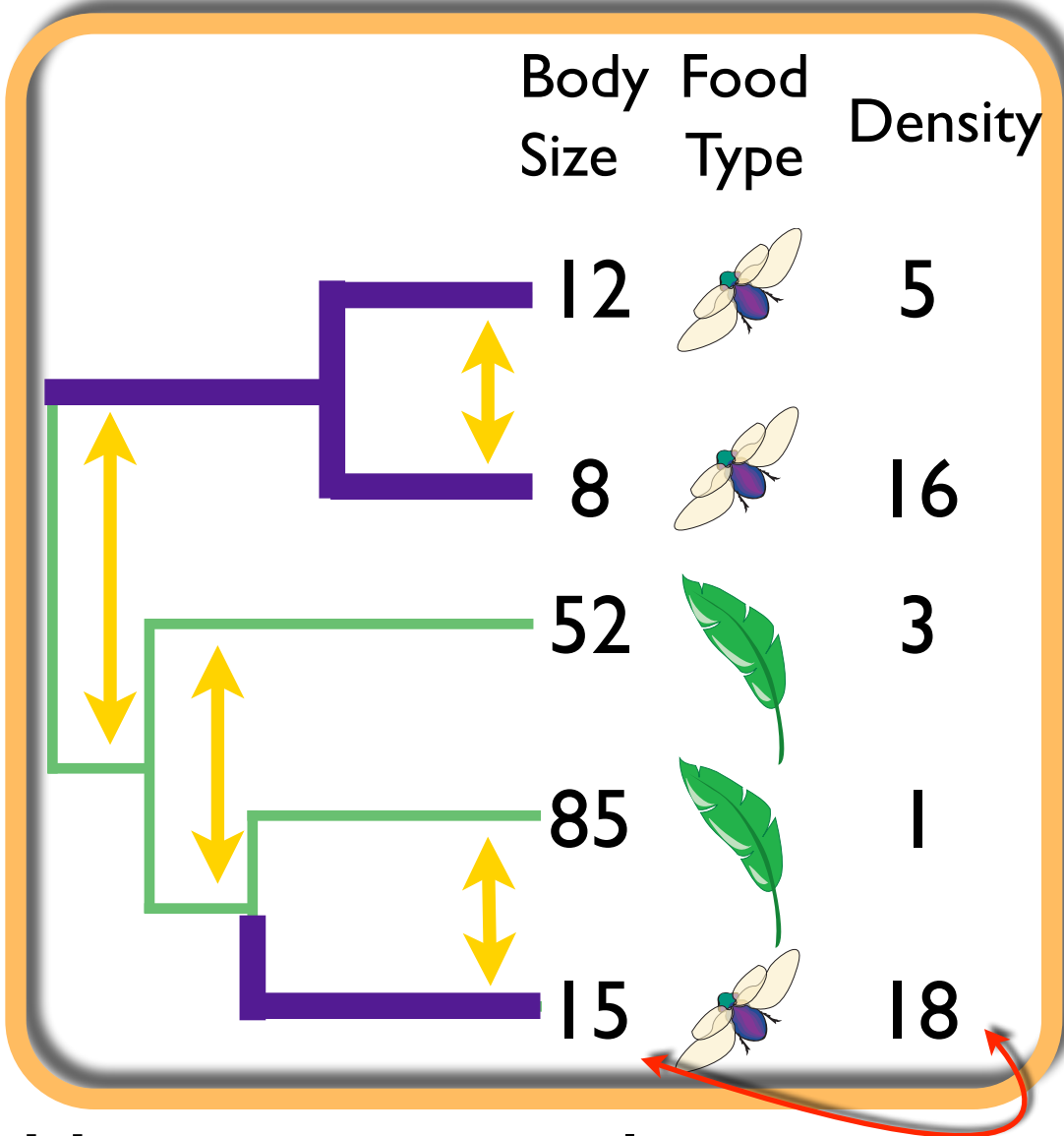
Aaron King

University of Michigan, Ecology & Evolutionary Biology

Marguerite Butler

University of Hawaii, Department of Zoology

(*Ornstein-Uhlenbeck for Comparative Hypotheses)

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
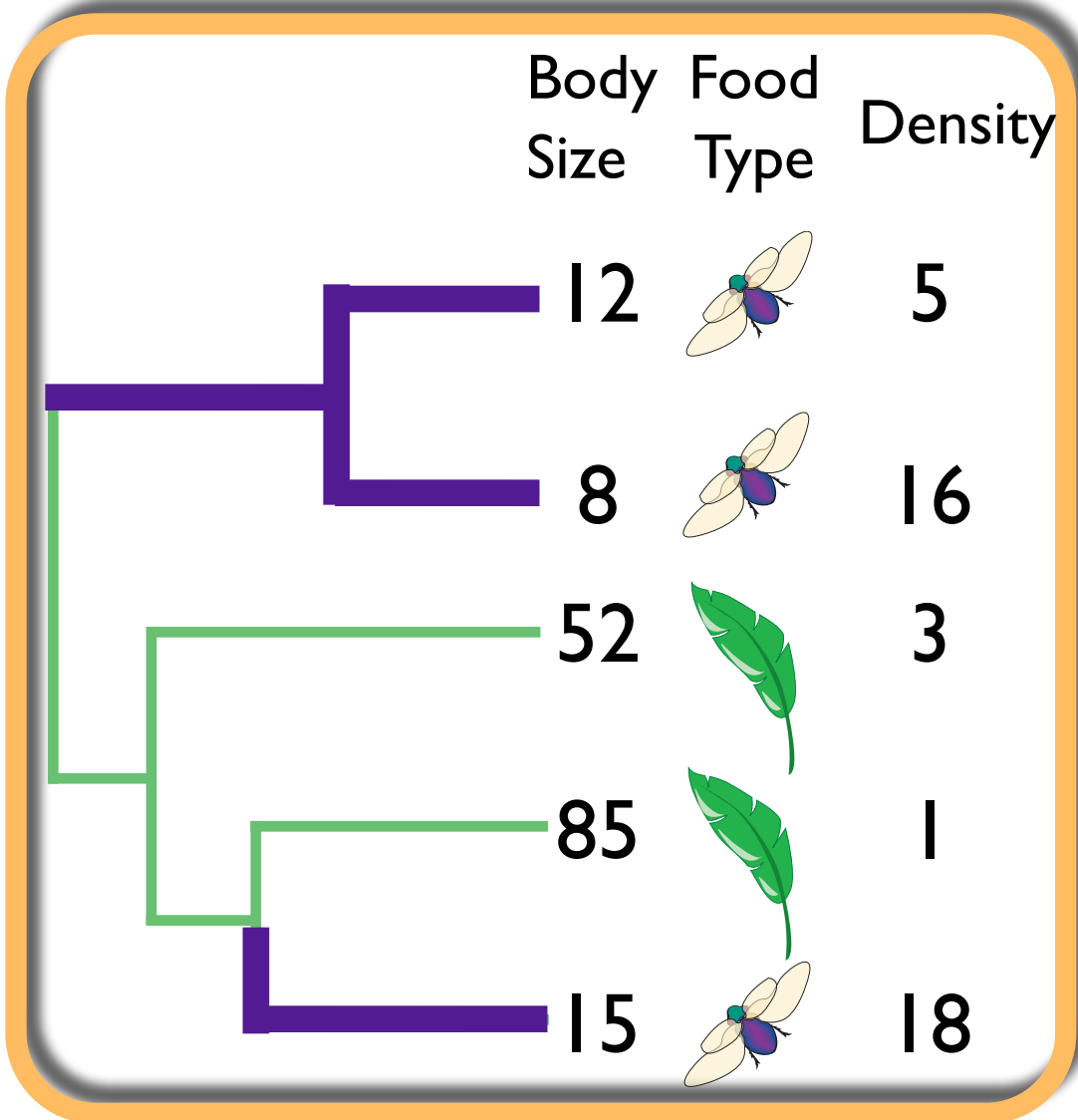
Why study correlated evolution?
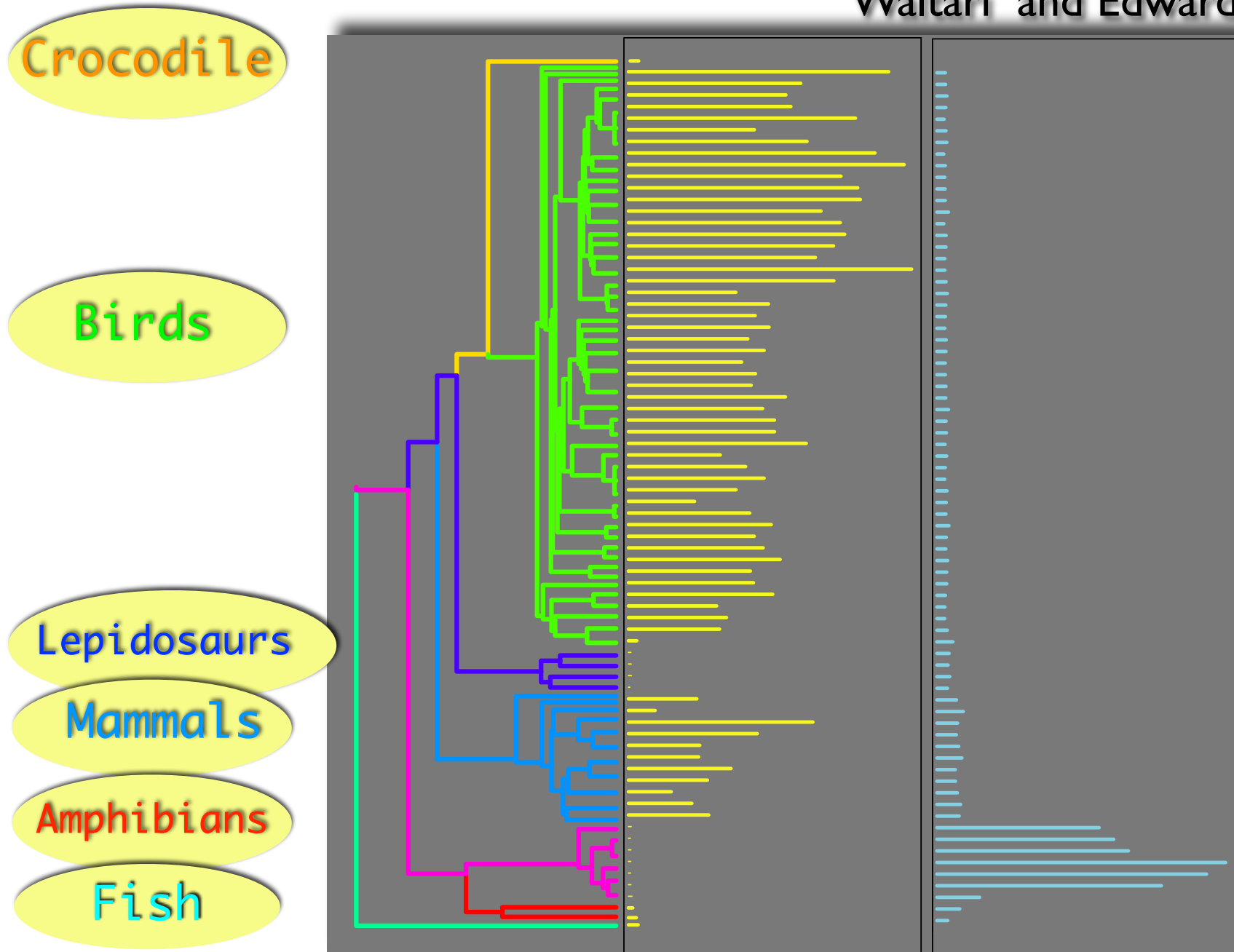
Helps us to understand patterns in diversity

Reveals important functional relationships

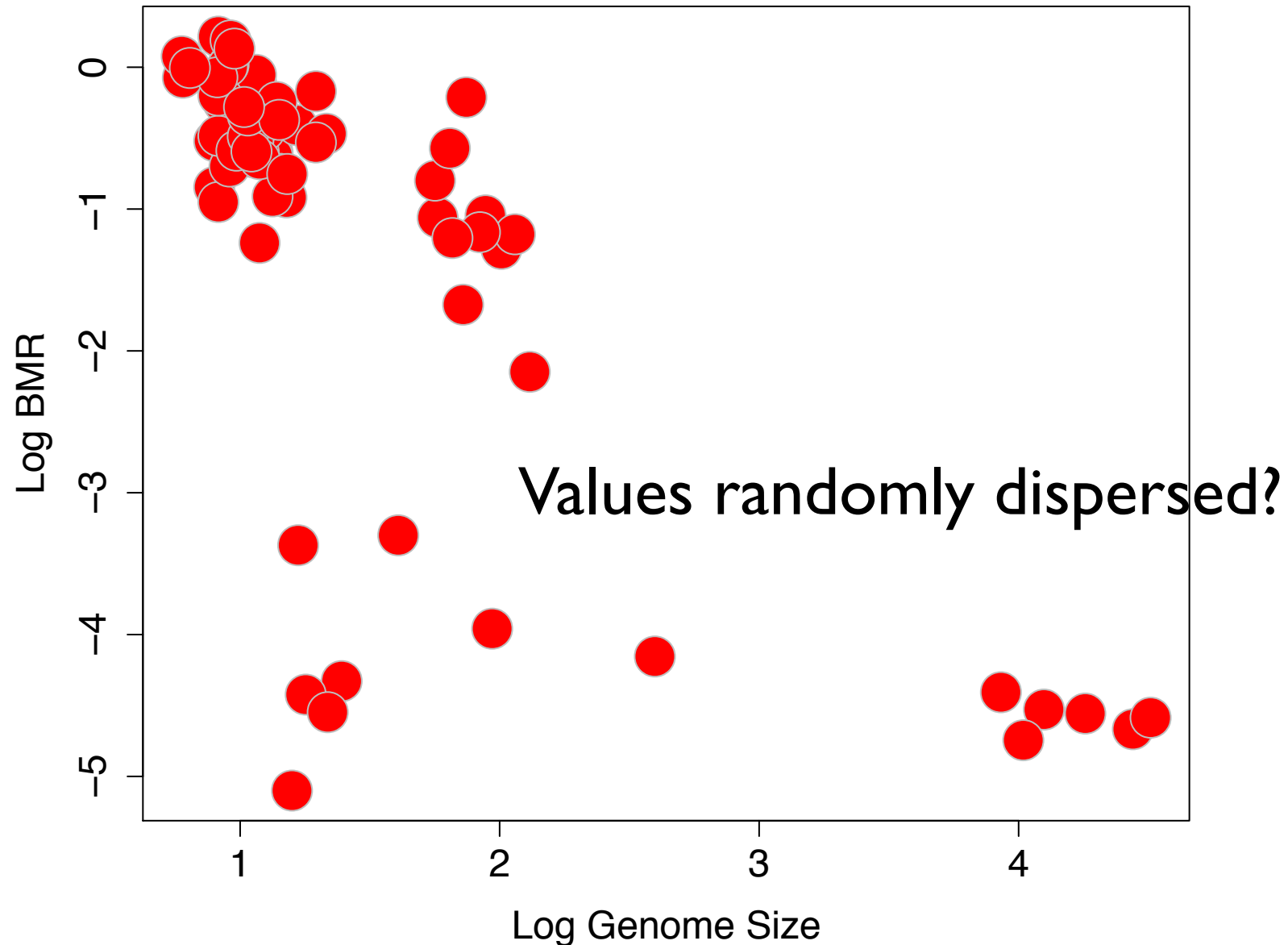Correlated selection can also prevent traits from reaching their individual optima

Here we develop a general modeling approach for multiple characters
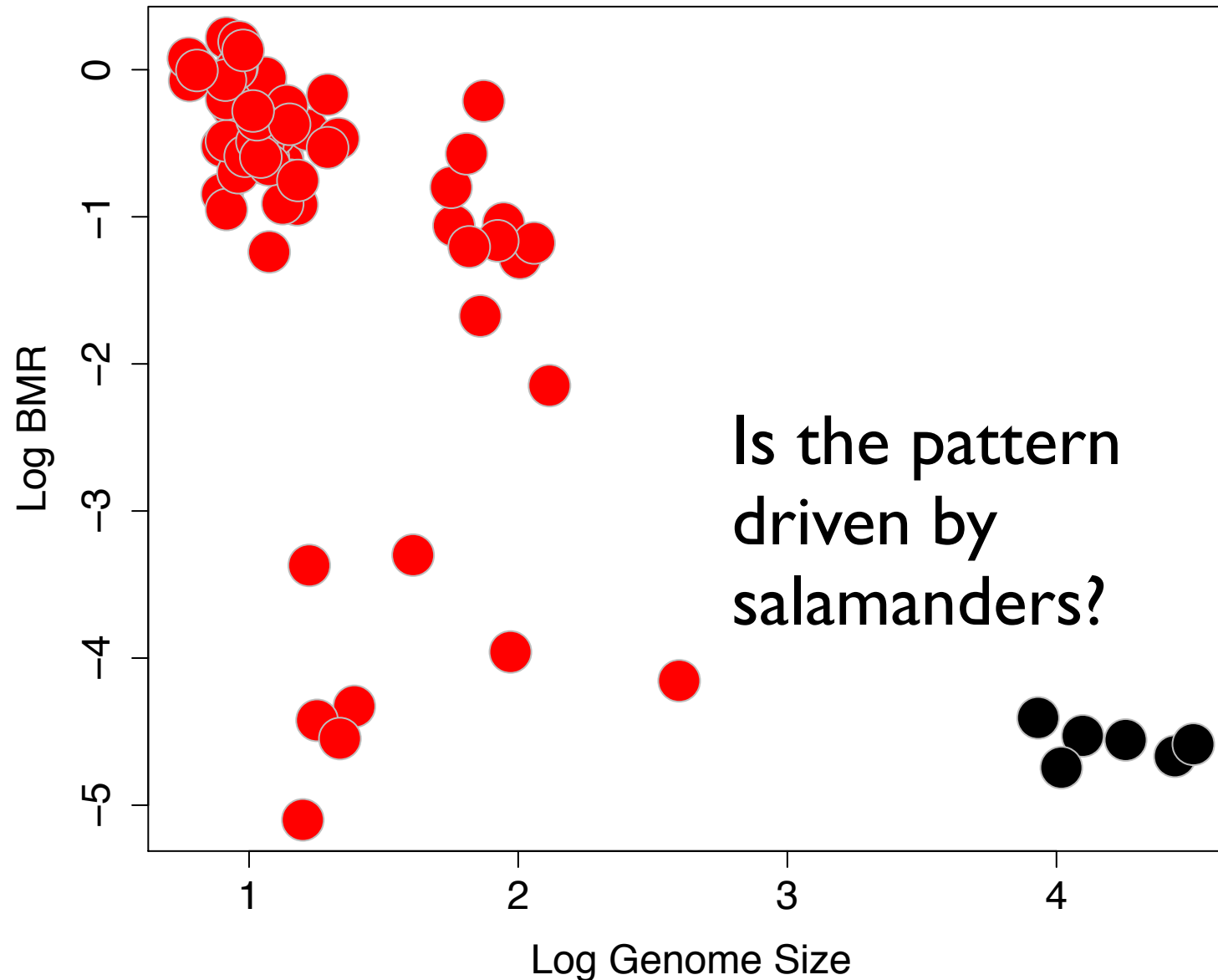
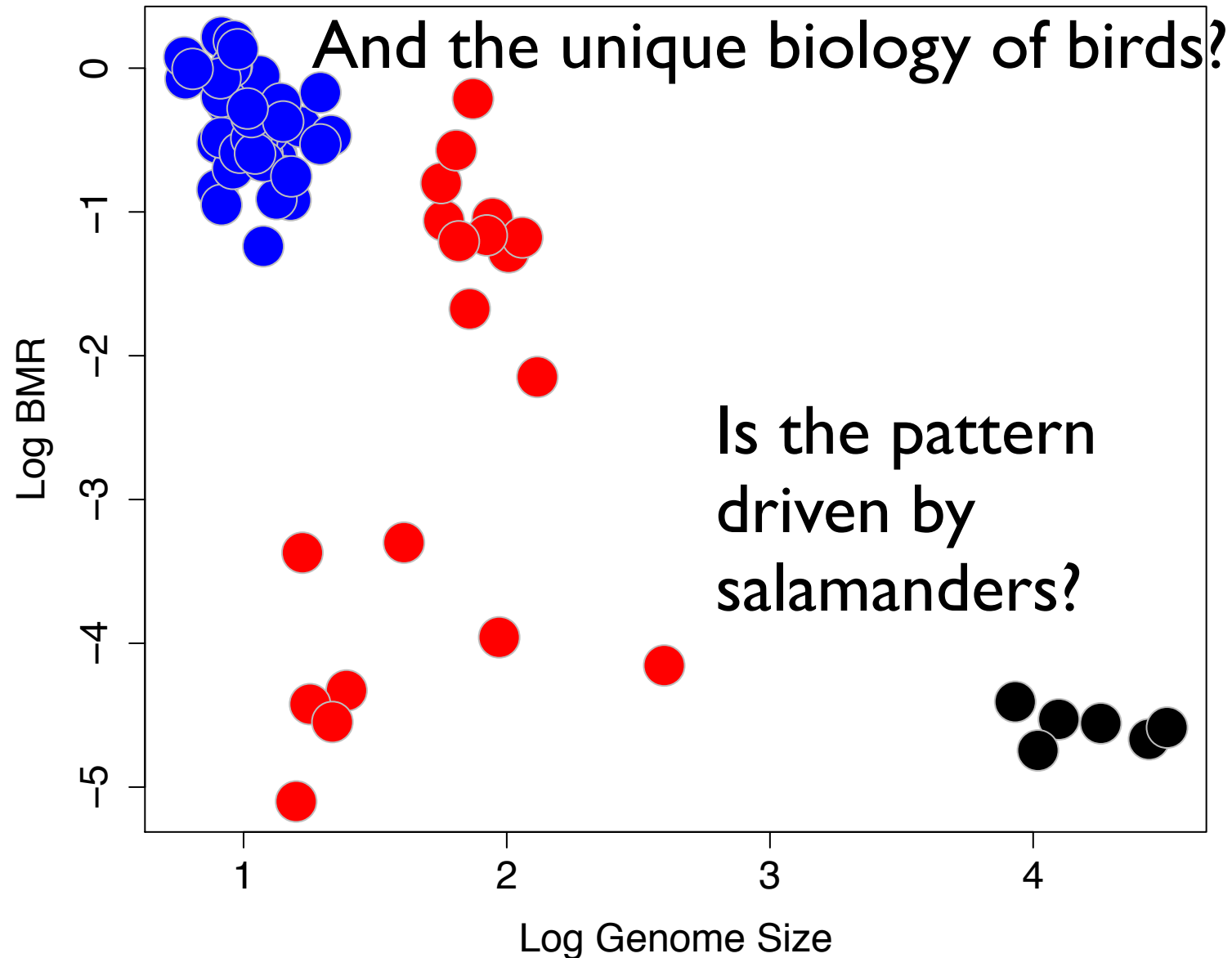# Basal Metabolic Rate and Genome Size

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?

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?

Log BMR (y-axis)
Log Genome Size (x-axis)

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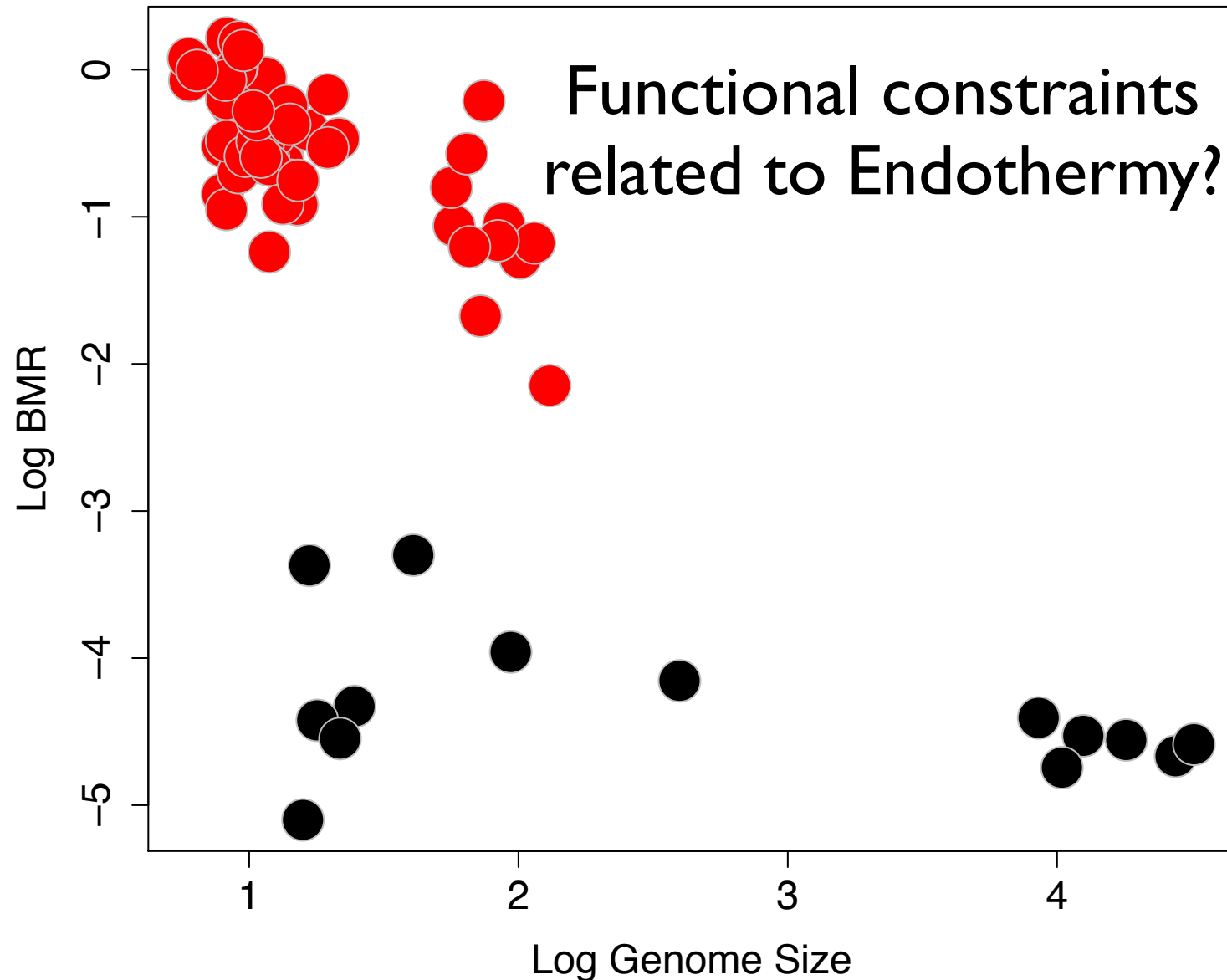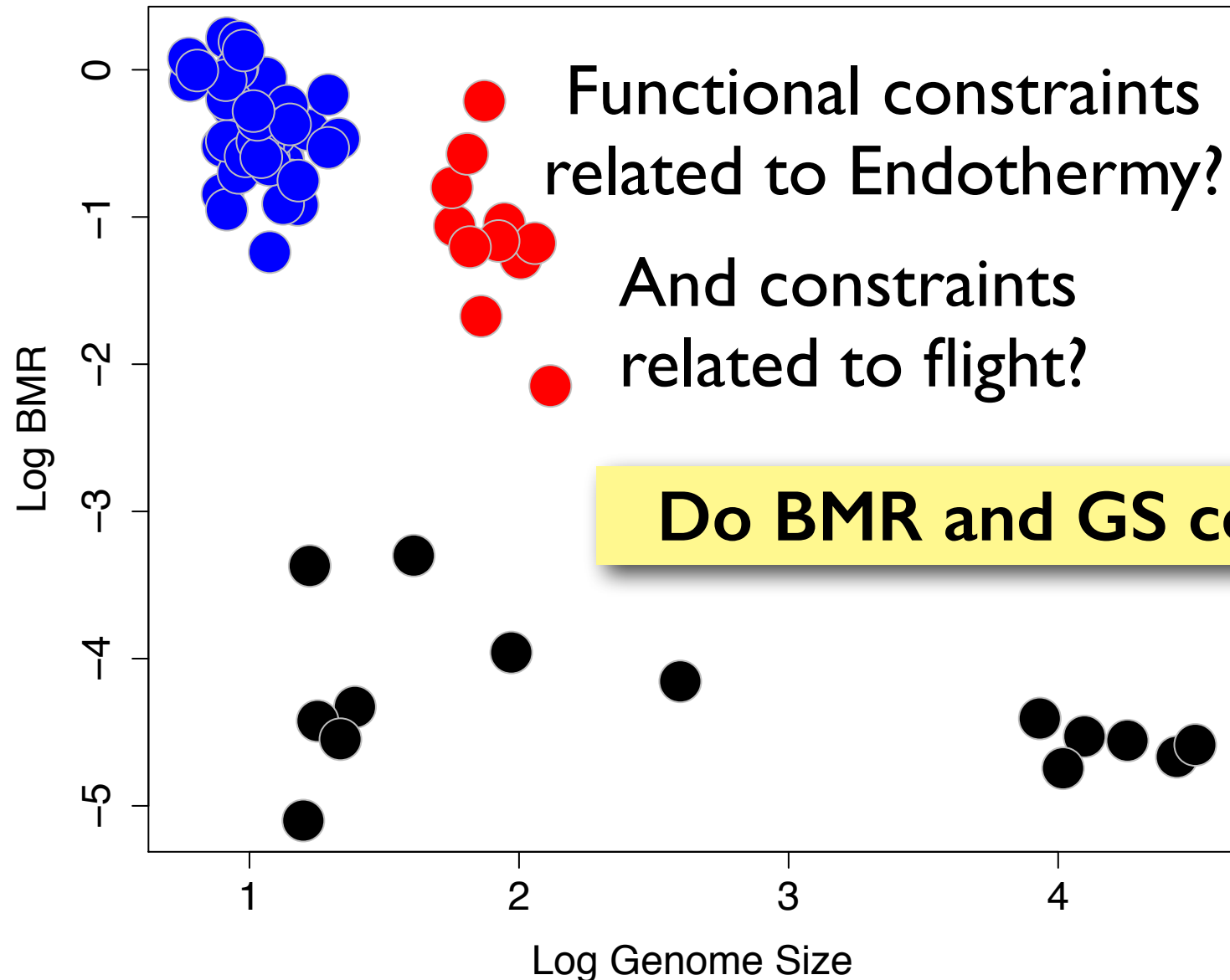
How do we explain the evolution of BMR and GS? Are they linked? What is the relationship?
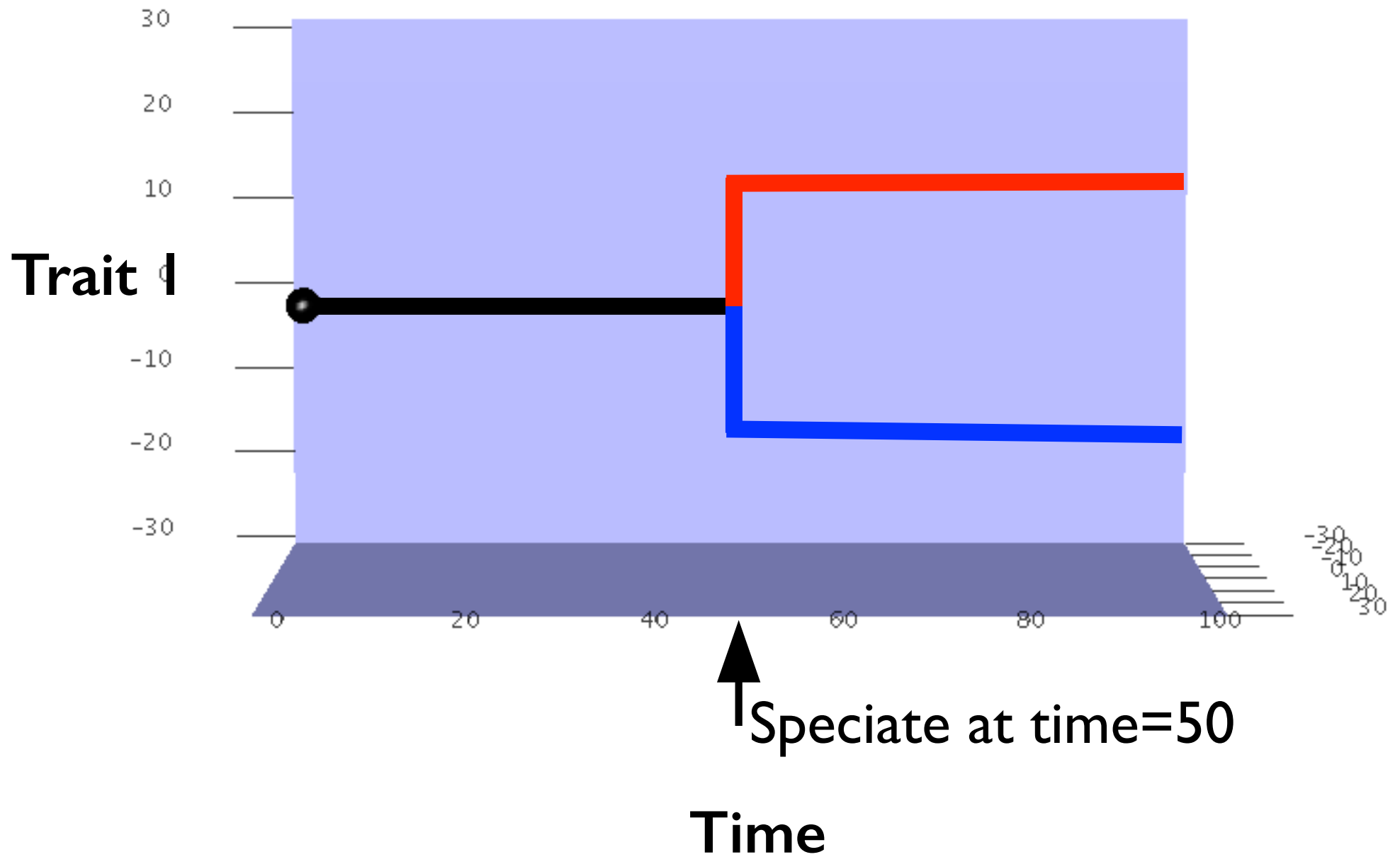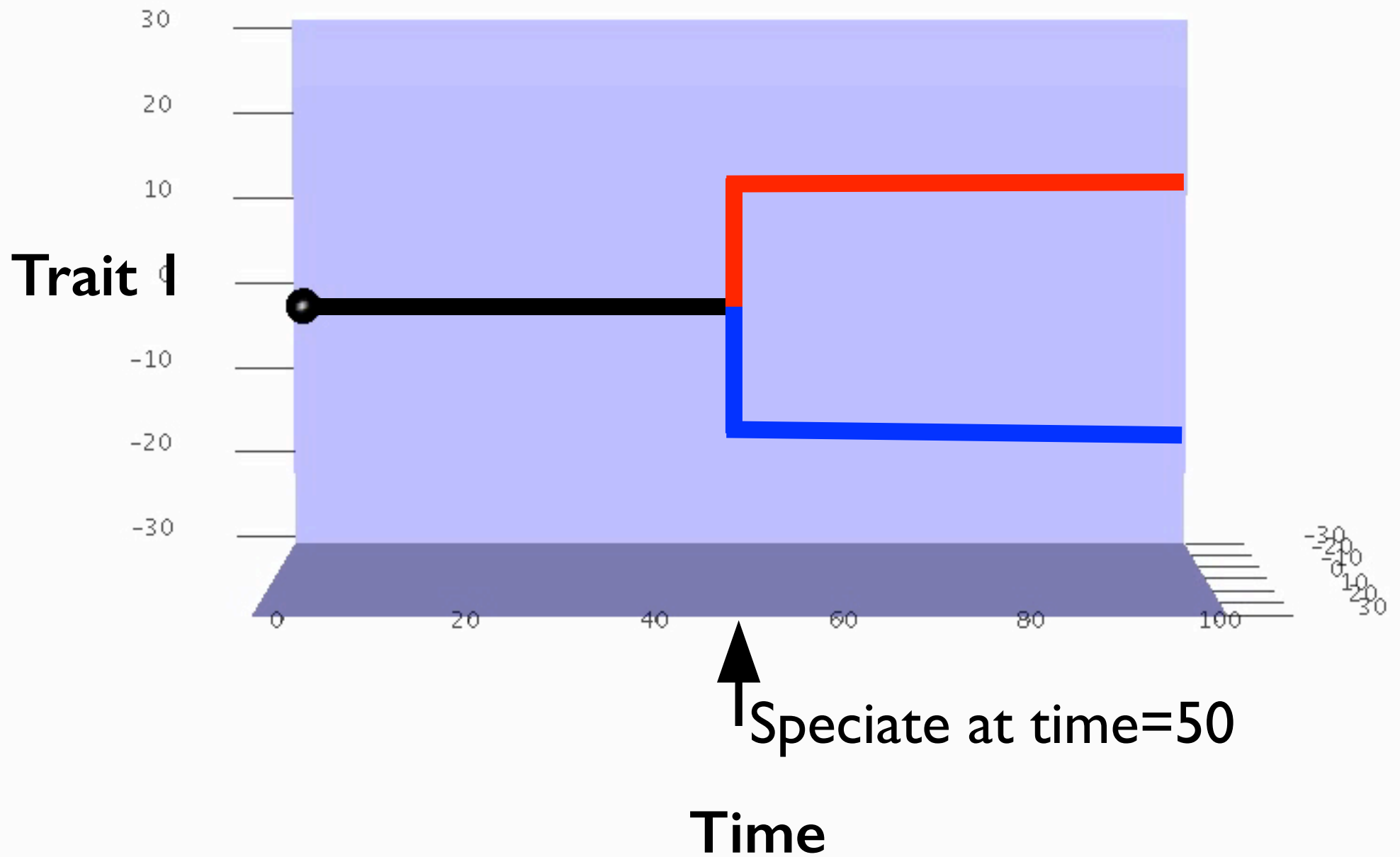

Functional constraints related to Endothermy?

And constraints related to flight?

Do BMR and GS covary?

(y-axis: Log BMR; x-axis: Log Genome Size)

Illustrating the Model

Brownian motion in 2D

Speciate at time=50

Time

Trait 1

Brownian motion in 2D

Trait 1

Speciate at time=50

Time

Brownian motion in 2D
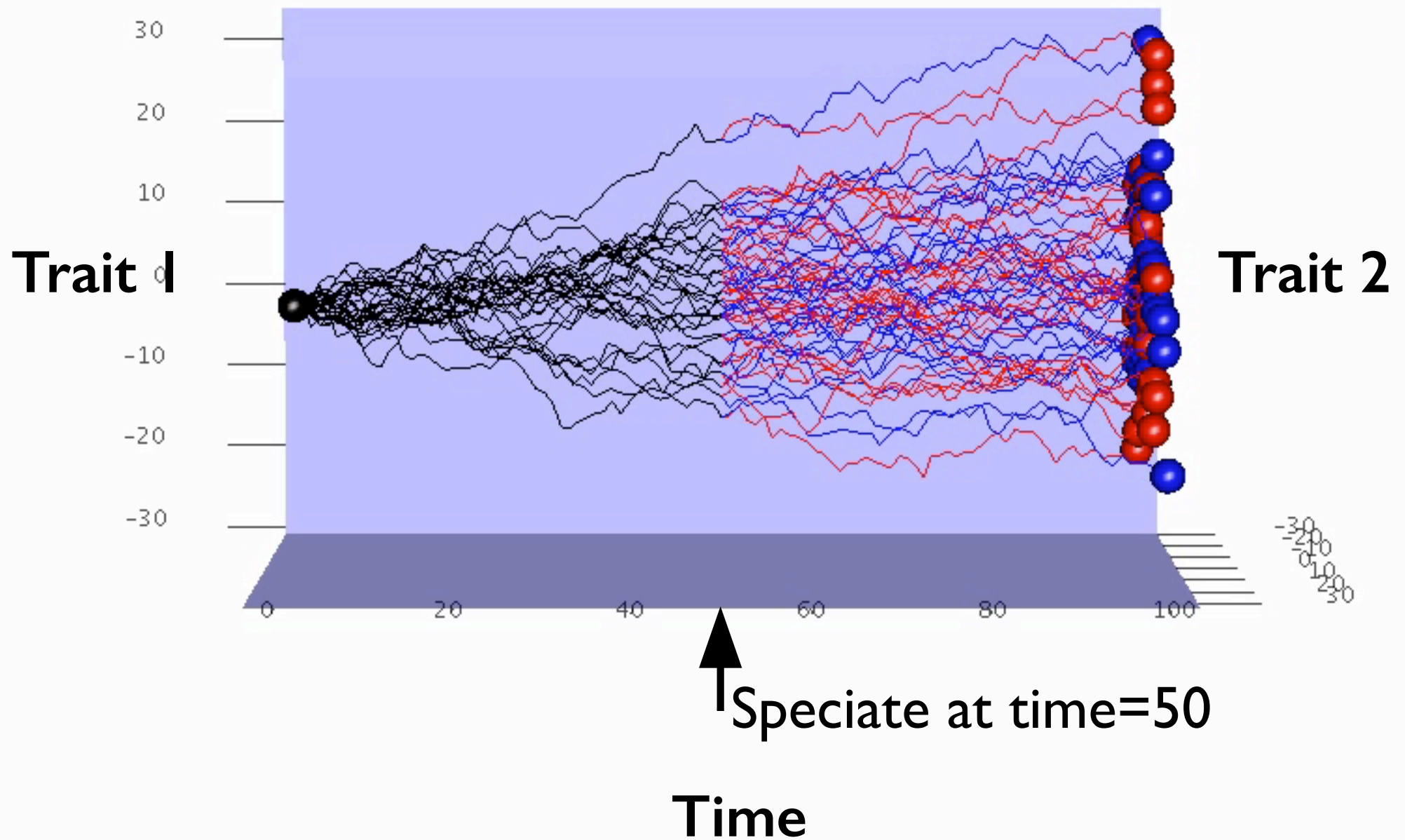
Brownian motion in 2D
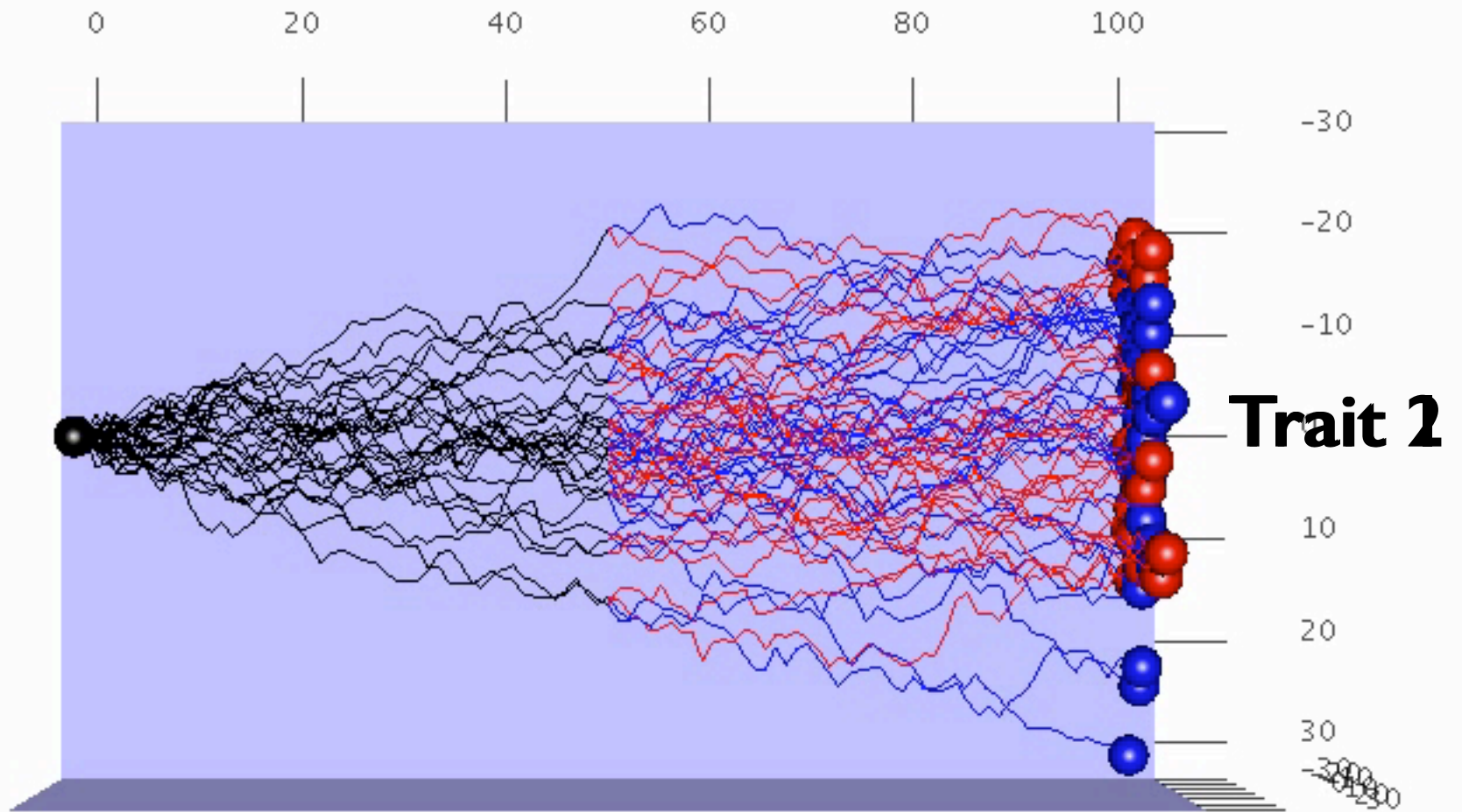
Trait 2

Trait 2

Time

Orstein-Uhlenbeck with strong selection

Trait 1

Optimum 1

Optimum 2

Starting value

Speciate at time=50

Time
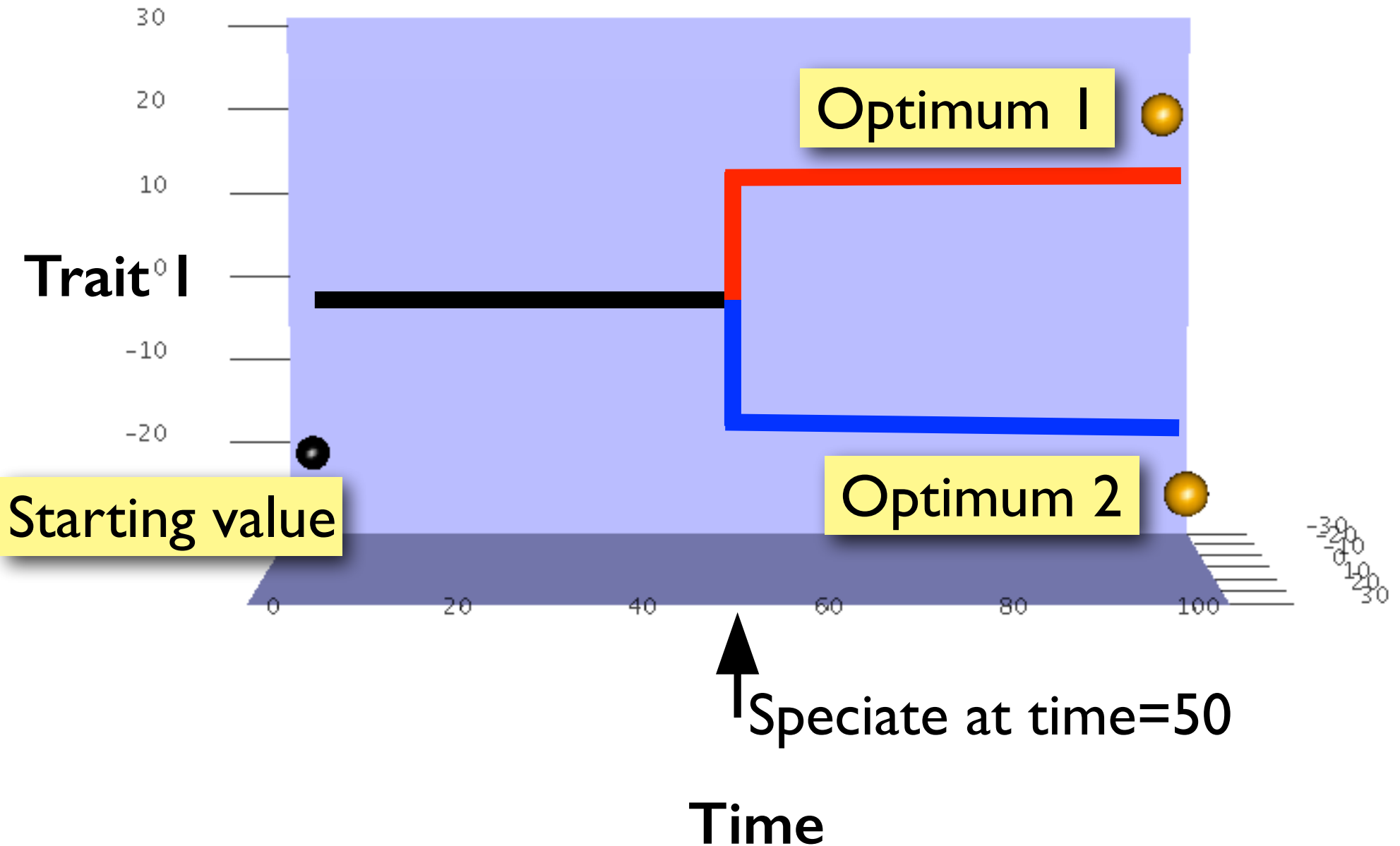
Orstein-Uhlenbeck with strong selection

With two traits in a multivariate model, *(overlapping text: "With one trait evolving,")*
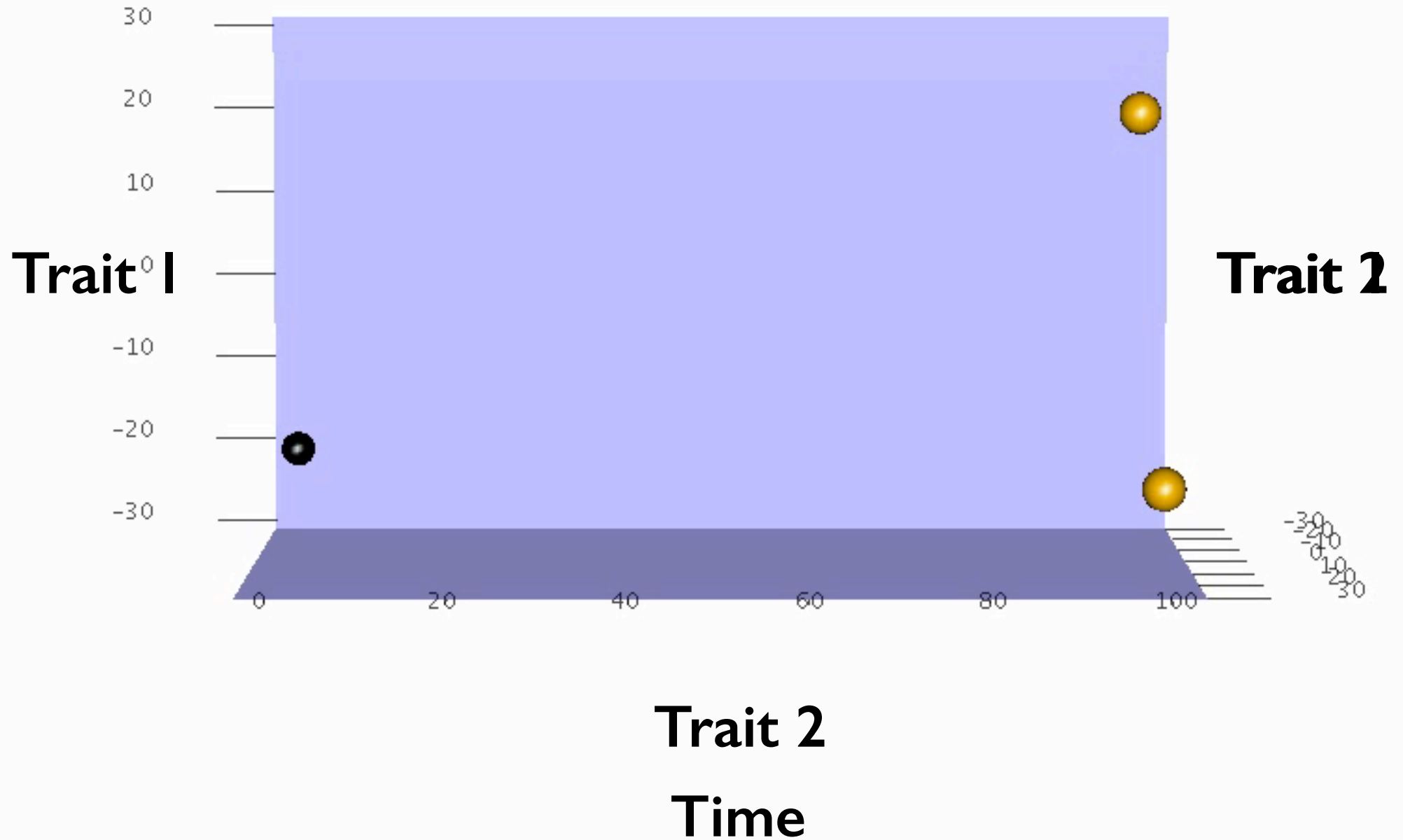
$$dX(t) = \alpha \left(\theta - X(t) \right) dt + \sigma \, dB(t)$$

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)

Illustrating the off-diagonal parameters

"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"

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{pmatrix} \quad \sigma = \begin{pmatrix} \sigma_{11} & 0 \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$$
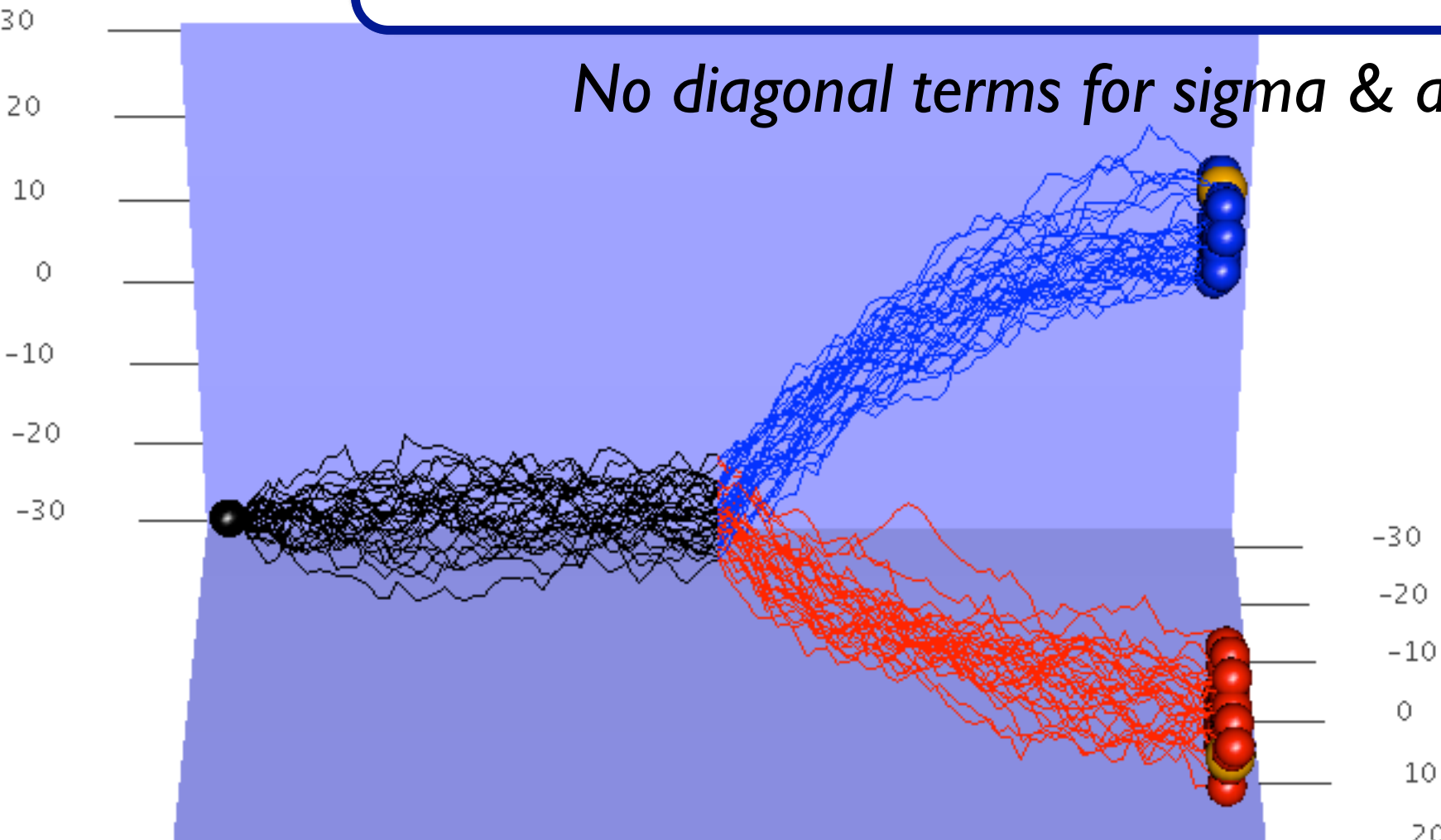
add off-diagonal sigma and alpha terms

Illustrating the off-diagonal parameters

"Double Univariate"

$$\alpha = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$
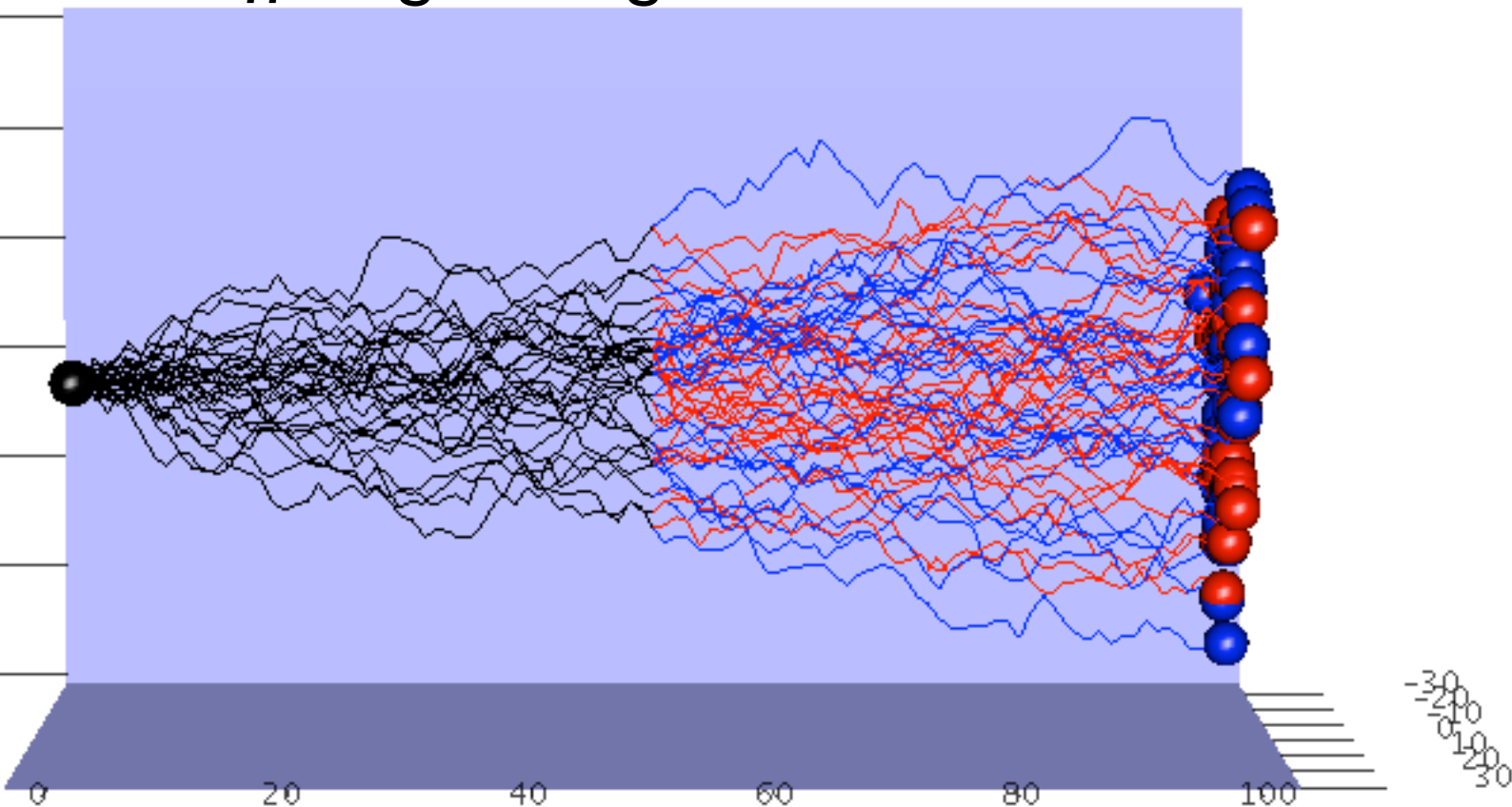
No diagonal terms for sigma & alpha

Illustrating the off-diagonal parameters

Correlated "drift"

$$\alpha = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \sigma = \begin{pmatrix} 10 & 0 \\ 5 & 10 \end{pmatrix}$$
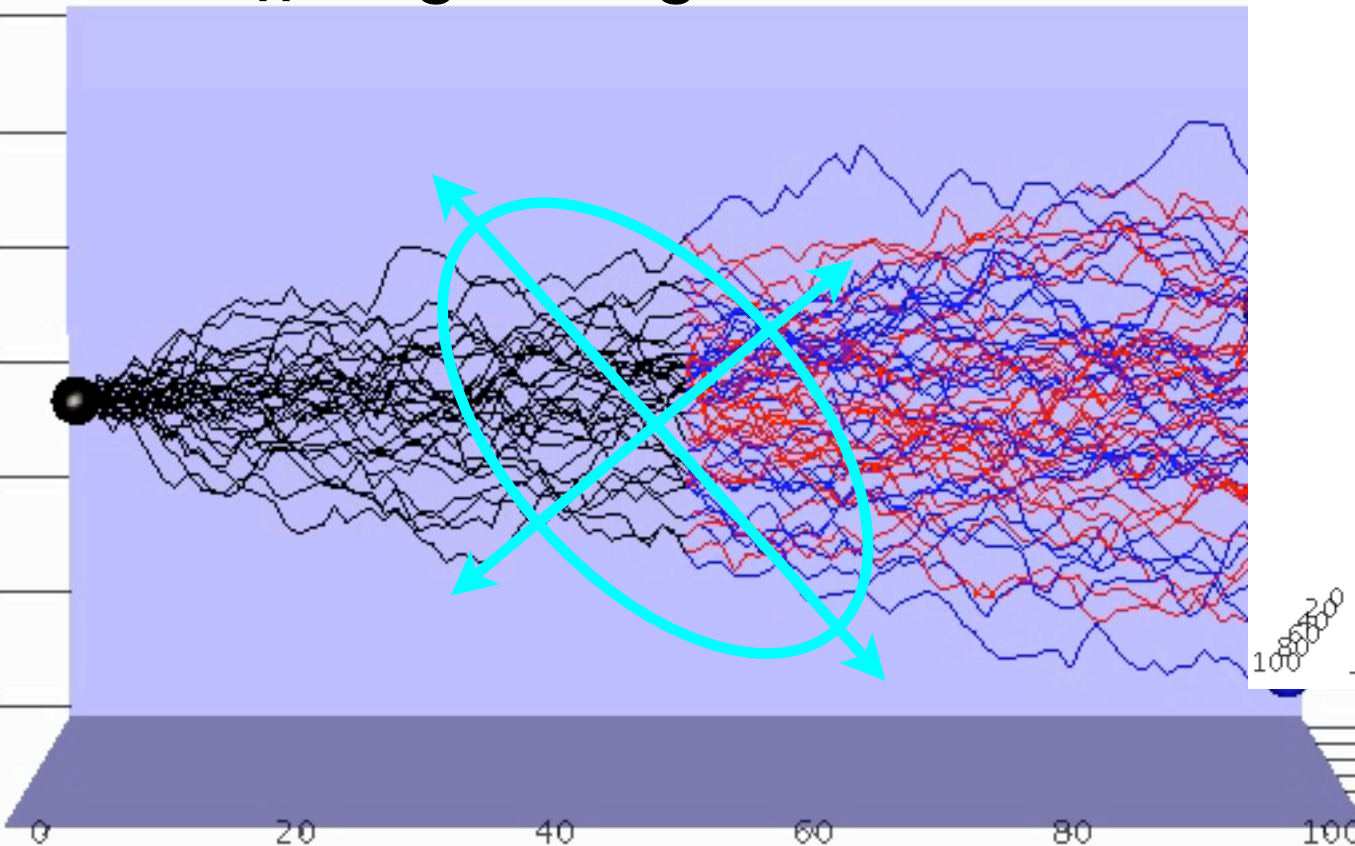
add off-diagonal sigma term

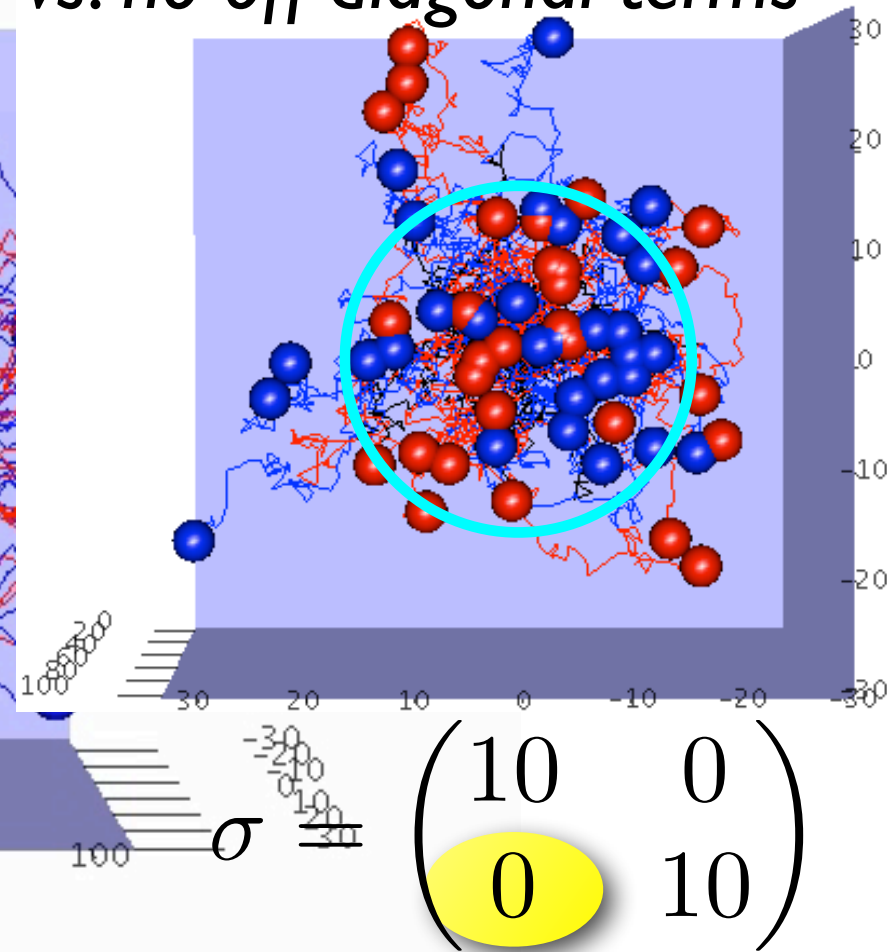# Illustrating the off-diagonal parameters

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*

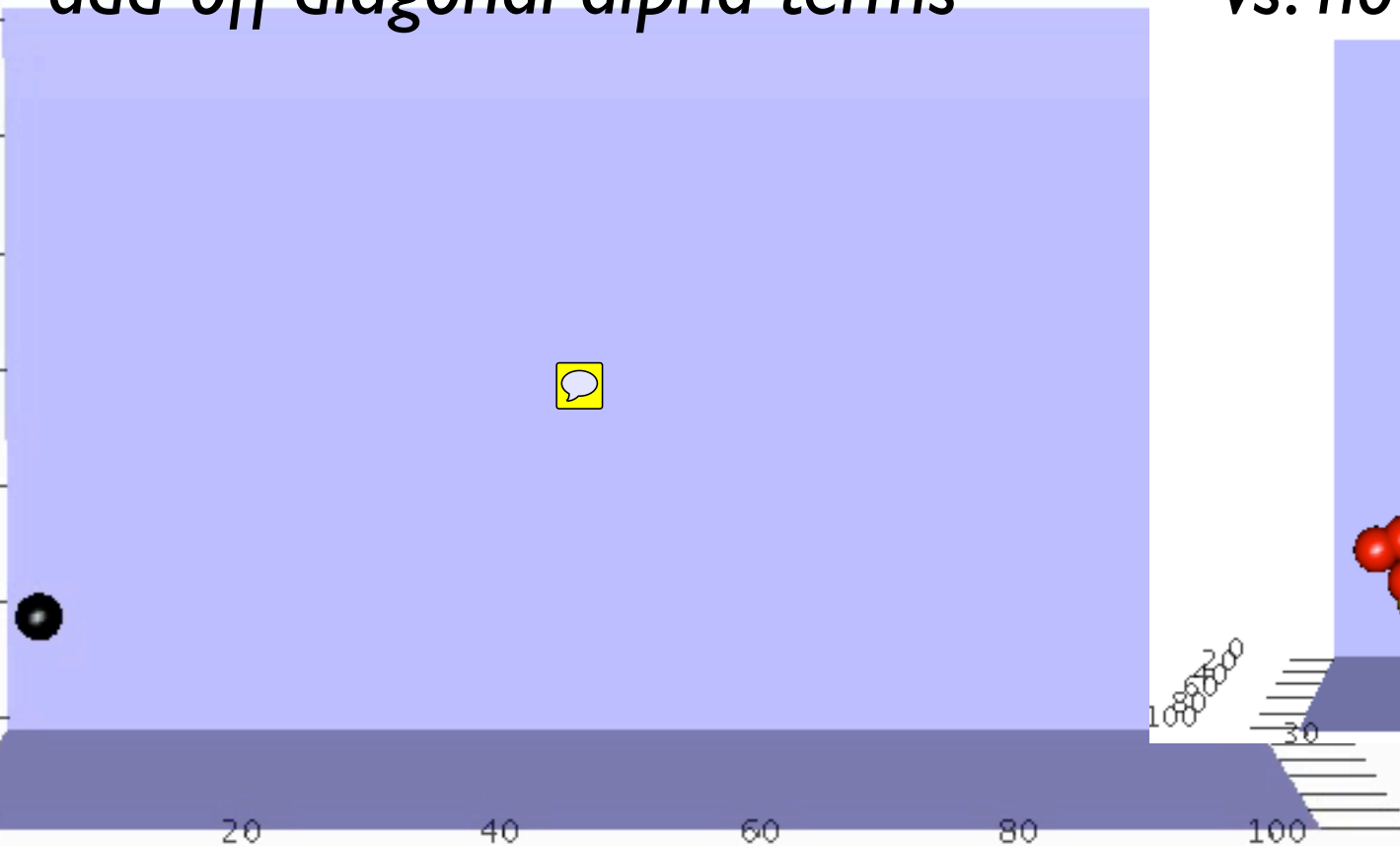$$\sigma = \begin{pmatrix} 10 & 0 \\ 0 & 10 \end{pmatrix}$$

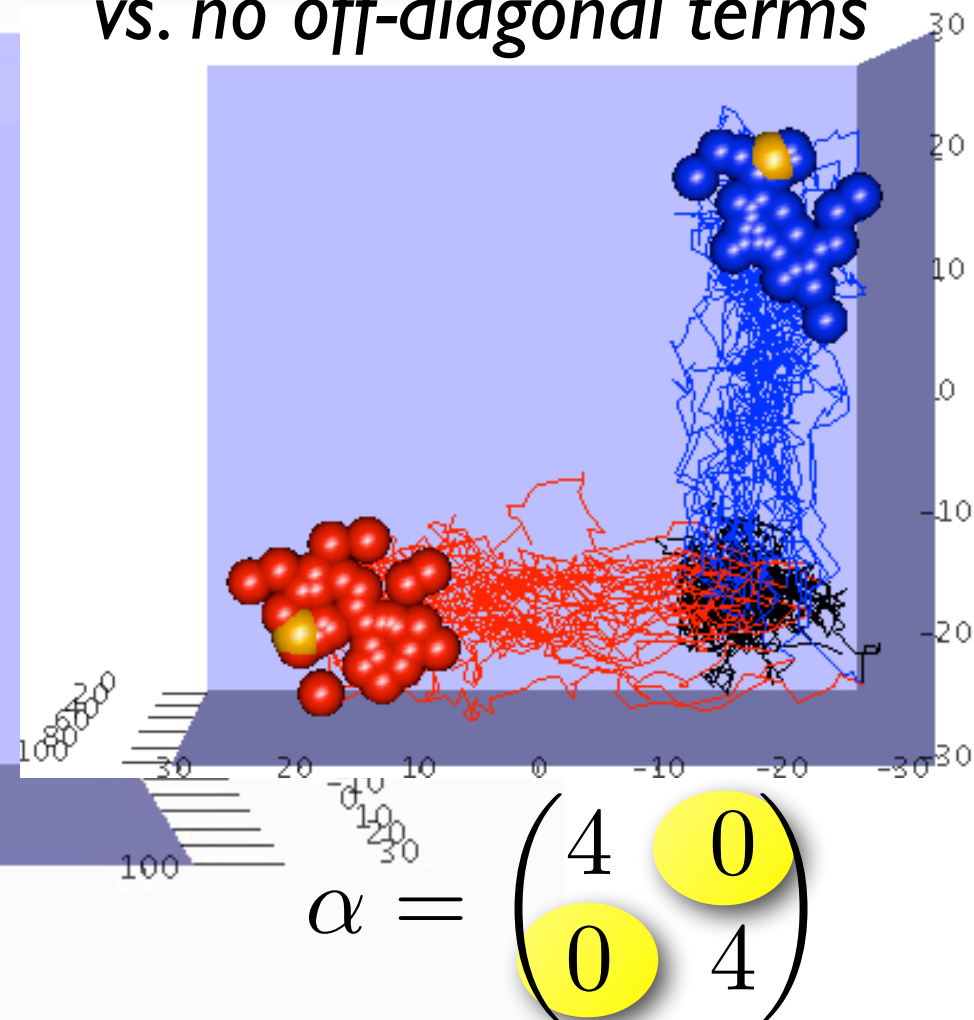# *Illustrating the off-diagonal parameters*

Correlated "selection"

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

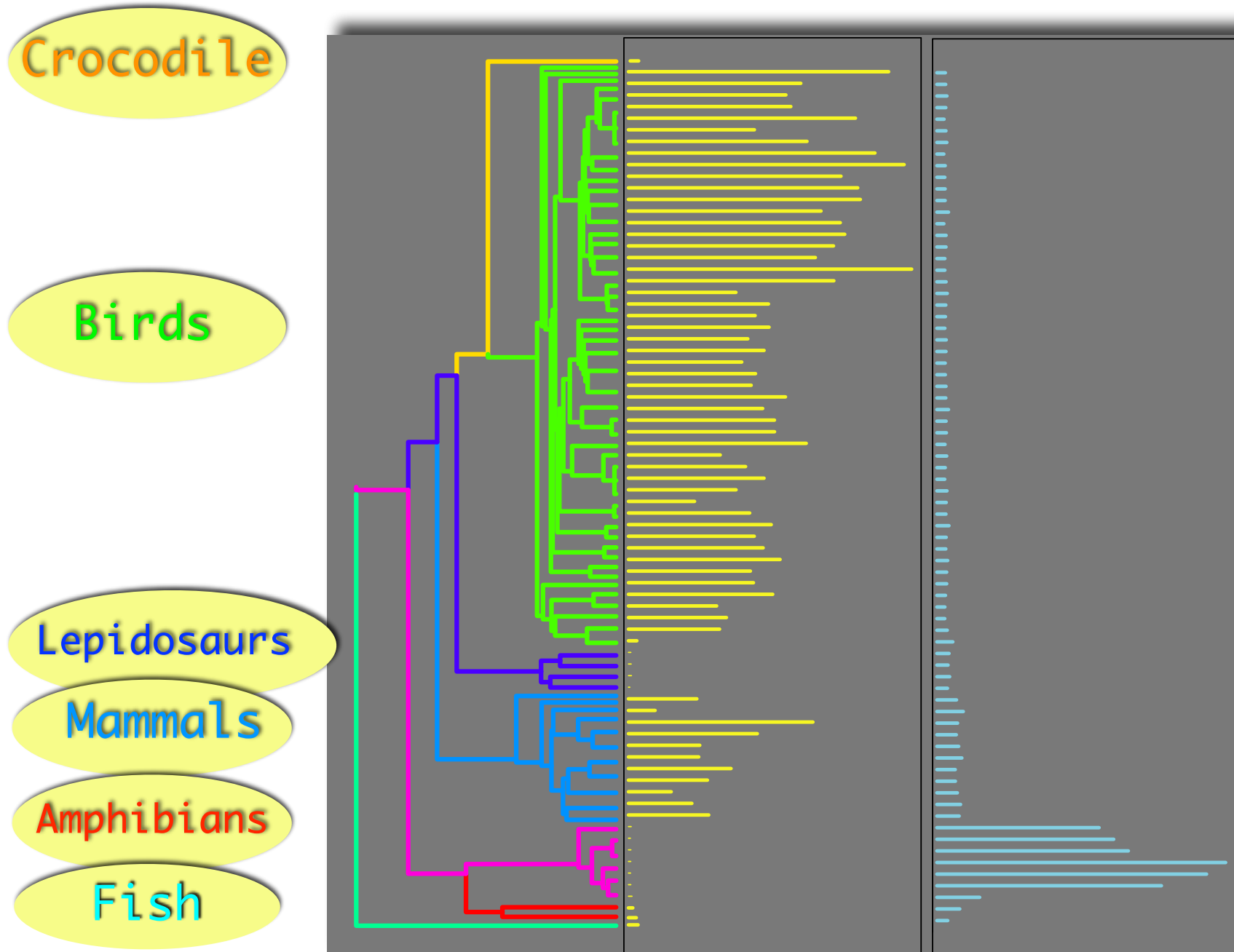*add off-diagonal alpha terms*

vs. no off-diagonal terms

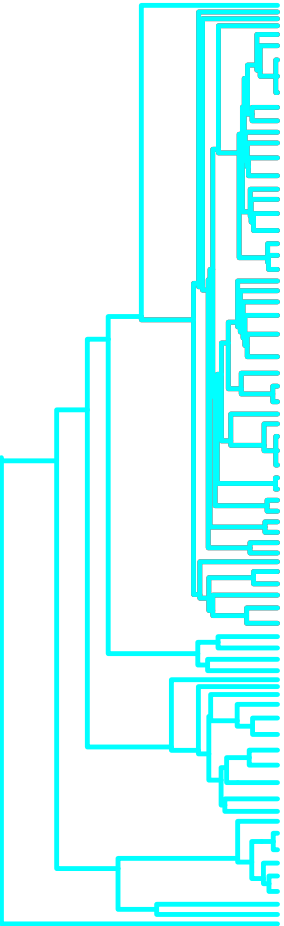$$\alpha = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}$$

Back to our example

Basal Metabolic Rate and Genome Size

Waltari & Edwards 2002

Crocodile

Birds

Lepidosaurs

Mammals

Amphibians

Fish

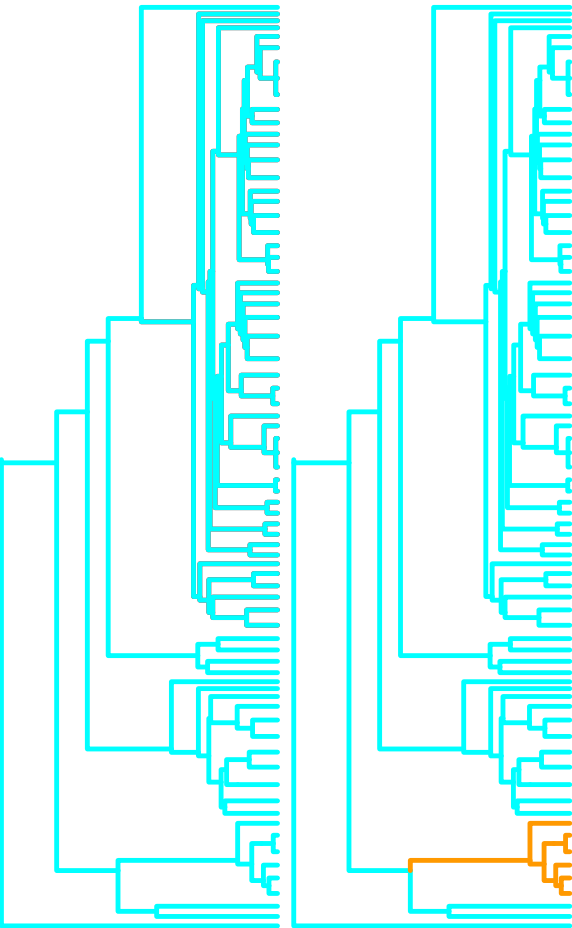# Alternative Adaptive Regime Models

BM

Alternative Adaptive Regime Models

BM

Salamanders
Vertebrates

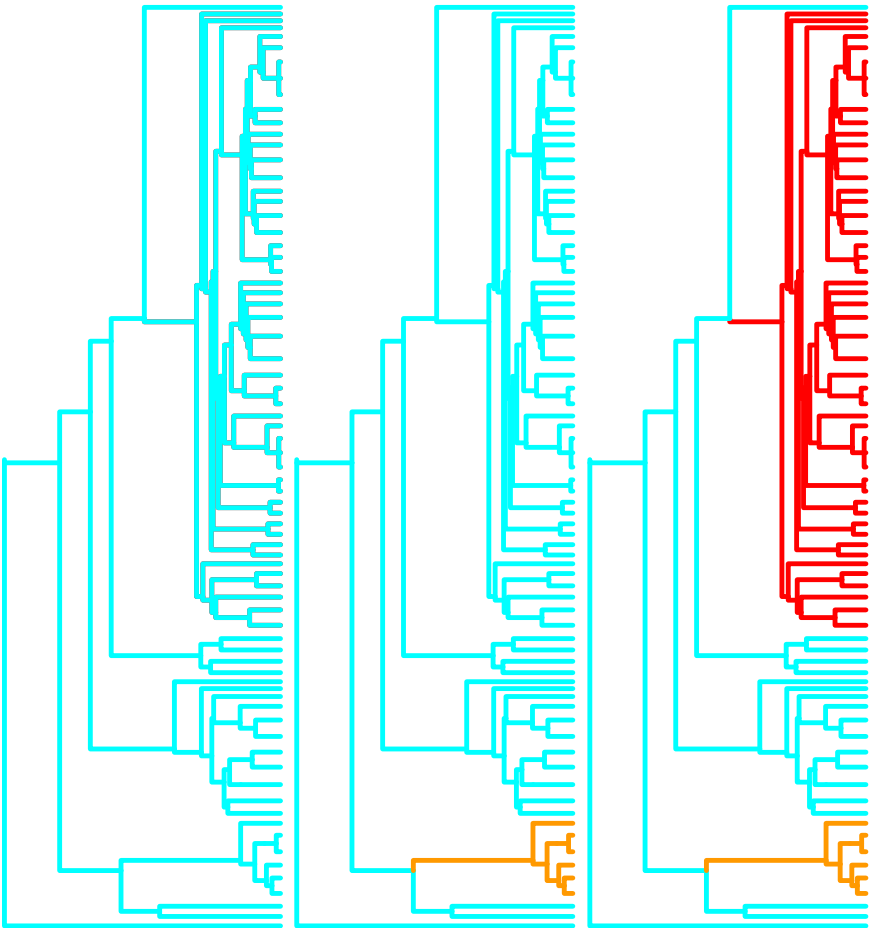# Alternative Adaptive Regime Models

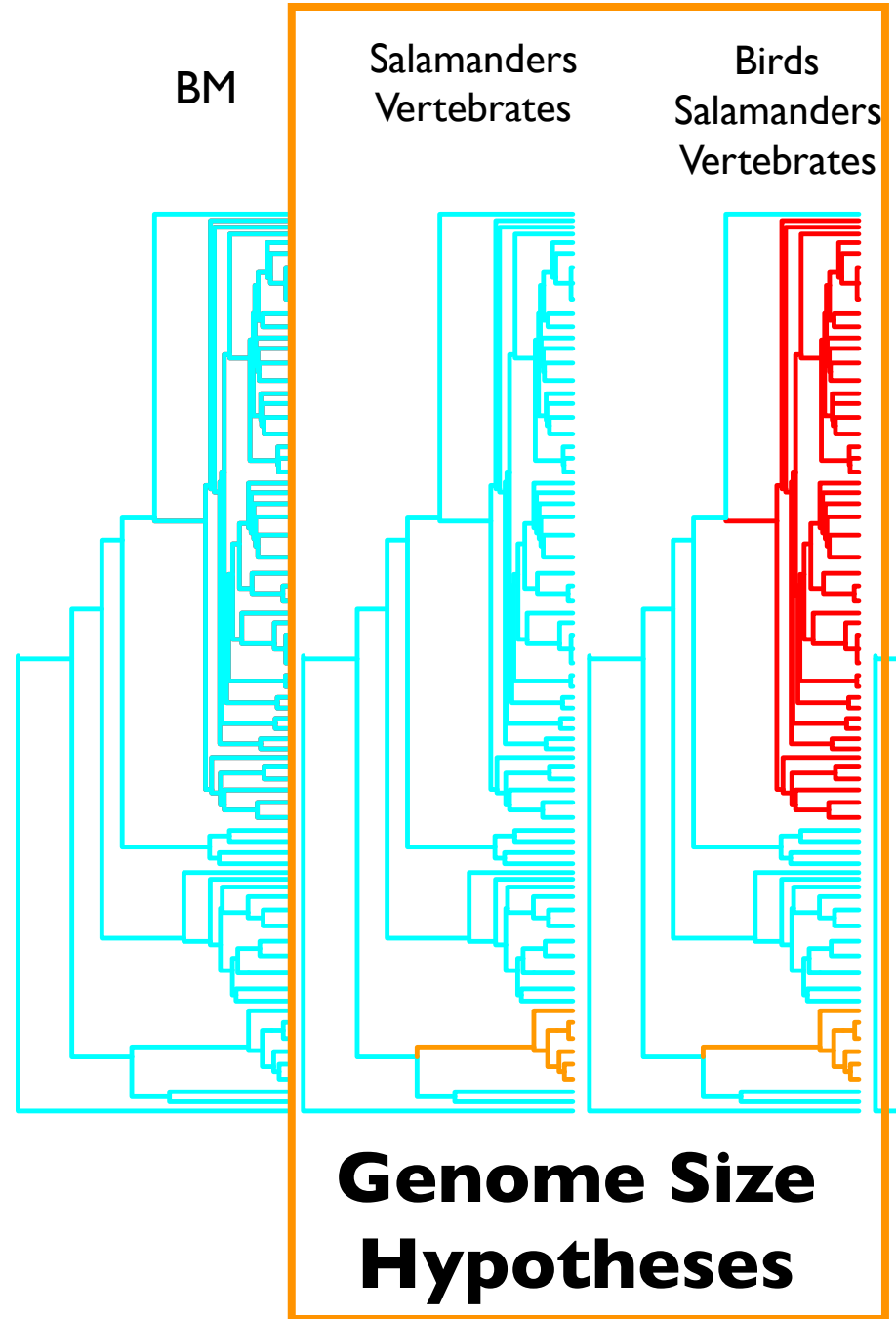BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Alternative Adaptive Regime Models
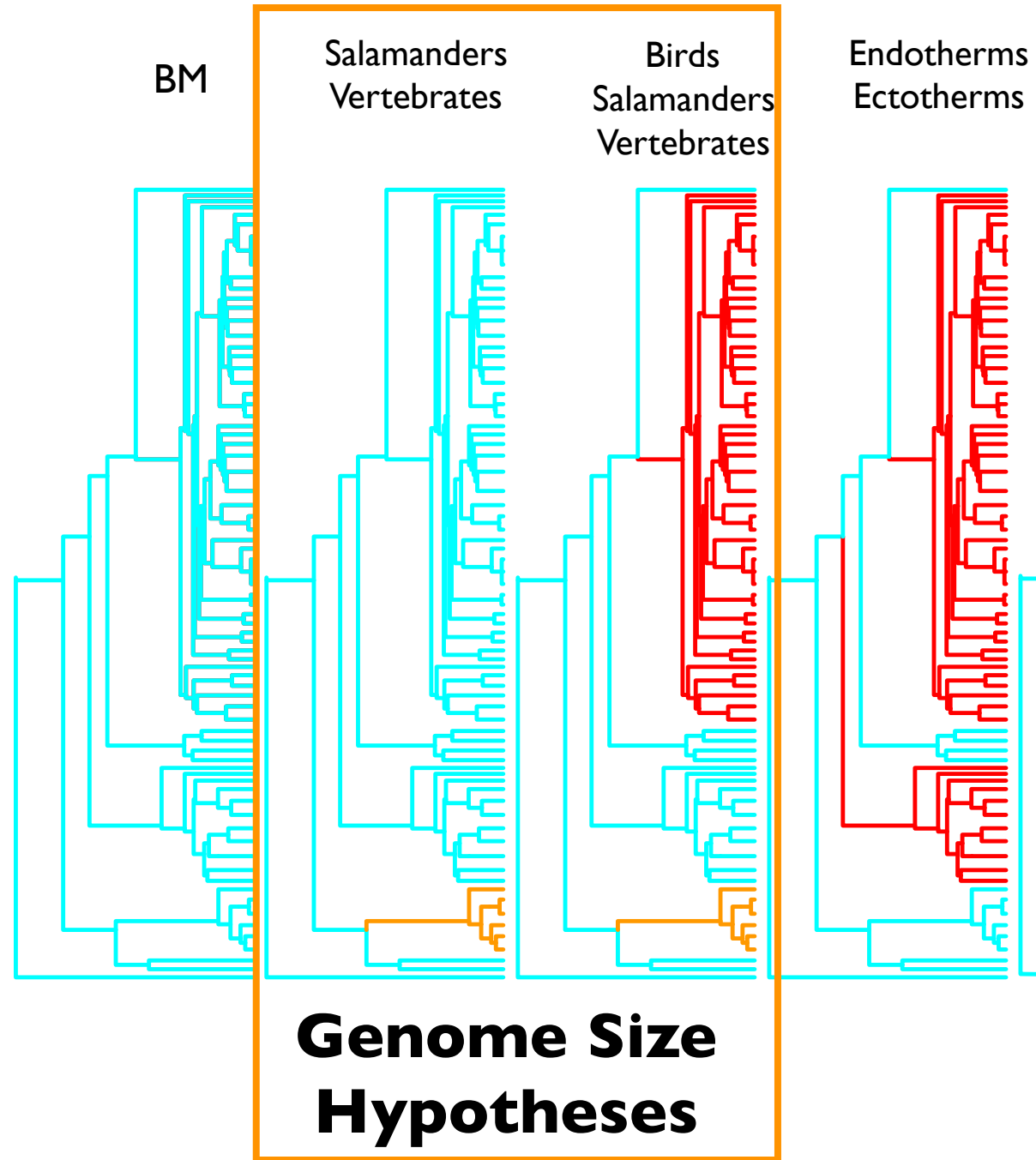
Alternative Adaptive Regime Models

BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

**Genome Size
Hypotheses**

Alternative Adaptive Regime Models
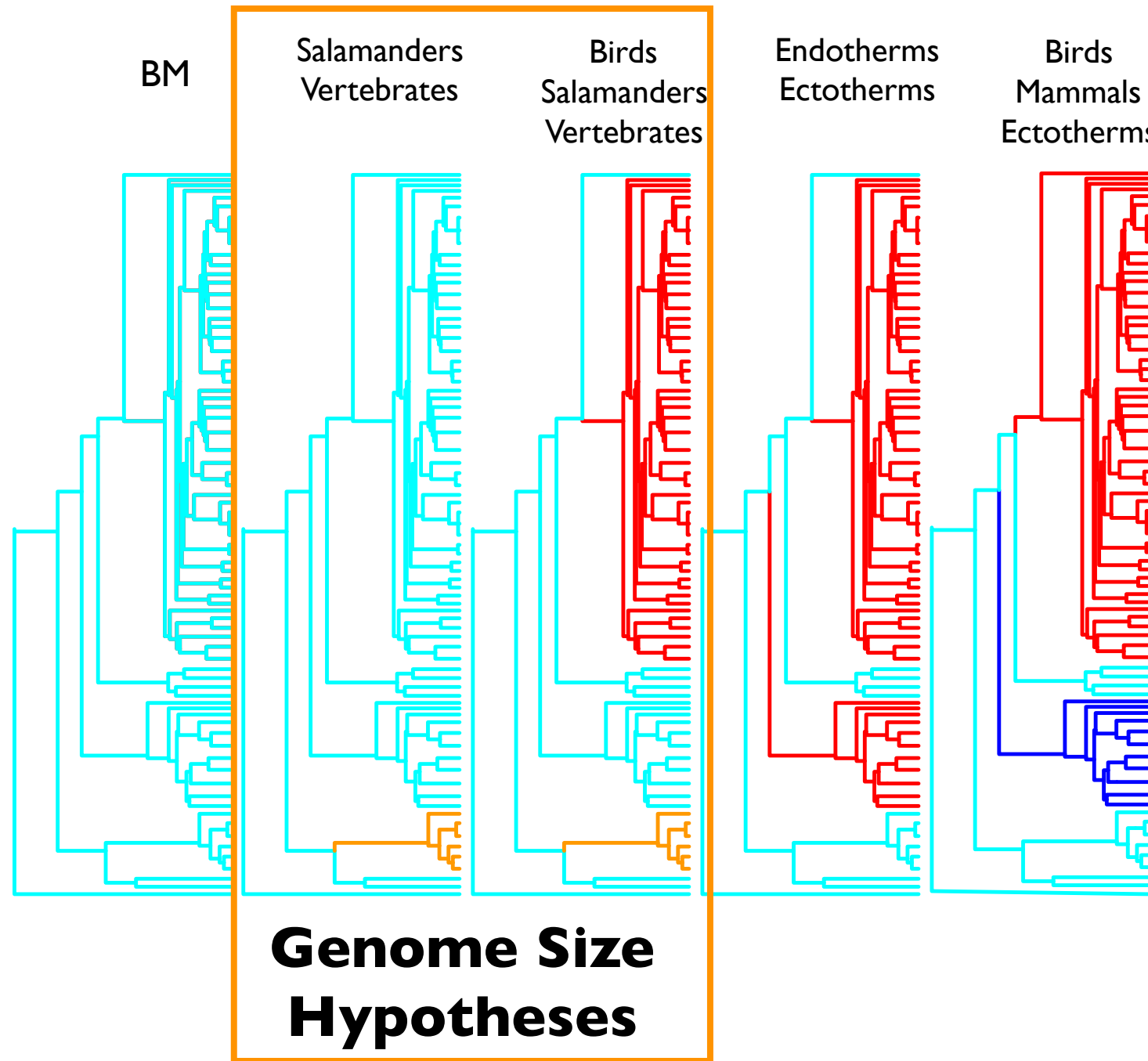

BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

Birds
Mammals
Ectotherms

**Genome Size
Hypotheses**

Alternative Adaptive Regime Models

Alternative Adaptive Regime Models

BM

Salamanders
Vertebrates

Birds
Salamanders
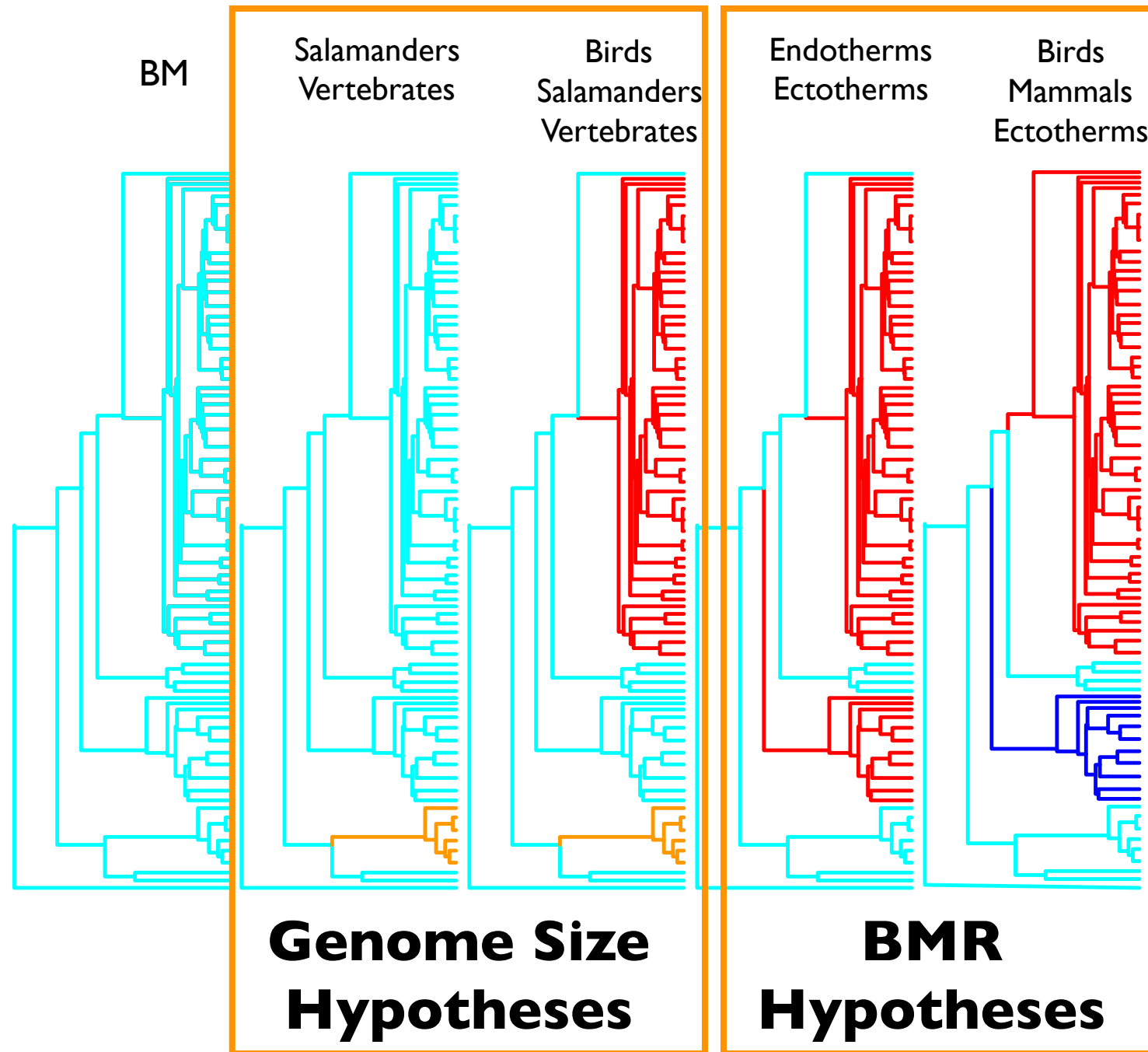Vertebrates

Endotherms
Ectotherms

Birds
Mammals
Ectotherms

Major
Taxa

**Genome Size
Hypotheses**
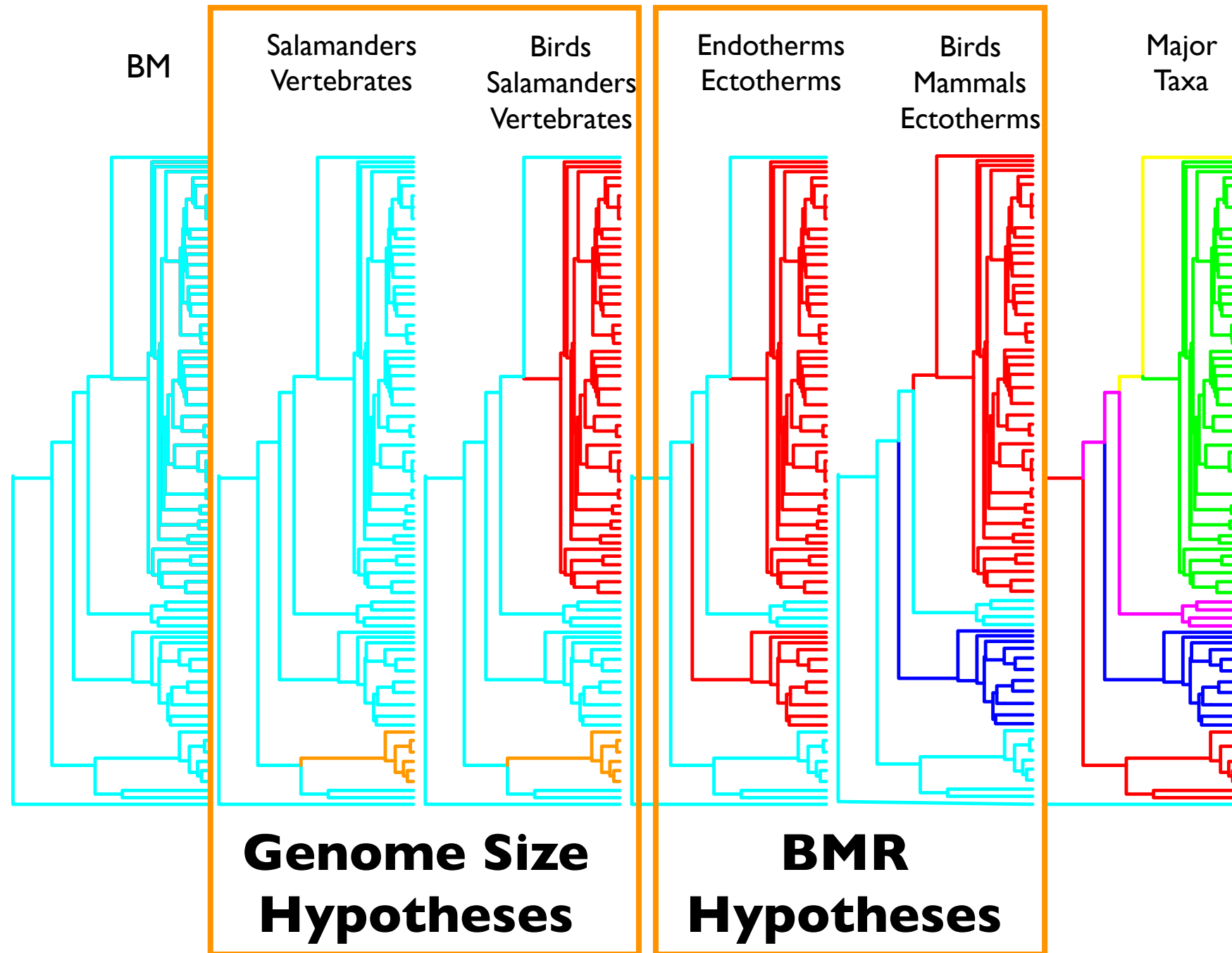
**BMR
Hypotheses**

Alternative Adaptive Regime Models

BM

Salamanders
Vertebrates

Birds
Salamanders
Vertebrates

Endotherms
Ectotherms

Birds
Mammals
Ectotherms

Major
Taxa

Major
Taxa
Salamanders

**Genome Size
Hypotheses**
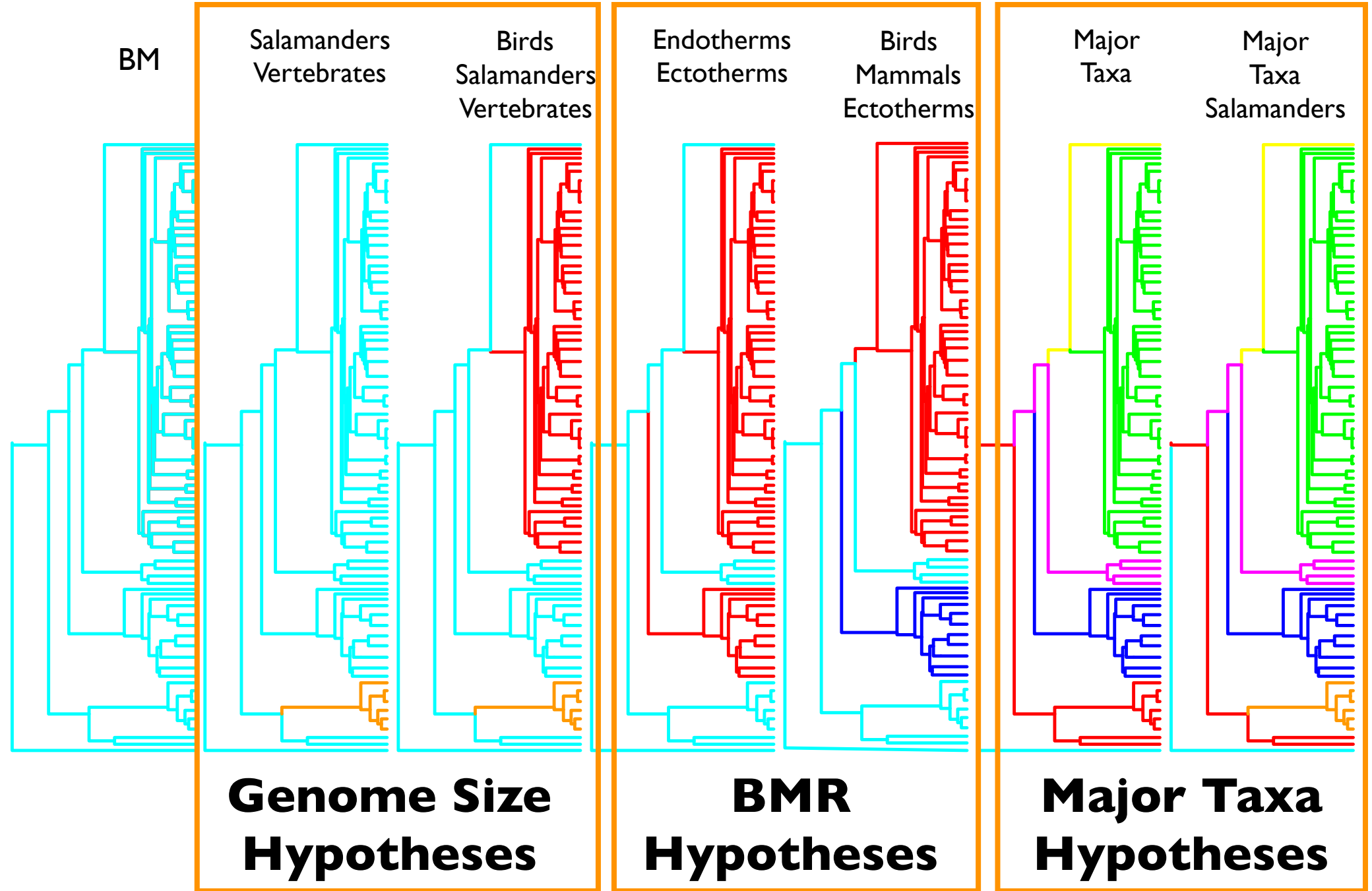
**BMR
Hypotheses**

Alternative Adaptive Regime Models

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

simplest bivariate model: no off-diagonal terms
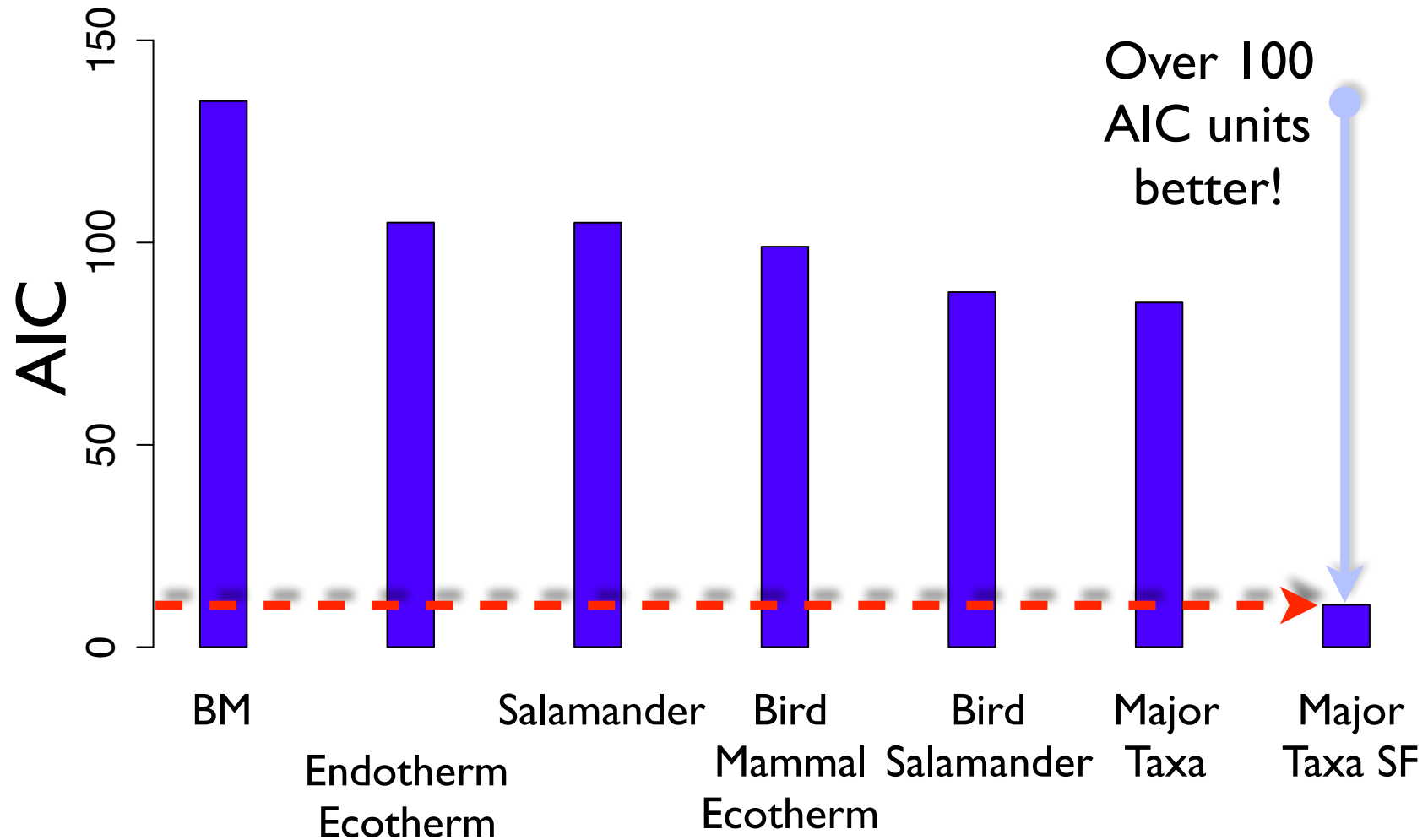
Over 100 AIC units better!

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

correlated drift: off-diagonal *sigmas* improve fit
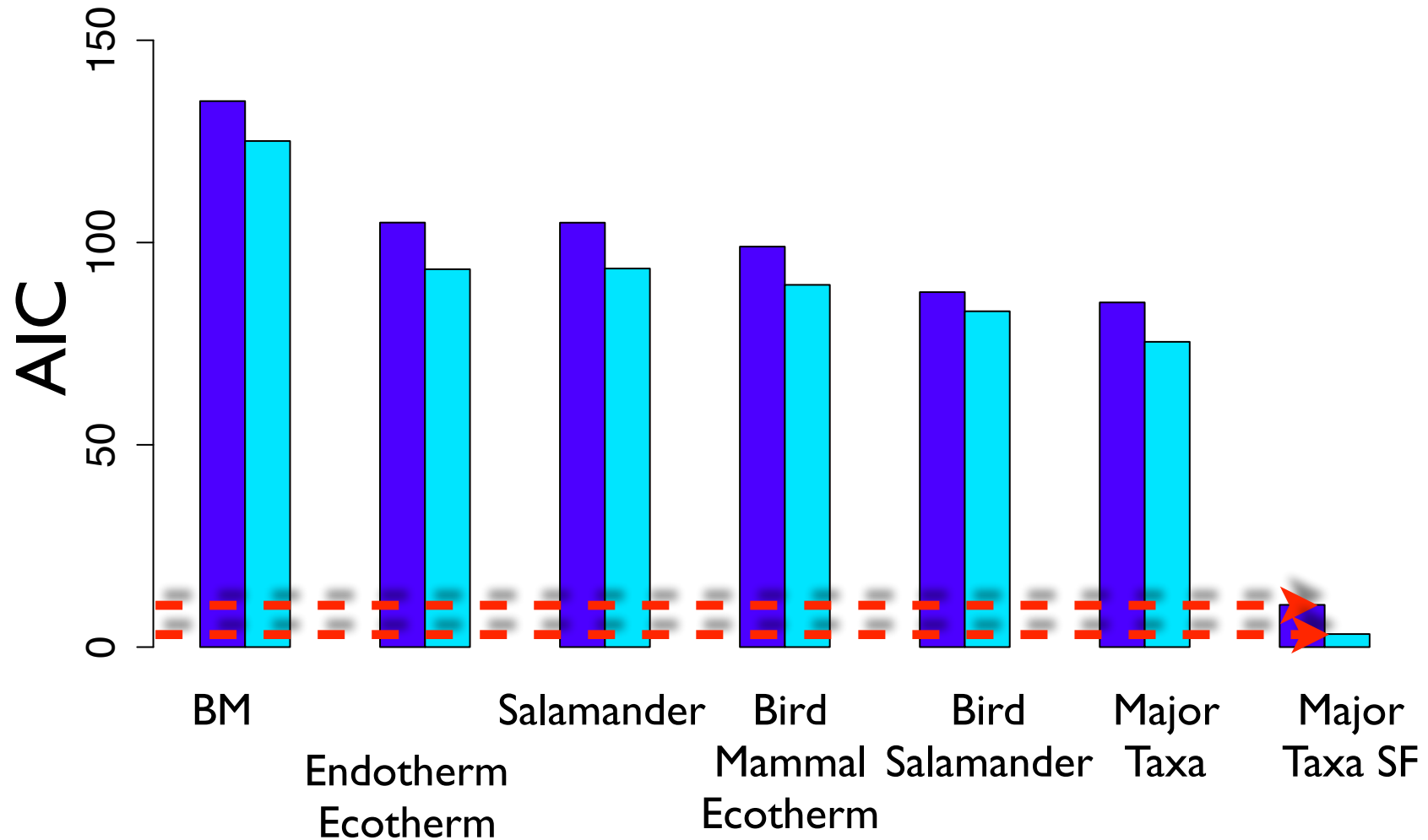
Separate adaptive regimes for each Major Taxon (splitting salamanders & frogs) was the Best Model

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

AIC

150

100

Endotherm
Ecotherm

Ecotherm
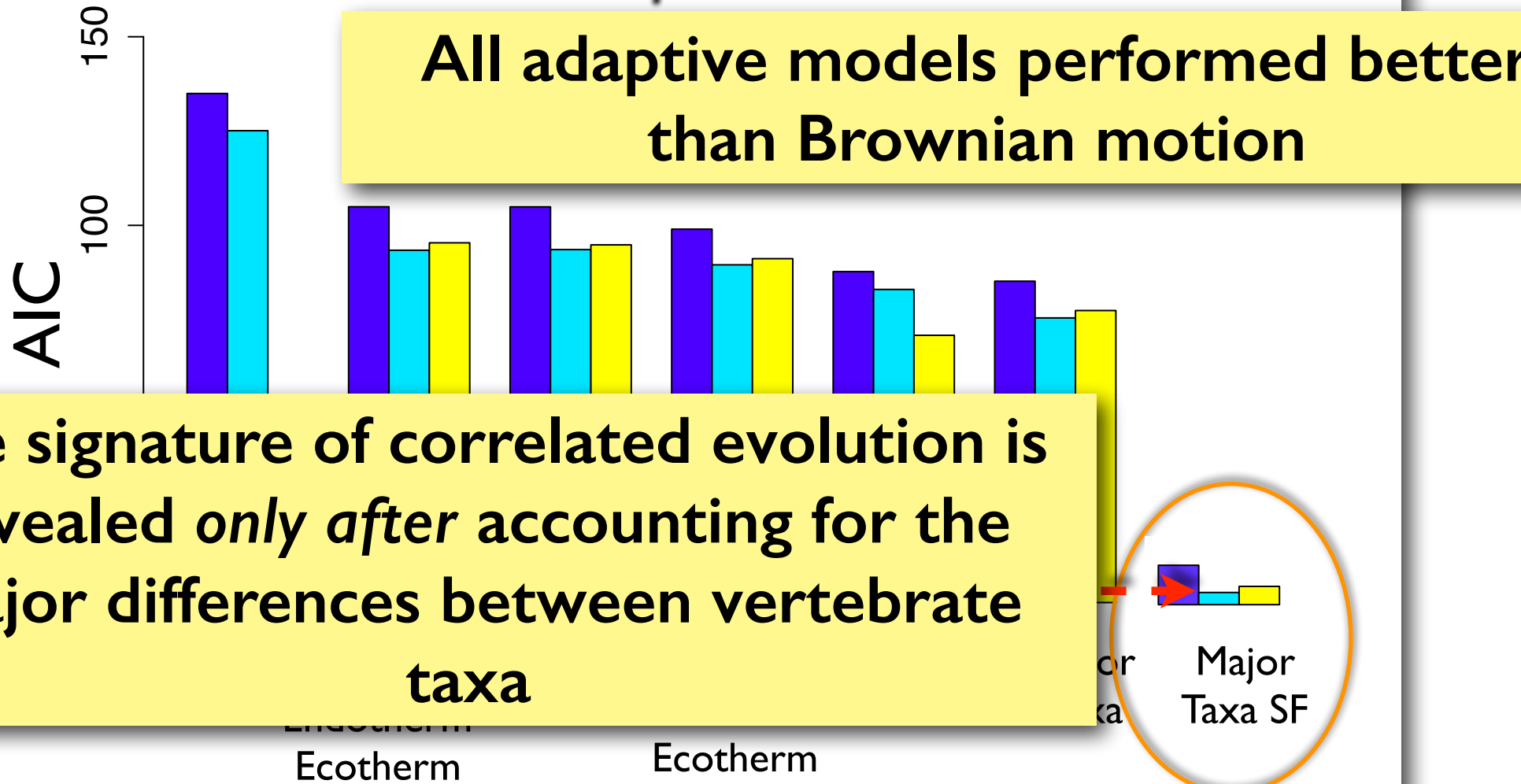
Major
Taxa SF
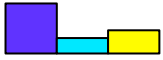
Results

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

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

Results

Profile Likelihoods for parameter estimates

$\sigma_{gs,bmr}$

negative

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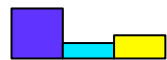
$$\alpha = \begin{pmatrix} 30.8 & 0 \\ 0 & 4.8 \end{pmatrix} \sigma = \begin{pmatrix} 1.1 & 0 \\ -0.44 & 0.91 \end{pmatrix}$$

α_{gs}

very large!

α_{bmr}

large
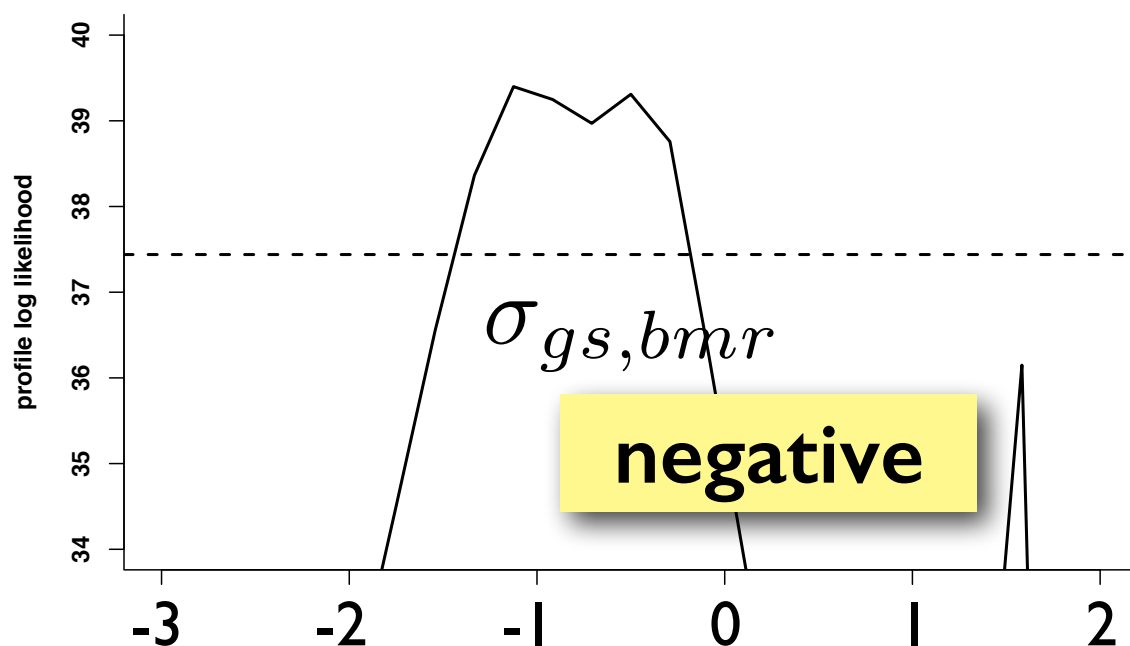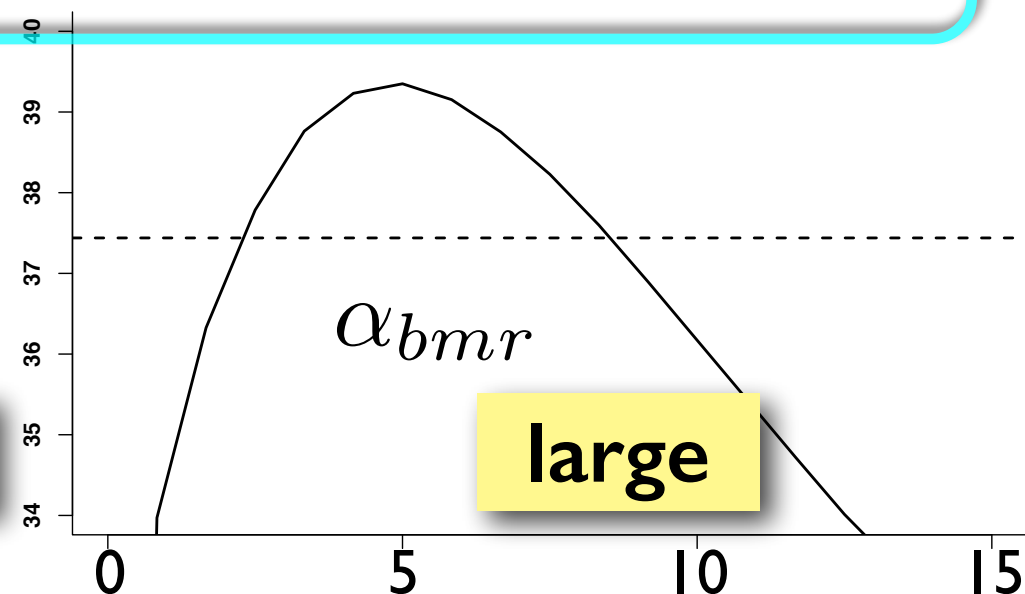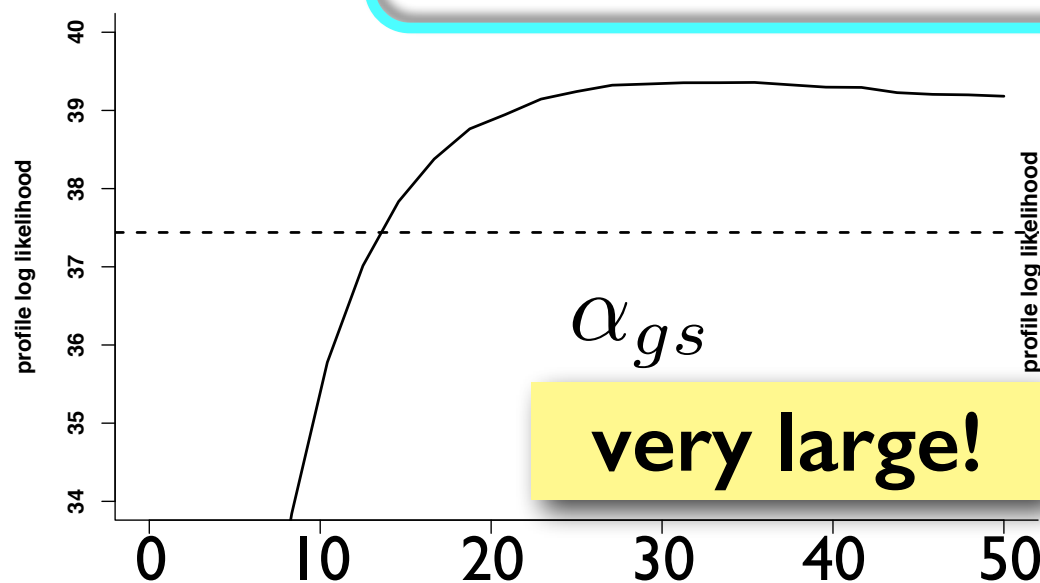
So... how to explain these data?

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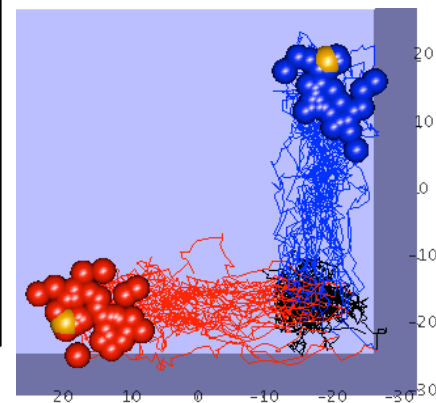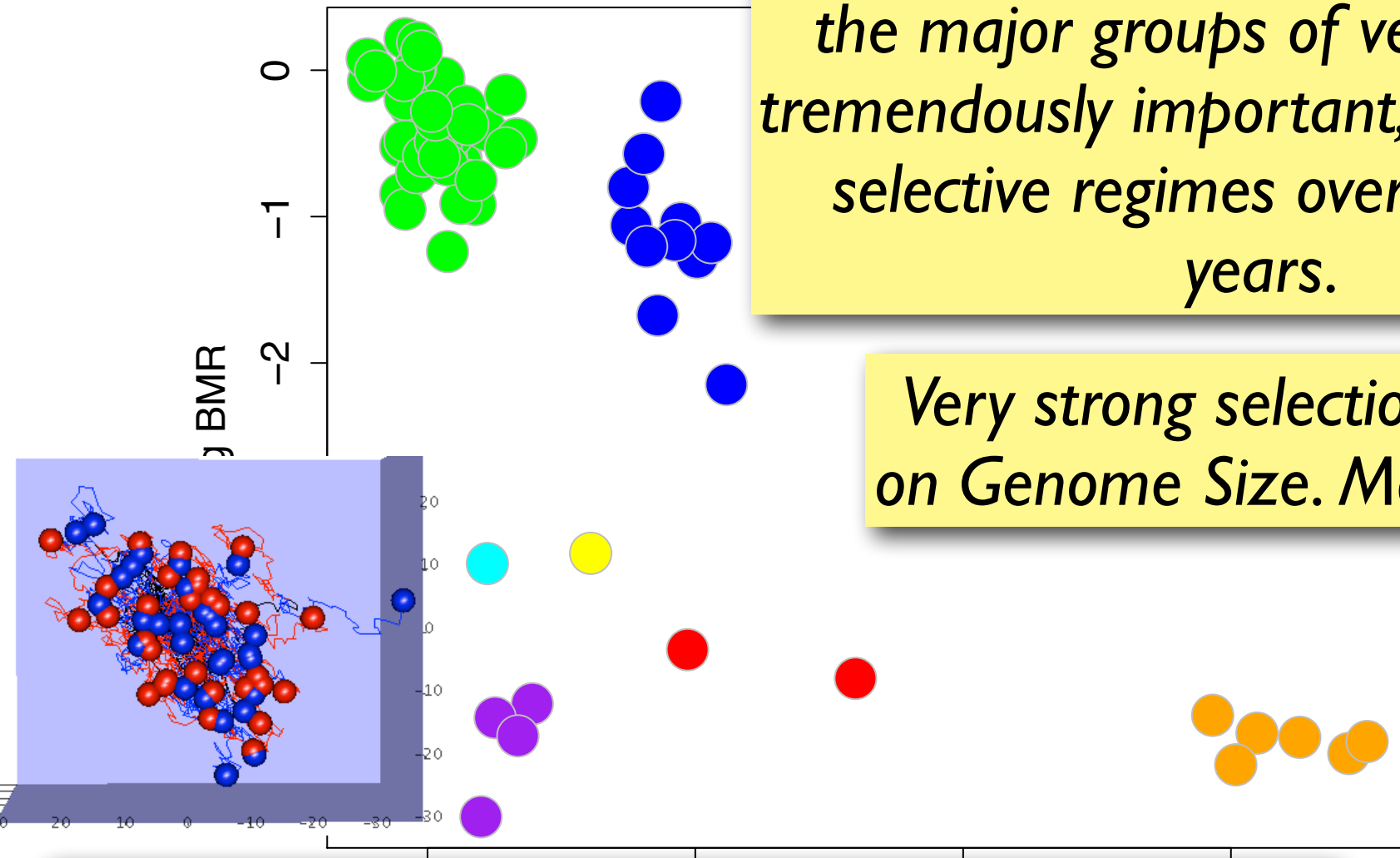
After accounting for major differences among taxa, strong signature of correlated evolution

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

drift

selection

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

Some Papers

Ané C. 2008. Analysis of comparative data with hierarchical autocorrelation. Ann. Appl. Stat. 2:1078–1102.

Bartoszek K., Pienaar J., Mostad P., Andersson S., Hansen T.F. 2012. A phylogenetic comparative method for studying multivariate adaptation. J. Theor. Biol. 314:204–215.

Beaulieu J.M., Jhwueng D.-C., Boettiger C., O'Meara B.C. 2012. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. Evolution 66:2369–2383.

Boettiger C., Coop G., Ralph P. 2012. Is your phylogeny informative? Measuring the power of comparative methods. Evolution 66: 2240–2251.

Butler, M. A. and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. American Naturalist, 164:683–695. URL http: //www.journals.uchicago.edu/AN/journal/issues/v164n6/40201/40201.html.

Cressler C., Butler M.A., and King A. A. 2015. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. Sys. Bio. 64(6):953-968. DOI: 10.1093/sysbio/syv043

Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. Evolution, 51:1341–1351. Hansen T.F., Pienaar J., Orzack S.H. 2008. A comparative method for studying adaptation to a randomly evolving environment. Evolution 62:1965–1977.

Felsenstein, J. 1985. Phylogenies and the comparative method. American Naturalist, 125:1–15. URL http://links.jstor.org/sici?sici=0003-0147%28198501%29125% 3A1%3C1%3APATCM%3E2.0.CO%3B2-Y. Felsenstein J. 1973. Maximum-likelihood estimation of evolutionary trees from continuous characters. Am. J. Hum. Genet. 25:471–492. Felsenstein, J. 2004. Inferring Phylogenies. Sinauer, Sunderland, Mass.

Lande R. 1976. Natural selection and random genetic drift in phenotypic evolution. Evolution 30:314–334. Lande R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. Am. Nat. 114:463–479.

O'Meara, B. C., C. Ane, M. J. Sanderson, and P. C. Wainwright. 2006. Testing for different rates of continuous trait evolution using likelihood. Evolution, 60:922–933.

Many Thanks...

Eric Waltari and Scott Edwards

Jeff Scales, Jason Pienaar, Justin Waguarnery, and Julio Rivera

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

UNIVERSITY OF MICHIGAN
ECOLOGY AND EVOLUTIONARY BIOLOGY

Department of
ZOOLOGY
UNIVERSITY OF HAWAI'I at MANOA

Try OUCH (our software) - available on CRAN @ www.r-project.org !

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