

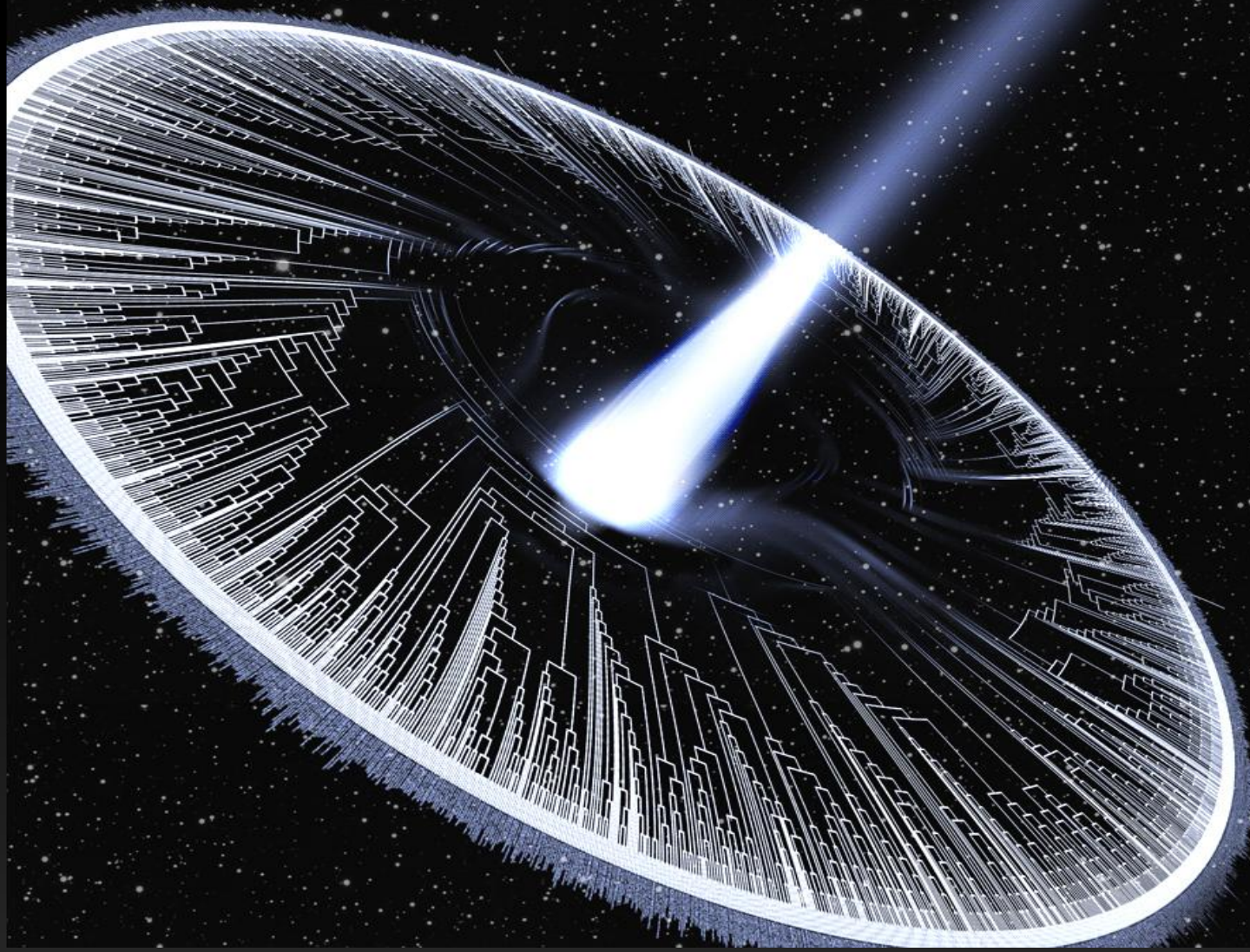
How phylogenies guide studies of trait inheritance & evolution

HDR Imageomics All-Hands

Josef Uyeda

Virginia Tech

juyeda@vt.edu



Outline

Lecture: Phylogenetic trait biology

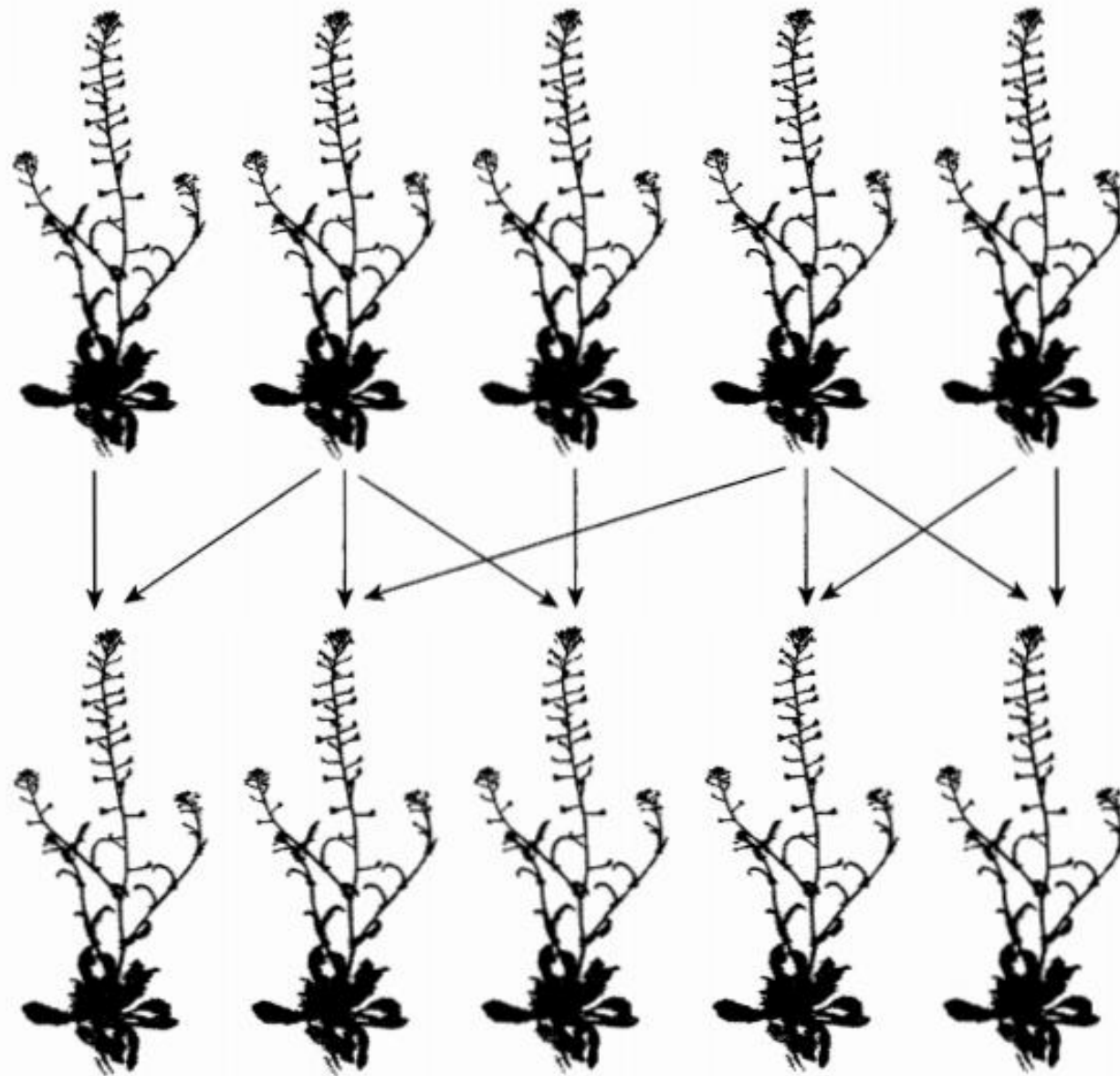
- Introduction to phylogenies and trait evolution
- Models of trait evolution
- Questions asked about trait evolution
- Phylogenetic signal
- The Phylogenetic Comparative Method

Computational Activity

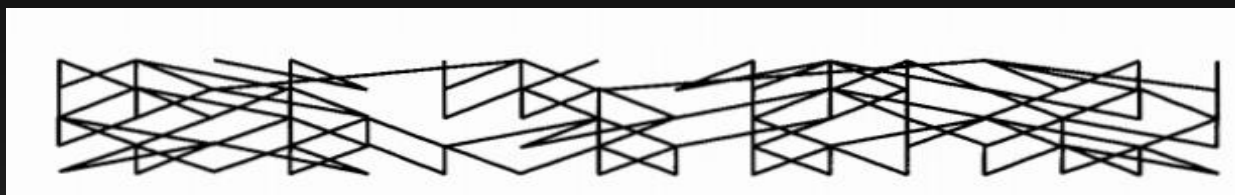
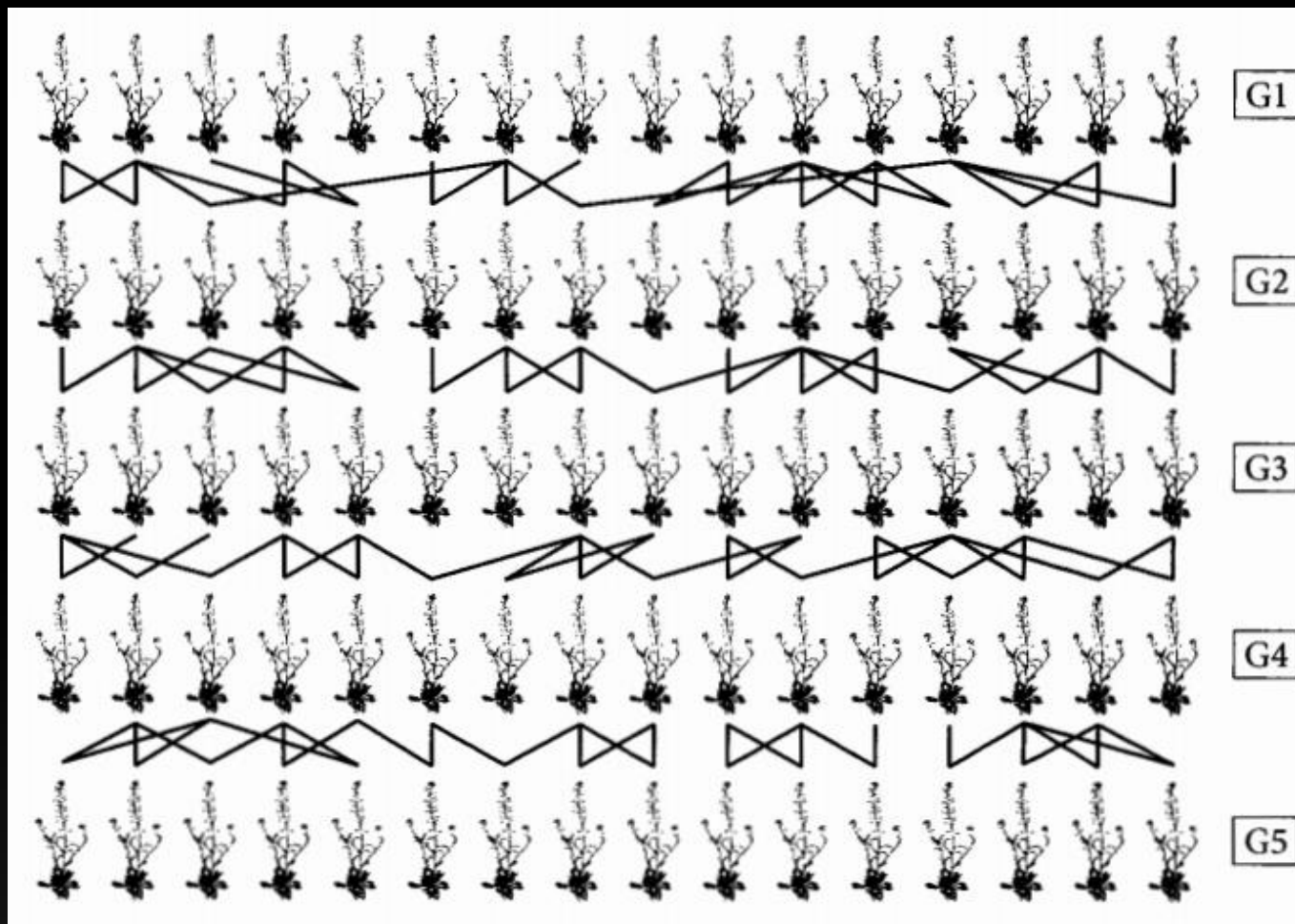
- Brownian Motion & Phylogenetic Signal - Univariate
- Geometric morphometrics
- PCA, PPCA, PACA - Aligning signal to phylogeny

Lecture: Character construction

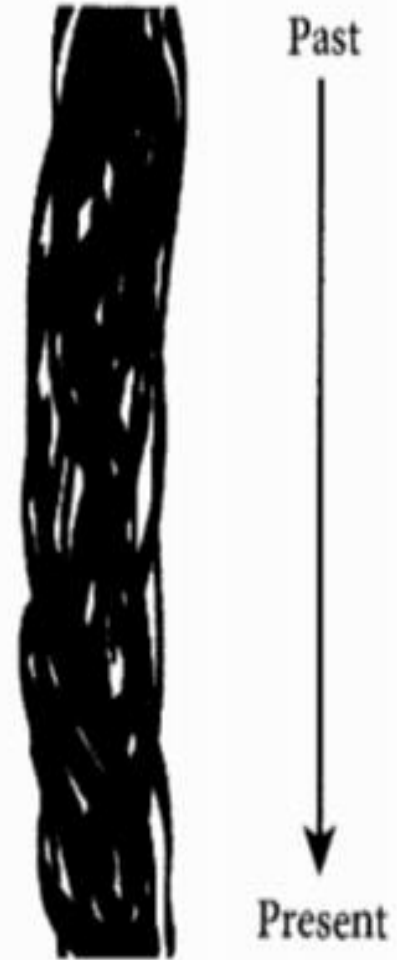
- Character construction in systematics
- Within & between species comparisons
- The two-scientist paradox



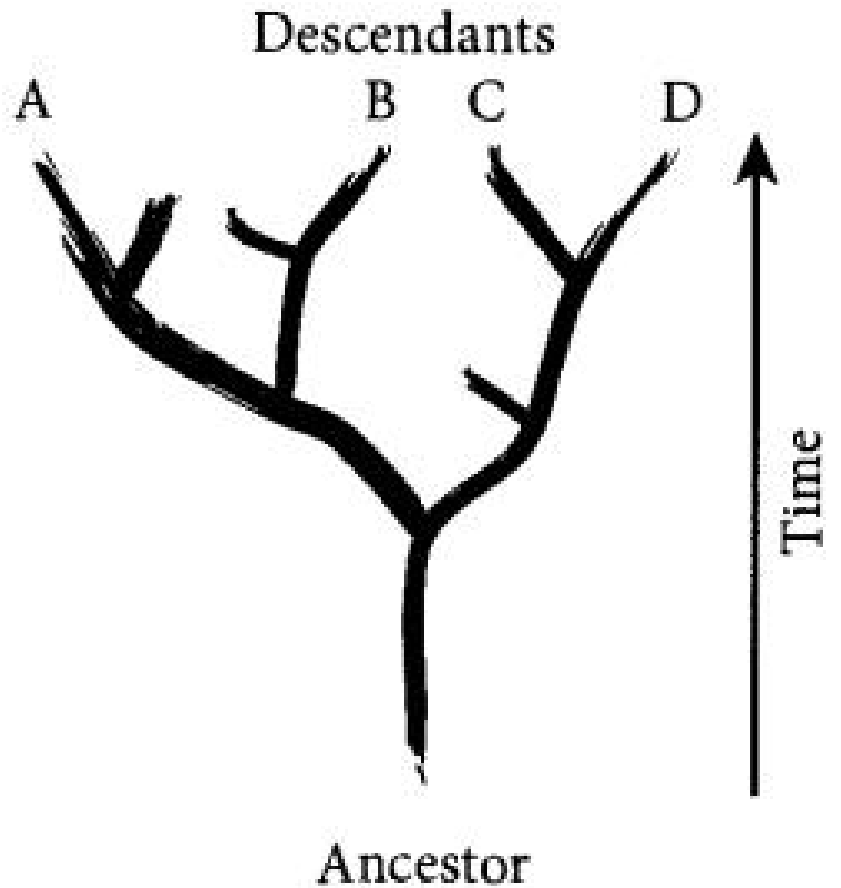
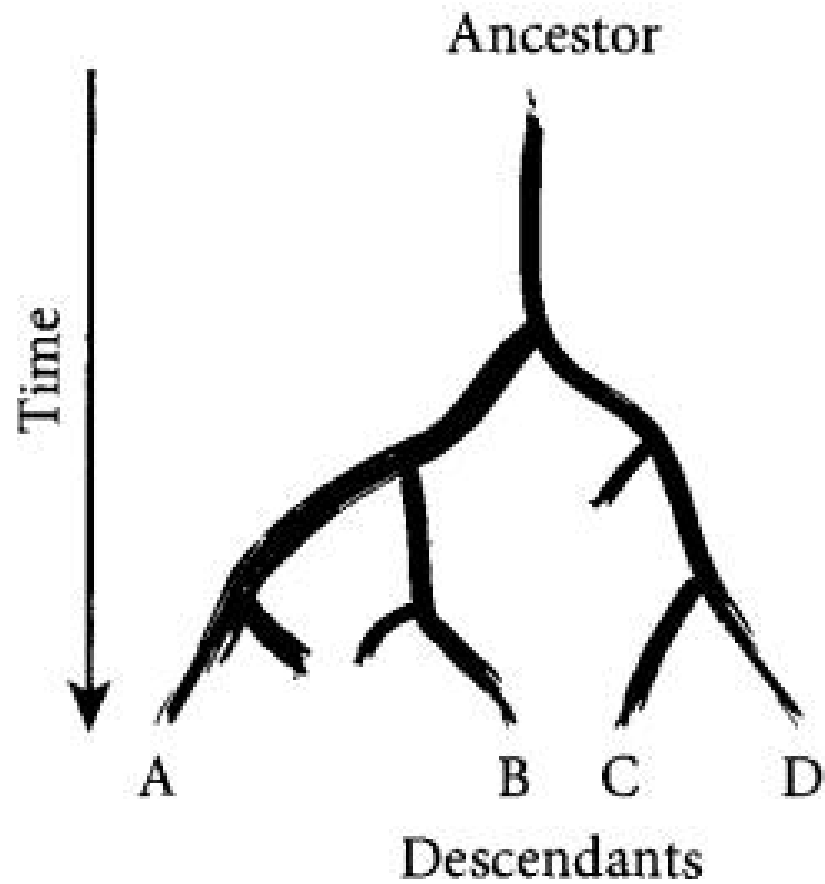
Baum & Smith, 2012



Baum & Smith, 2012



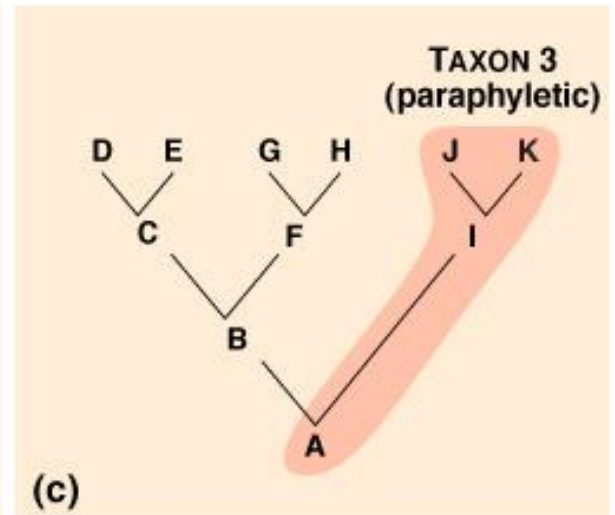
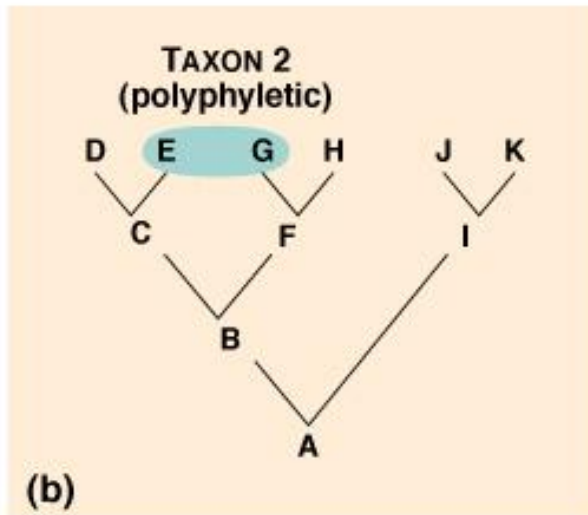
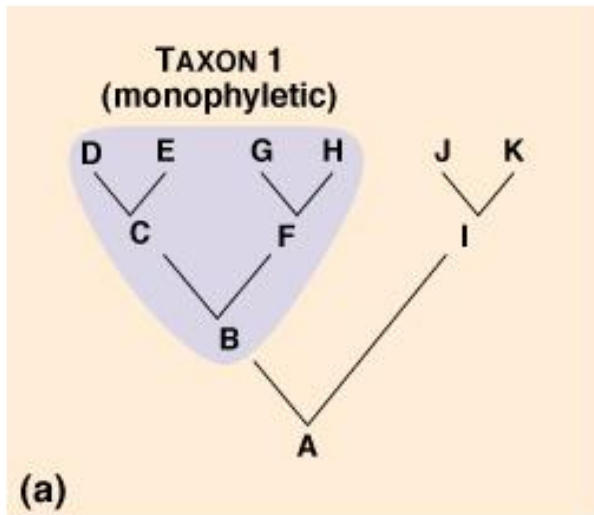
Baum & Smith, 2012



Baum & Smith, 2012

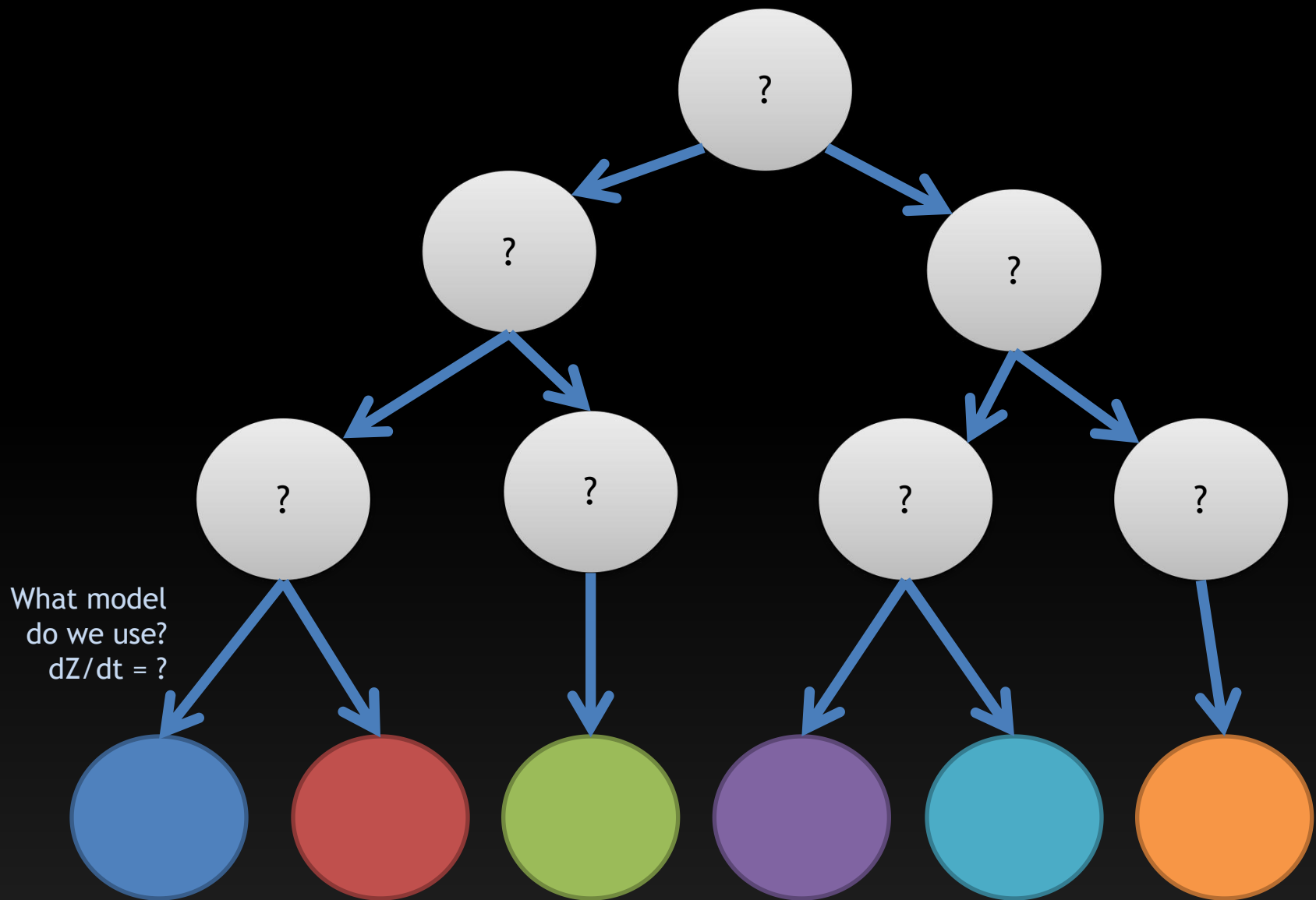
Taxonomy & Systematics

Goal: Defining monophyletic clades



Ingredients





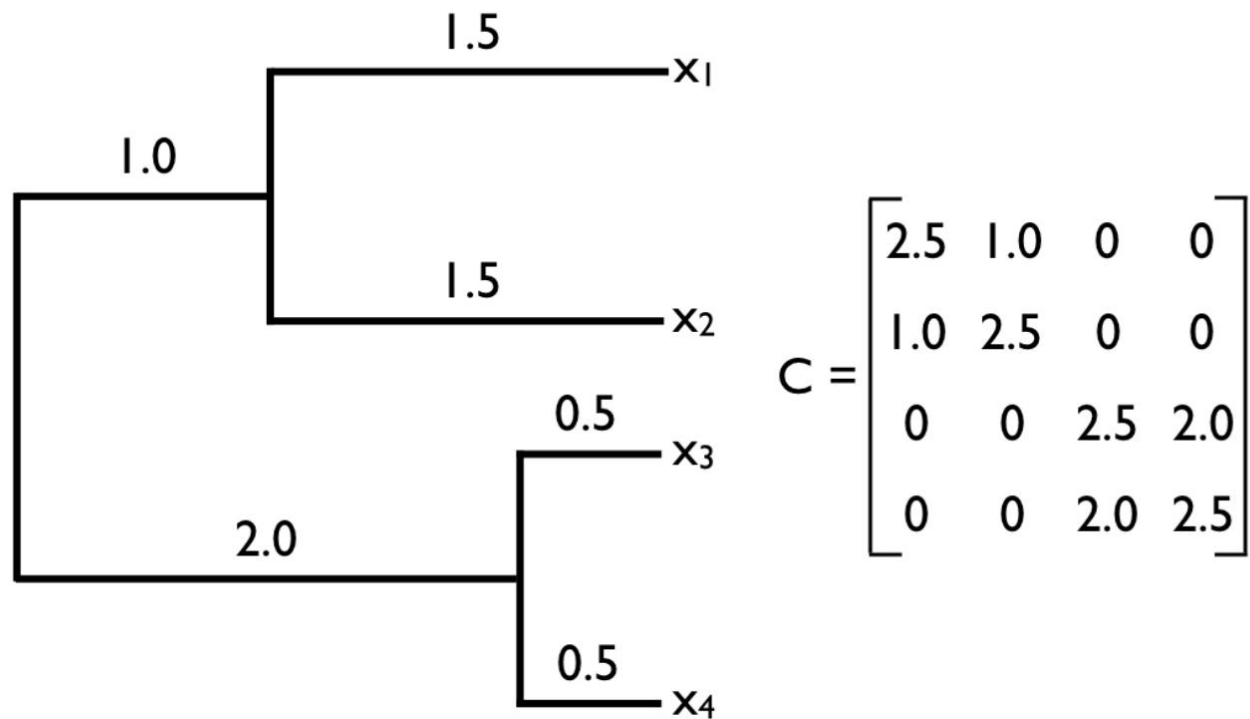


Figure 3.5. Example of a phylogenetic tree (left) and its associated phylogenetic variance-covariance matrix C (right).

Phylogenetics & PCMs basically use same 2 or 3 models for *everything*

Continuous traits:

Gaussian models (BM + extensions)

Discrete traits:

Continuous-Time Markov Chains (CTMCs)

Diversification:

Birth-Death models

Evolutionary Inferences from Phylogenies: A Review of Methods

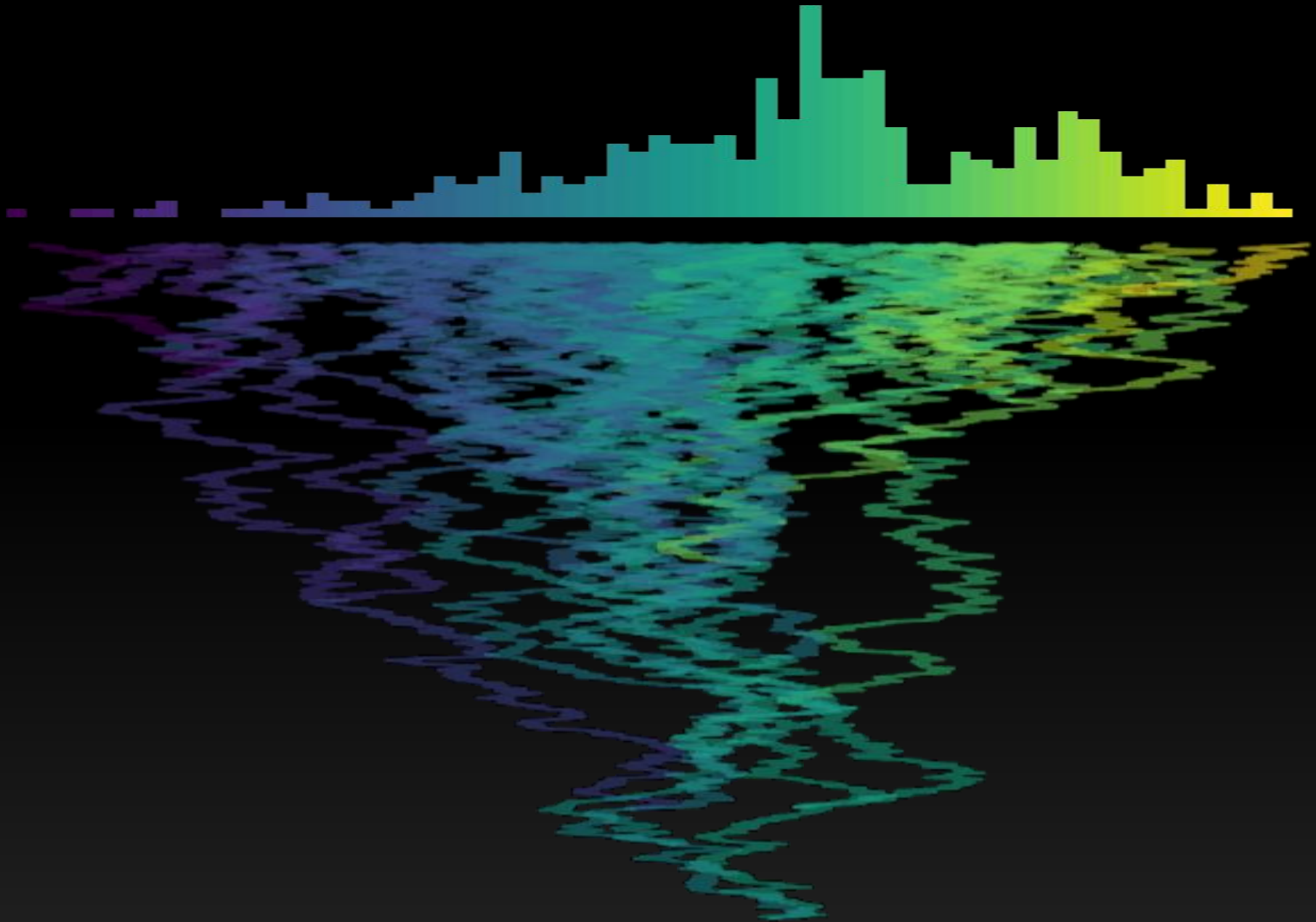
Annual Review of Ecology, Evolution, and Systematics

Vol. 43:267-285 (Volume publication date December 2012)

First published online as a Review in Advance on September 4, 2012

<https://doi.org/10.1146/annurev-ecolsys-110411-160331>

Brian C. O'Meara



Brownian Motion - continuous traits

Brownian Motion

Long history of usage in evolution

modeling genetic drift and random walks in fossil timeseries,
allele frequencies, and quantitative traits (Raup, Lande, Cavalli-Sforza)

Univariate:

Traits \sim mvnorm(root value, $\sigma^2 C$)

Multivariate:

Traits \sim mvnorm(root values, $R \otimes C$)

Discrete traits: Continuous time Markov Models

Q matrix

	A	B	C	D	E
A	-	r_{AB}	r_{AC}	r_{AD}	r_{AE}
B	r_{BA}	-	r_{BC}	r_{BD}	r_{BE}
C	r_{CA}	r_{CB}	-	r_{CD}	r_{CE}
D	r_{DA}	r_{DB}	r_{DC}	-	r_{DE}
E	r_{EA}	r_{EB}	r_{EC}	r_{ED}	-

Frequency vector

A	f_A
B	f_B
C	f_C
D	f_D
E	f_E

Can be adapted
to many kinds of
biological
scenarios

a GTR nucleotides

	A	G	C	T
A	-	r_{AG}	r_{AC}	r_{AT}
G	r_{AG}	-	r_{GC}	r_{GT}
C	r_{AC}	r_{GC}	-	r_{CT}
T	r_{AT}	r_{GT}	r_{CT}	-

b Binary correlation

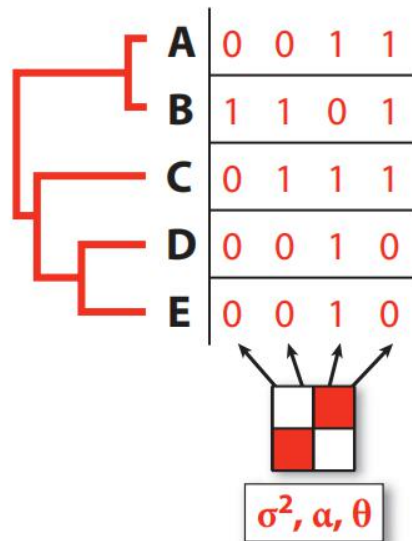
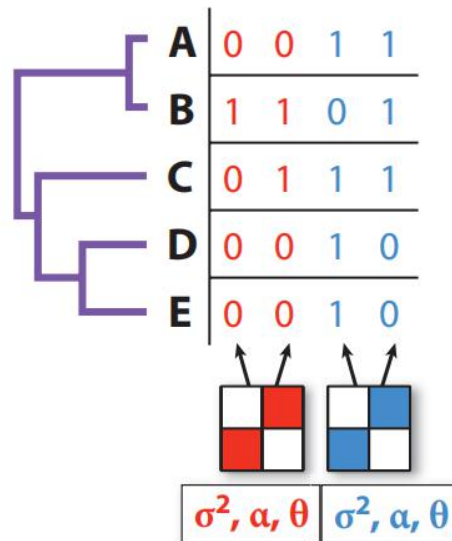
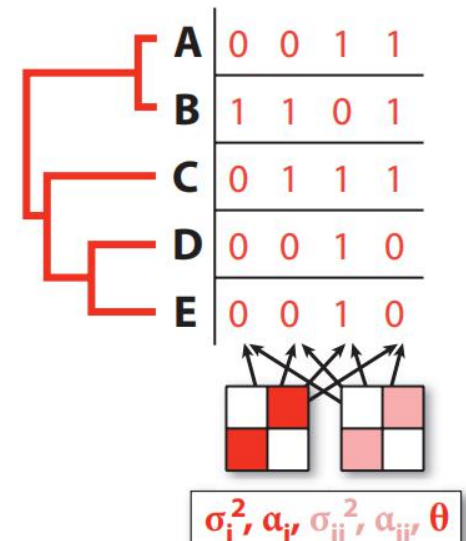
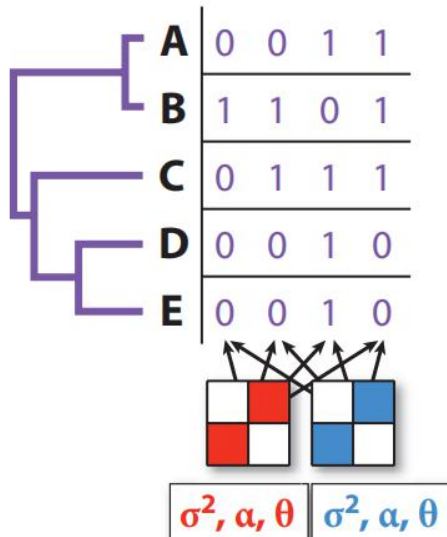
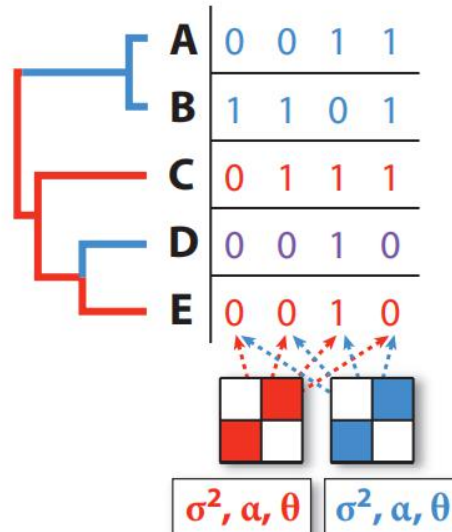
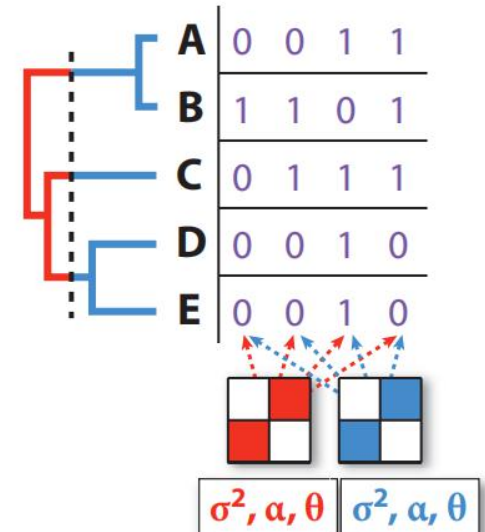
	00	01	11	10
00	-	r_A	0	r_B
01	r_C	-	r_D	0
11	0	r_E	-	r_F
10	r_G	0	r_H	-

c Covarion

	A+	T+	A-	T-
A+	-	r_{AT}	δ	0
T+	r_{TA}	-	0	δ
A-	$k\delta$	0	-	0
T-	0	$k\delta$	0	-

d Ordered transitions

	0	1	2	3
0	-	r_{01}	0	0
1	r_{10}	-	r_{12}	0
2	0	r_{21}	-	r_{23}
3	0	0	r_{32}	-

a No heterogeneity**b Partitioning by character****c Discrete gamma****d Mixture model****e Branch heterogeneity****f Time heterogeneity**

I. Plethodontidae



II. Serpentes



III. Mammals



Slope

>0.80

0.77

0.73

0.70

0.66

I

II

I

II

III

**What are the biological
questions we ask?**

Phylogeny + single trait we can ask:

Is there phylogenetic signal?

What is the model that best captures the transmission of information down the tree?

Is there evidence of constraints/adaptation?

What are the ancestral states?

Is the rate of evolution constant across branches?

Does the model change in different parts of the tree? ...across different sets of traits?

Does change accumulate early or late in the tree?

i.e. Questions about Tempo & Mode of evolution

“Phylogenetic Signal”

Deceptively simple idea

Many different statistics used

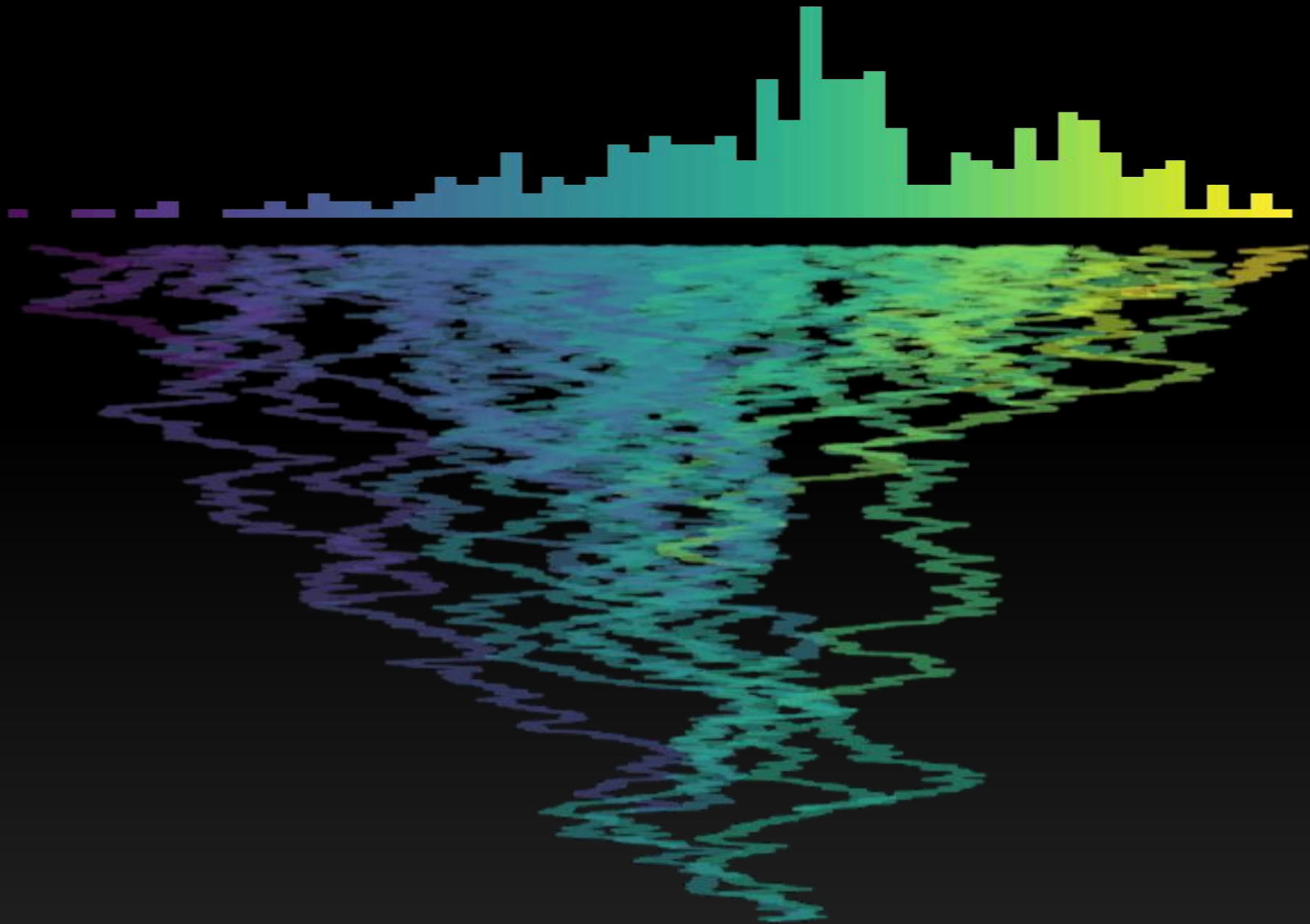
Phylogenetic heritability

Pagel's Lambda

Blomberg's K

Phylogenetic half-life

Phylogenetic-D



Measured relative to BM

Why doesn't every trait have phylogenetic signal?

Saturation of state space (e.g. too many mutations in DNA)

Multiple hits/convergence (e.g. Heliconius butterflies)

Constrained evolution

Adaptation to an optimal state

Strong environmental effects (basically convergence again)

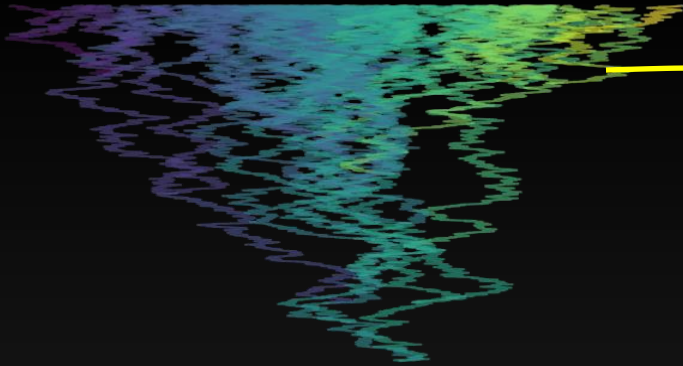
Incorrectly measuring/embedding states

...or even a few outliers

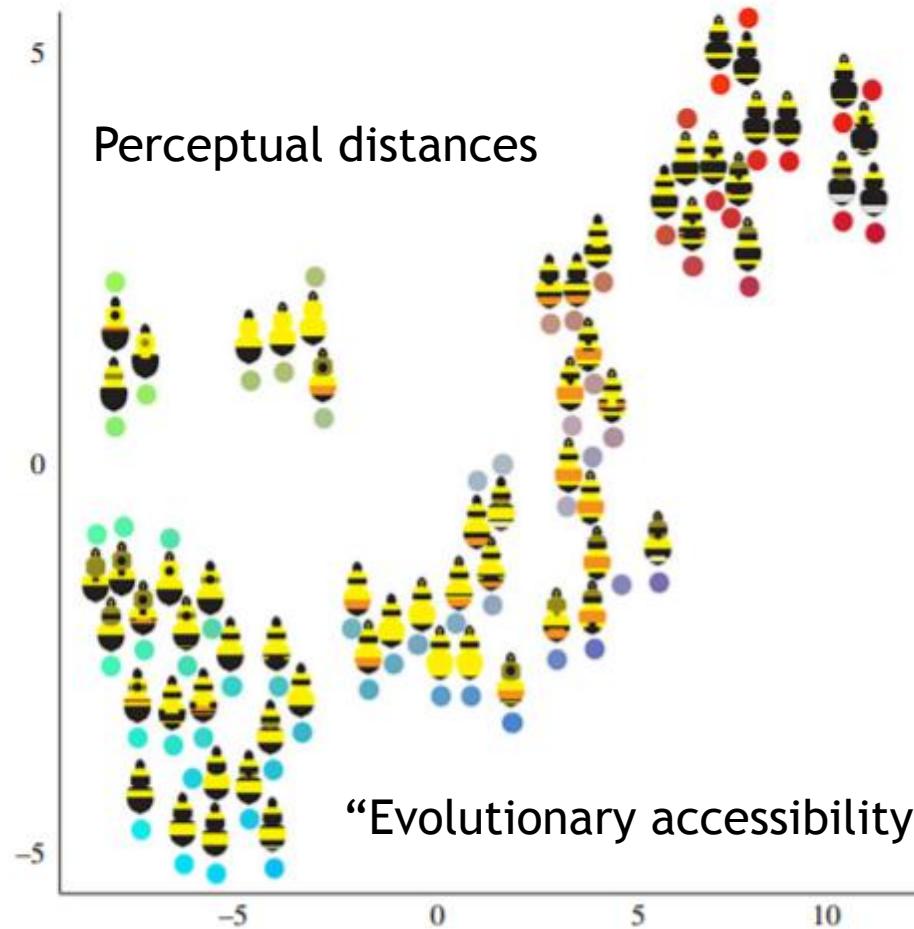
There's always a handful of weirdos in every dataset



There's always a handful of weirdos in every dataset

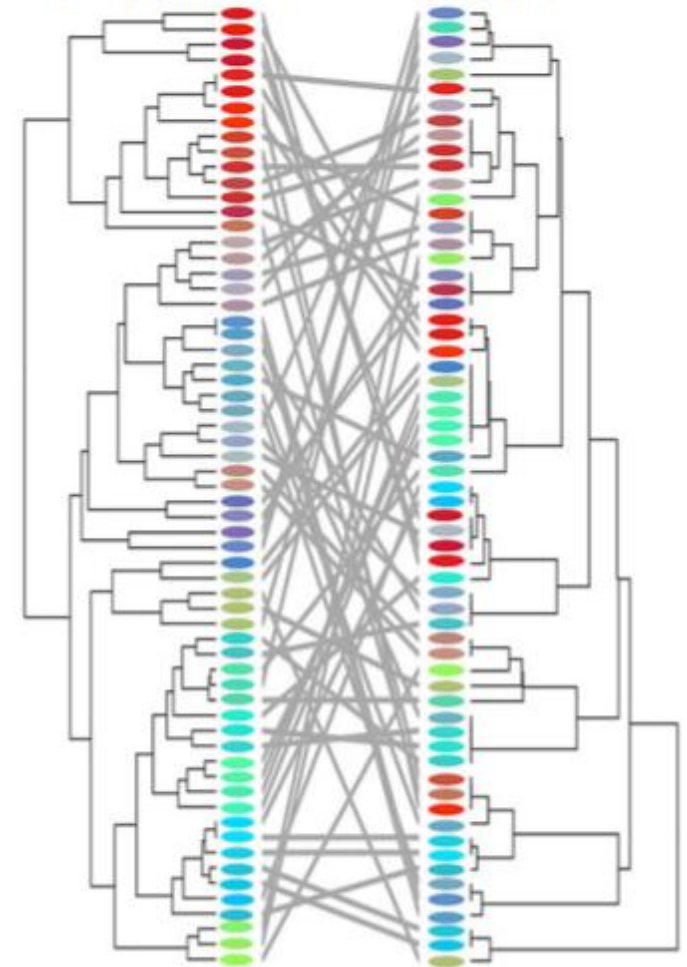


(a)



(b) perceptual colour

phylogeny



Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum

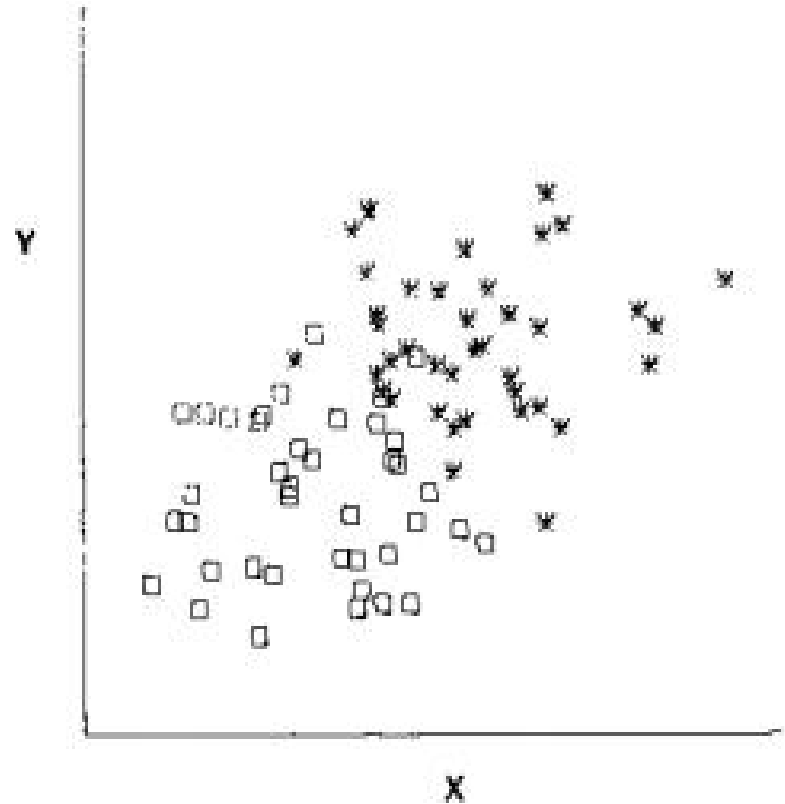
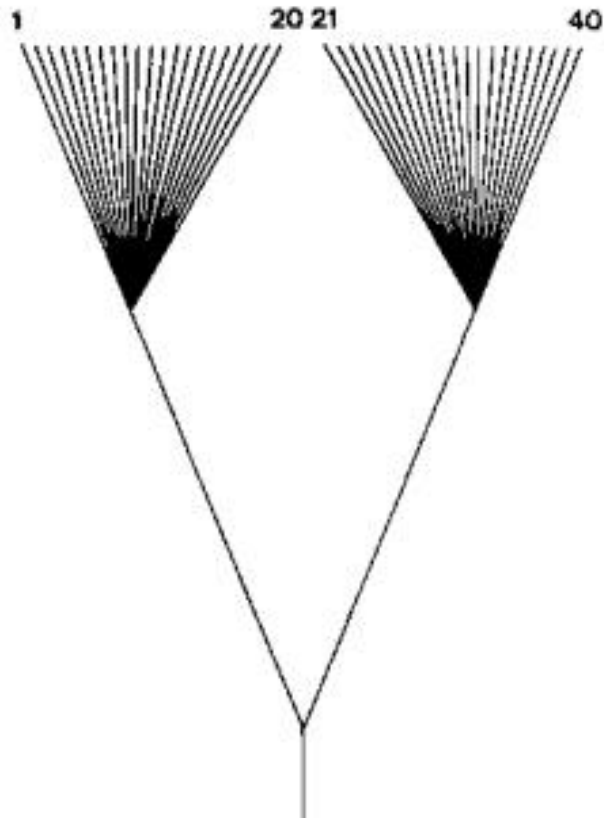
Briana D. Ezray¹, Drew C. Wham², Carrie E. Hill² and Heather M. Hines^{1,2}

“The Comparative Method”

Studying trait associations

**What causes
adaptations?**





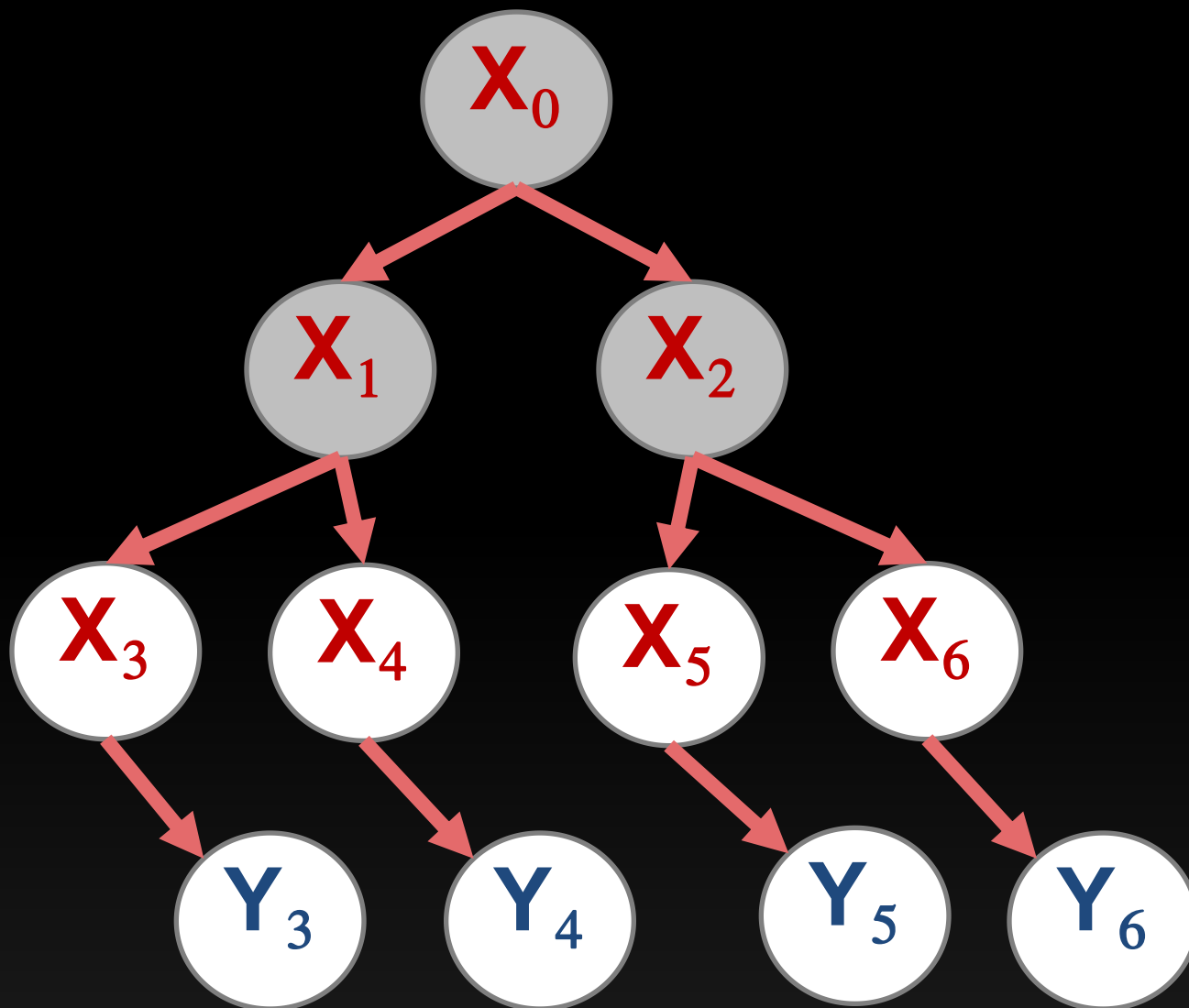
Felsenstein 1985

Why can't we do "regular" linear regression again?



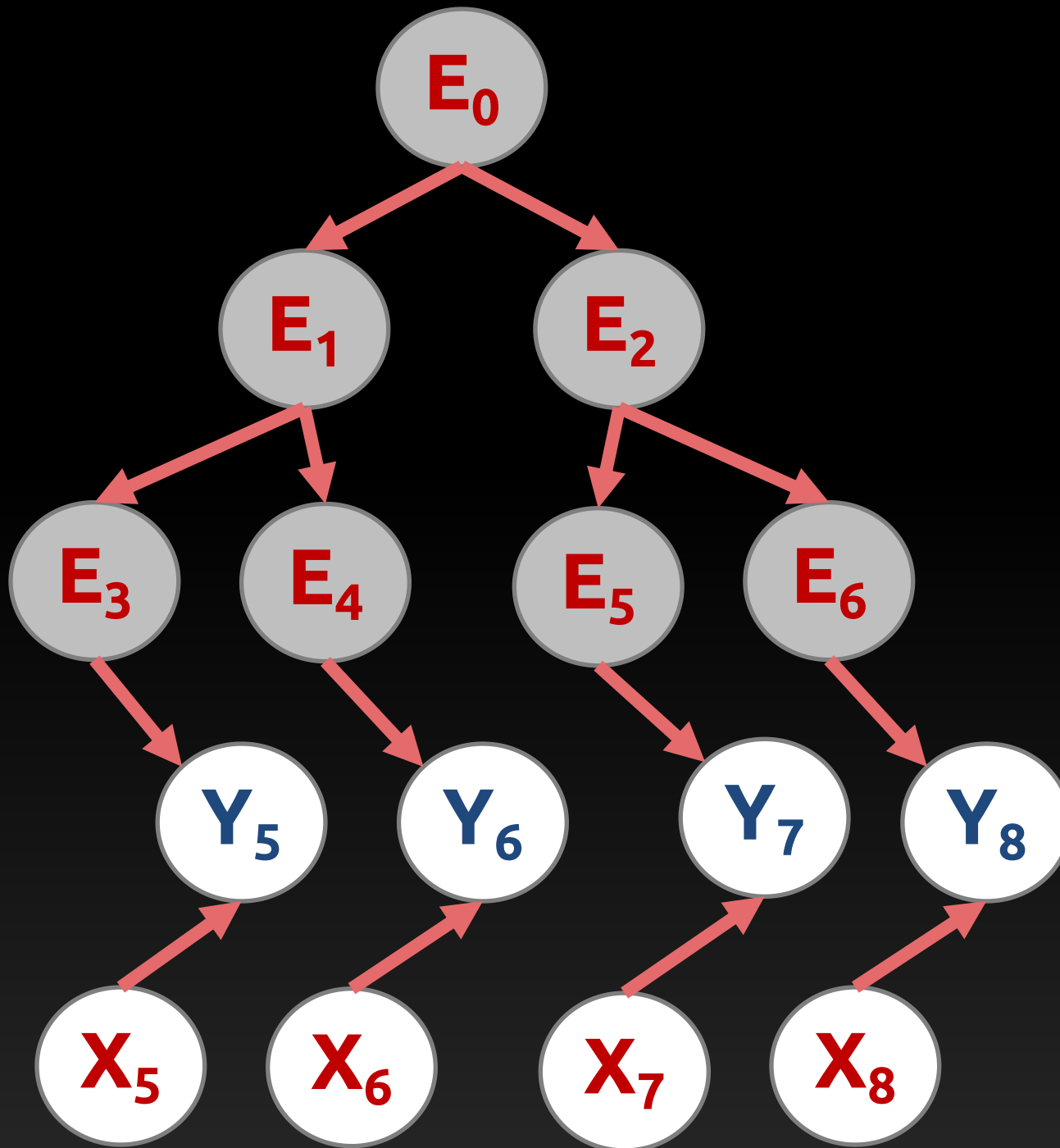
**Non-independence
of species**

$$Y \sim X$$



Why can't we do "regular" linear regression again?

**Non-independence
of *residual error variation***

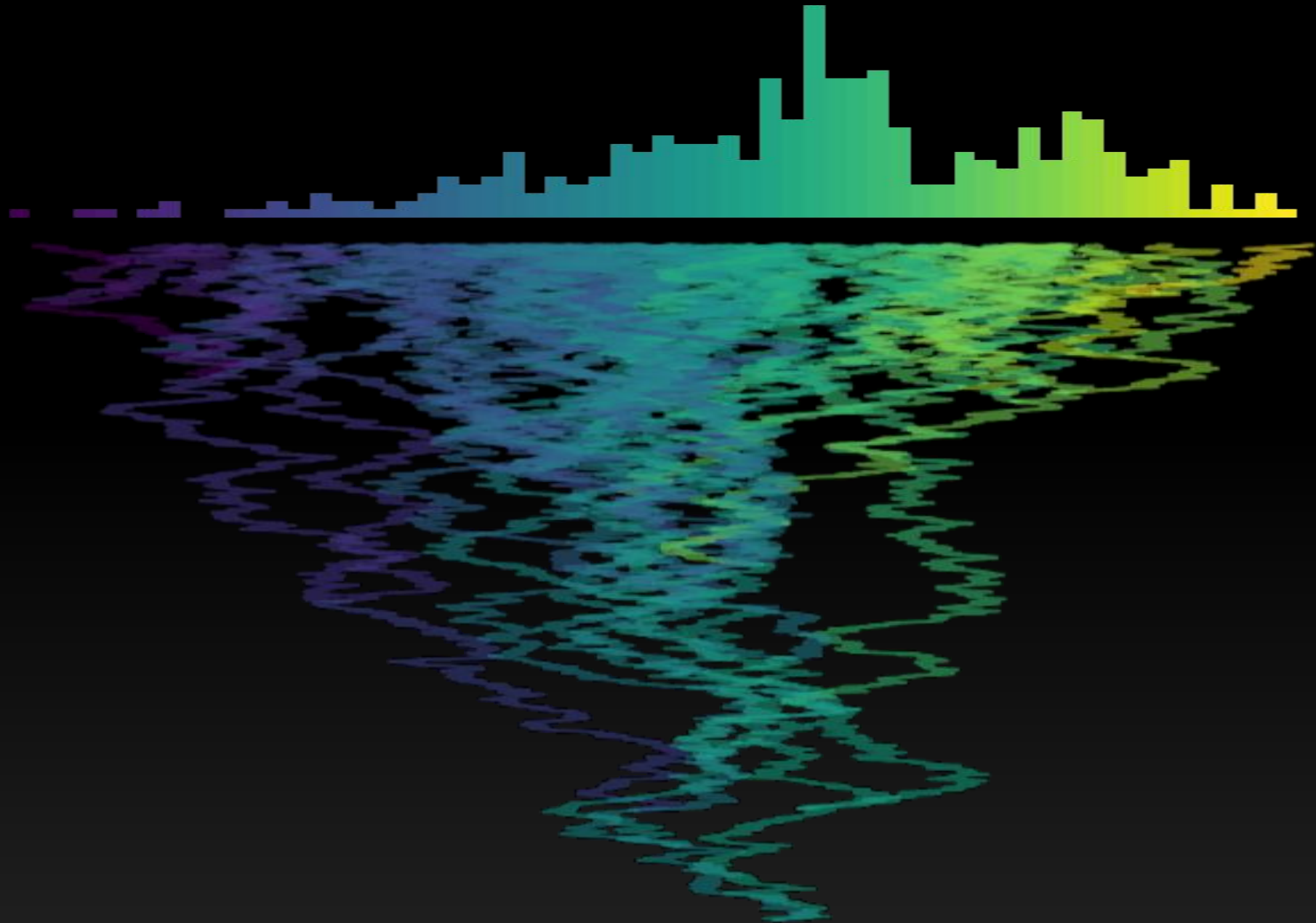


$$Y \sim X$$

**Pedigree/
Phylogeny
&
Model for random
effects
(e.g. Mendelian
inheritance or BM)**

**Response
traits**

**Predictor traits/
Treatment/
Fixed effects**



Brownian Motion

Questions asked about trait associations:

Are traits coevolving/correlated?

Does the relationship vary in different clades?

Is the relationship similar between and within species?

Is there an evolutionary time lag in response to the environmental change?

After accounting for predictors, is there any remaining phylogenetic signal in the residuals?

Part II - Character construction

Comparative methods are usually a two step process

1. Character construction - Measuring traits

Done idiosyncratically based on homology hypotheses of researcher

Not repeatable

Subject to observer bias

Redundancy/limited in scope

2. Character analysis & modeling

Phylogenetics, systematics & taxonomy

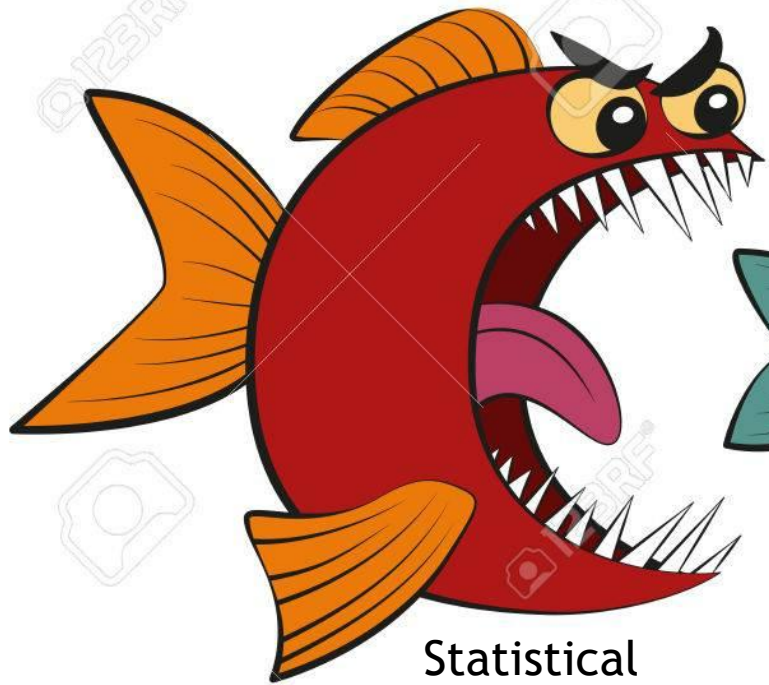
Phylogenetic comparative analyses

Character construction:

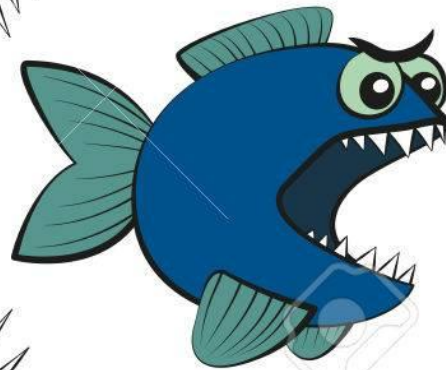
How we decide what traits to study and how we measure them



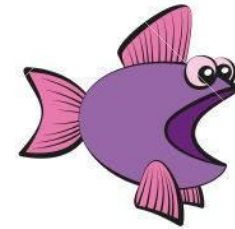
History of phylogenetic inference



Statistical
Phylogenetics



Cladistics



Phenetics



Evolutionary
Taxonomy

Cladistics

Model: Parsimony

**Define homologous characters that
characterize monophyletic groups**

**Convergence/homoplasy a mistake
in character construction**

Homology

Similarity due to common ancestry

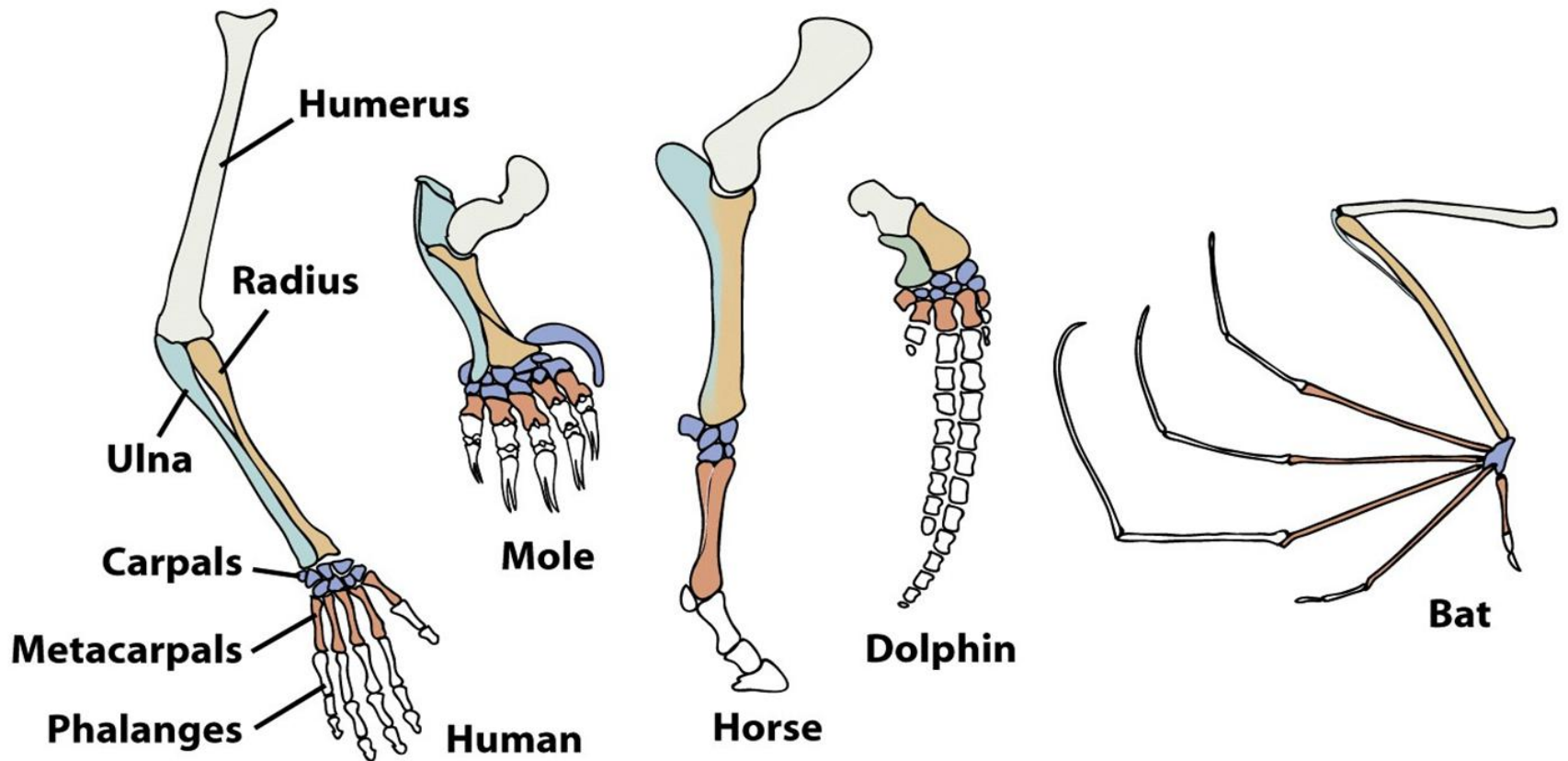


Figure 2-18 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Statistical Phylogenetics (for DNA)

Model: Continuous-Time Markov Model

**Align homologous DNA sequences &
assume sites evolve according to CTMC**

**Homoplasy/saturation an expected
outcome to be modeled**

Statistical Phylogenetics (phenotypes)

Model: CTMC (Mk) model

Align homologous phenotypes

**Hybrid approach to defining
homology & mixed feelings
about homoplasy....**

```
R> library(rphenoscape)
```

```
R> get_phenotypes("caudal fin")
```

"uroneural 1 increased length"

"uroneural 1 decreased size"

"uroneural 1 absent"

"uroneural 2 elongated"

"uroneural 2 increased length"

"uroneural 2 increased size"

"uroneural decreased size"

"uroneural increased length"

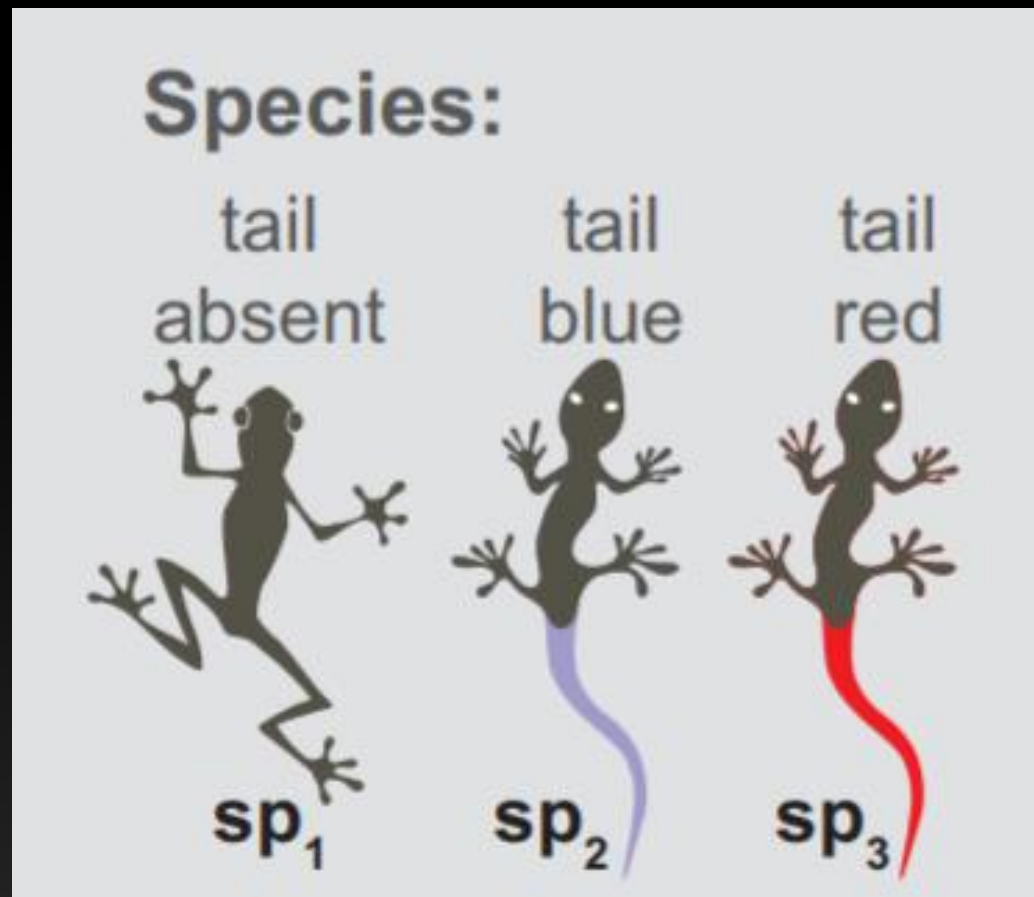
"uroneural fused with upper hypural set"

"uroneural fused with hypural"



**The “unseen” work of trait
biology is deciding how to
define and measure/code
characters**

Maddison's Tail Color Problem

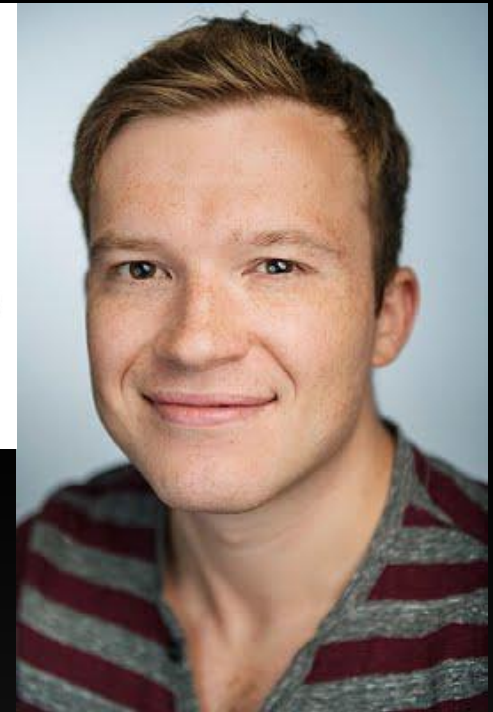


Integration of Anatomy Ontologies and Evo-Devo Using Structured Markov Models Suggests a New Framework for Modeling Discrete Phenotypic Traits

SERGEI TARASOV^{1,2}

¹National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996, USA; and ²Department of Biological Sciences, Virginia Tech, 4076 Derring Hall, 926 West Campus Drive, Blacksburg, VA 24061, USA
E-mail: sergxf@yandex.ru.

Received 19 September 2017; reviews returned 6 January 2019; accepted 15 January 2019
Associate Editor: Emma Goldberg

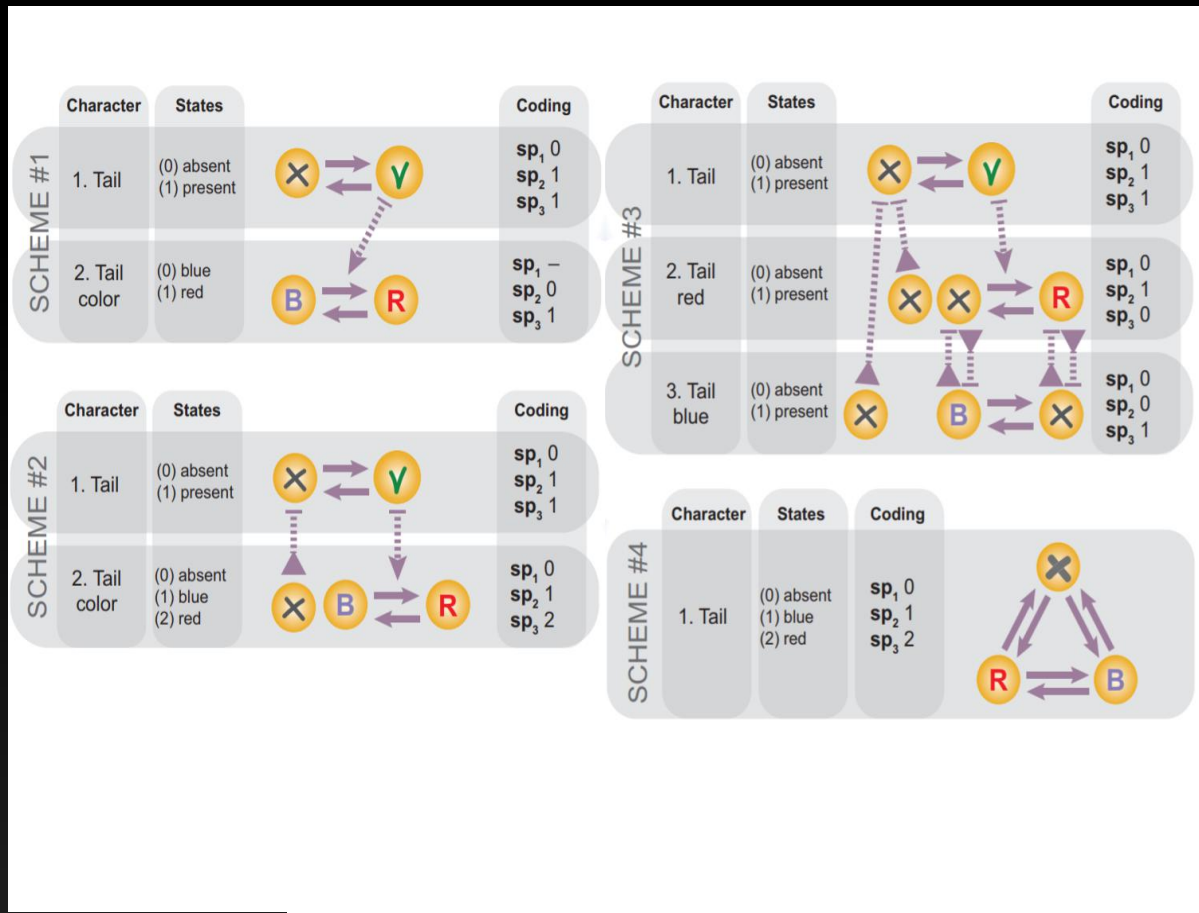


Sergei Tarasov
Finnish Museum of
Natural History

Fixing CTMC for Evo-Devo

1. Structured Markov Models:
Dependencies among traits
2. Hidden-State Markov Models:
Unobserved states with same
observable state

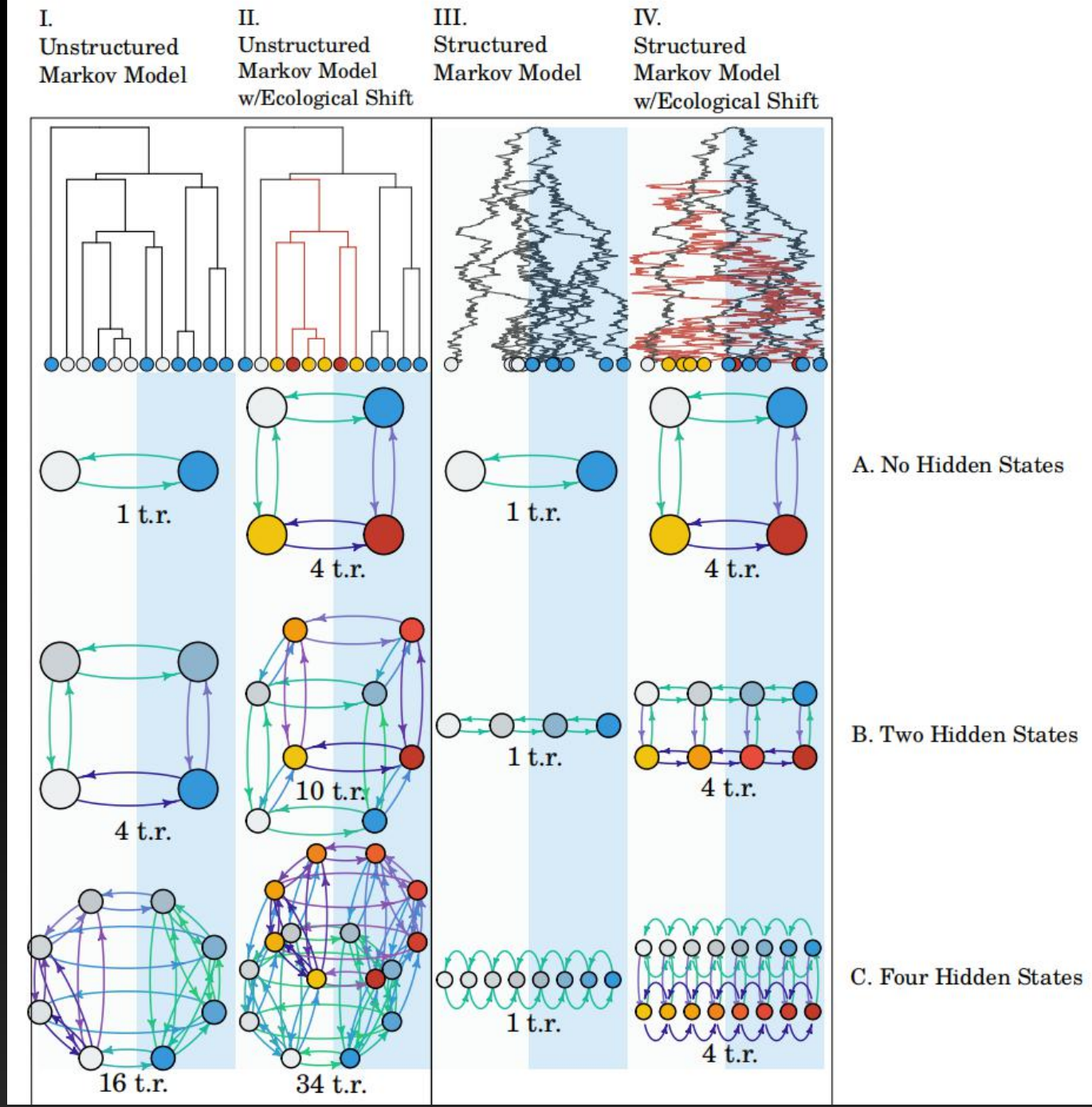
Maddison's Tail Color Problem



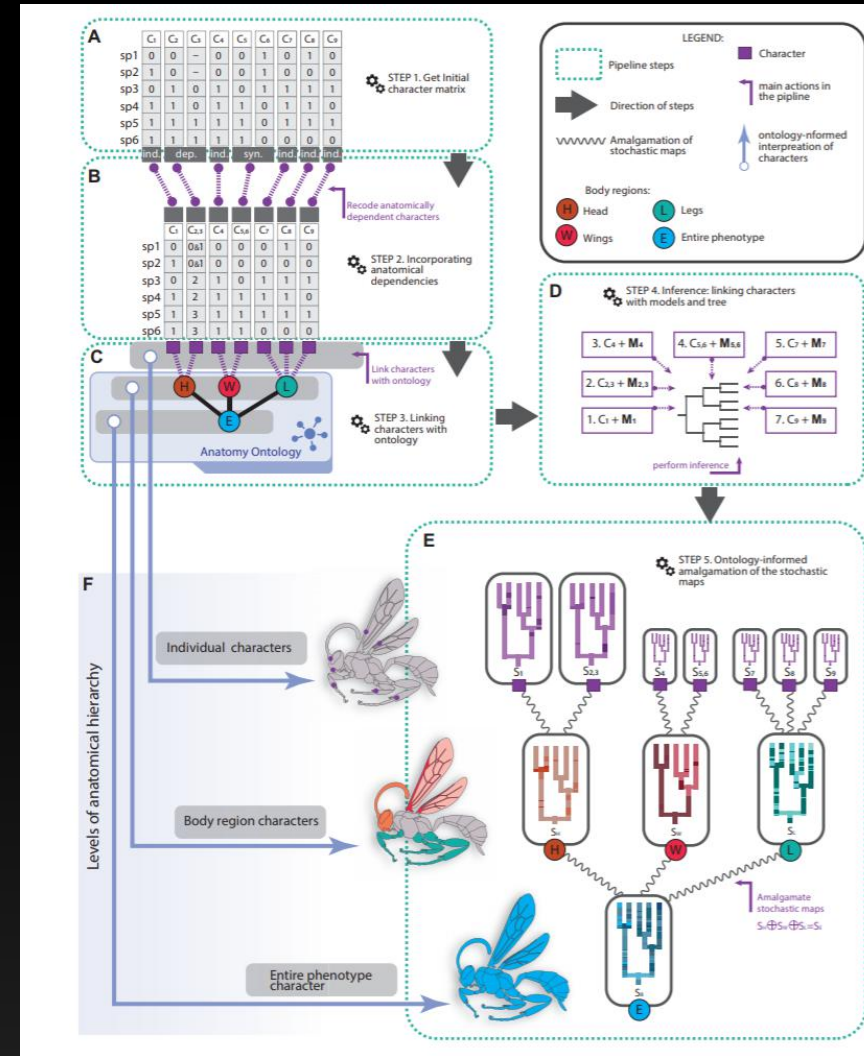
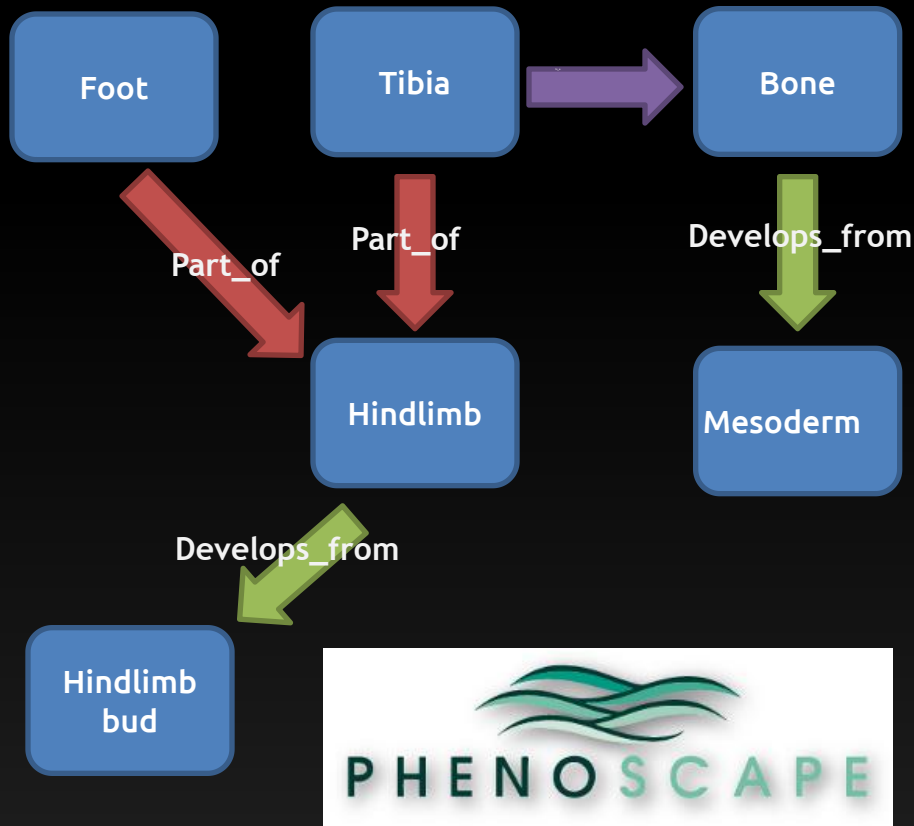
Discrete traits:

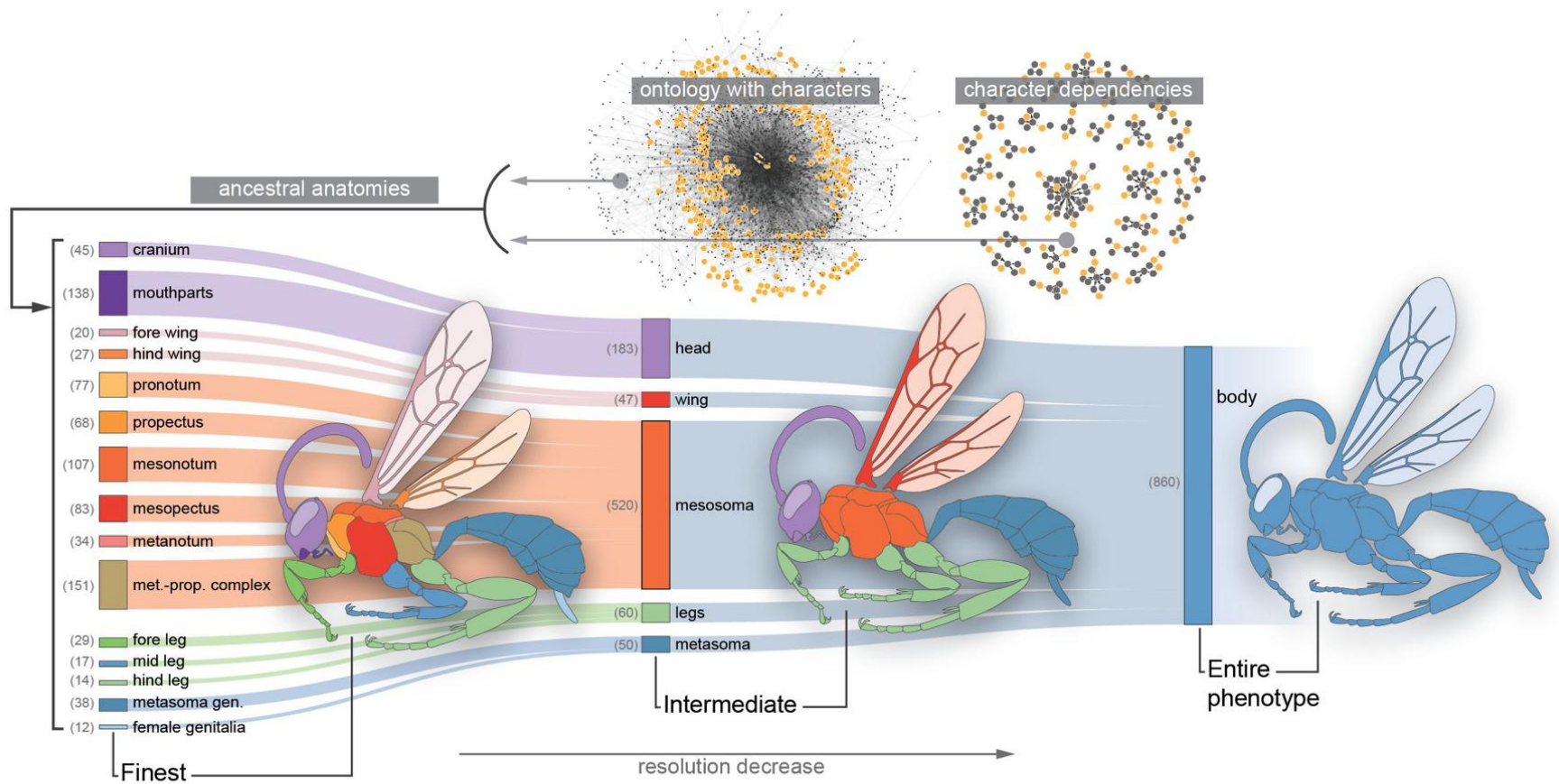
Unstructured models blow up w/increasing # of states

Structured models use knowledge of evolutionary state space to limit complexity.

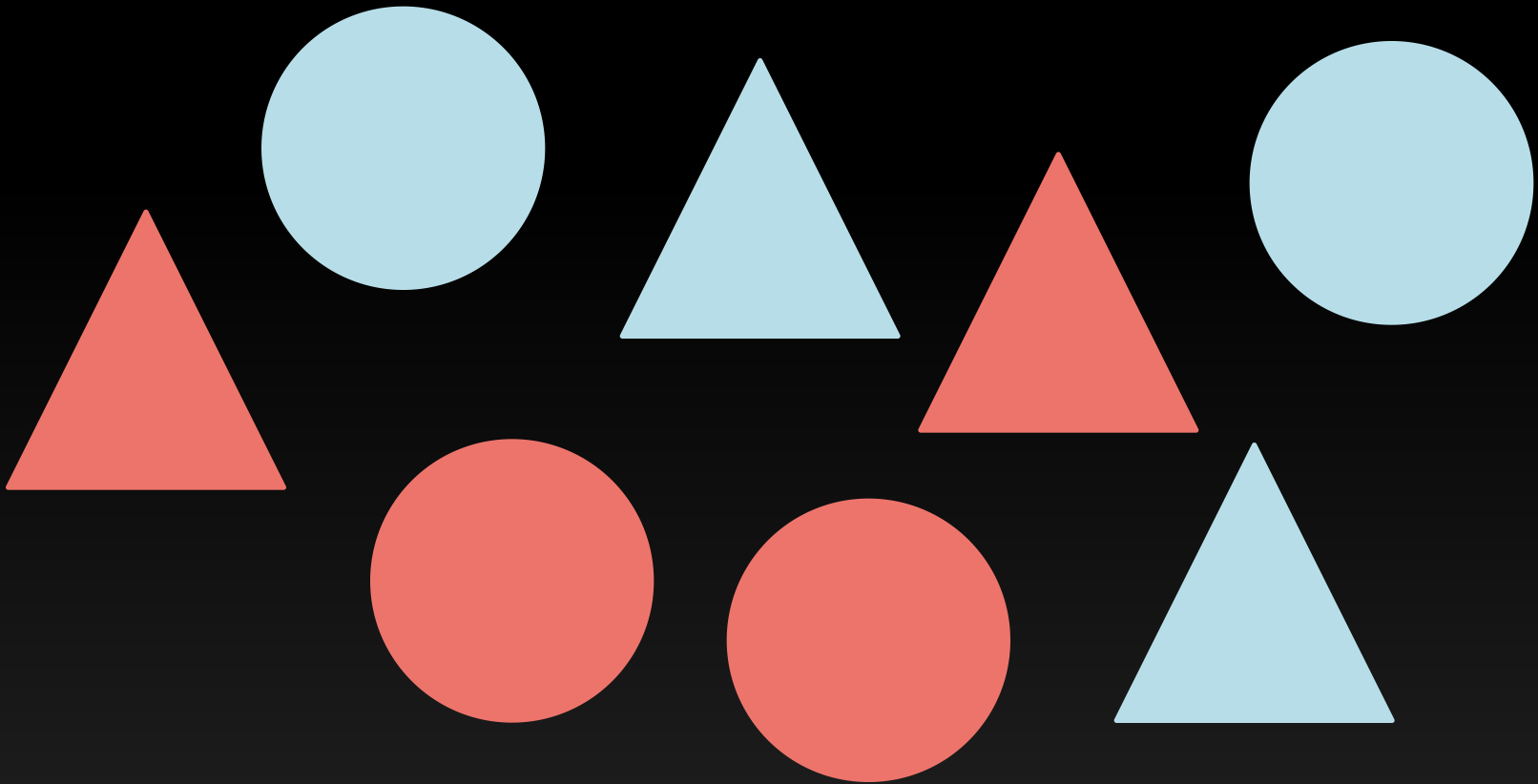


Phenotypic ontologies

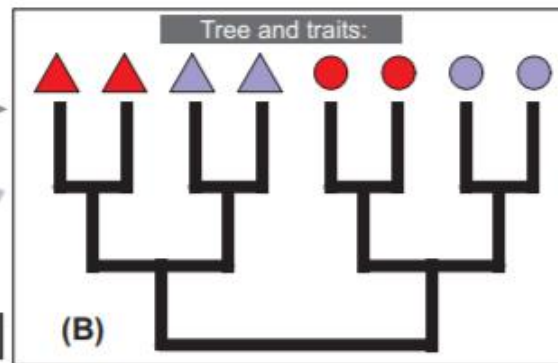
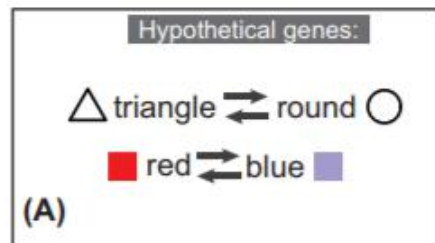




Two-Scientist Paradox (Tarasov, 2018)

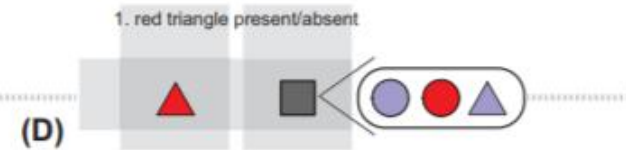
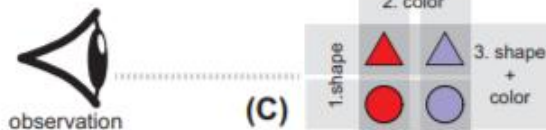


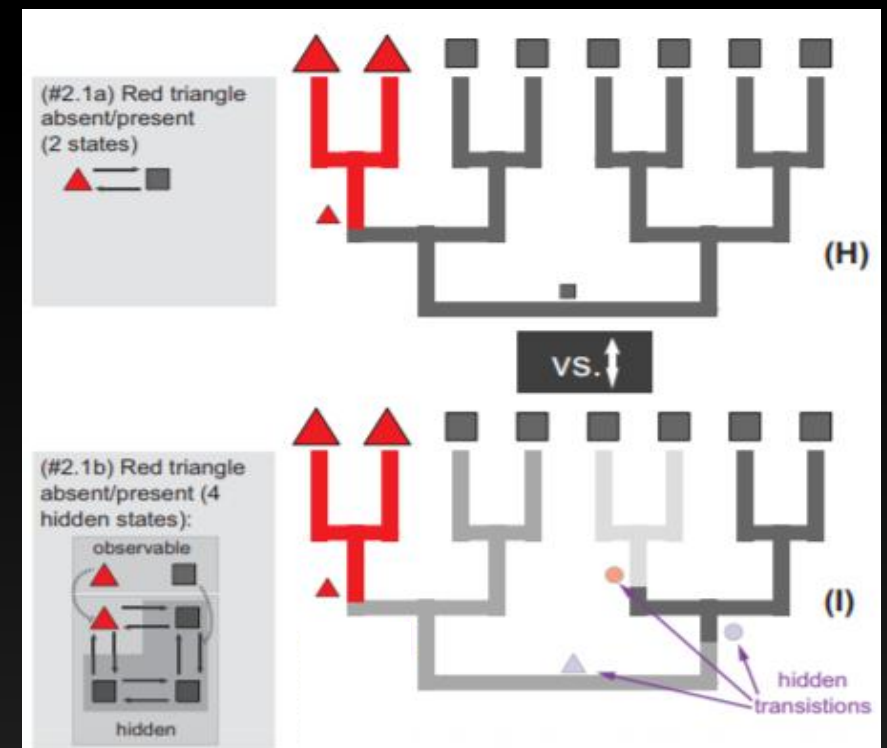
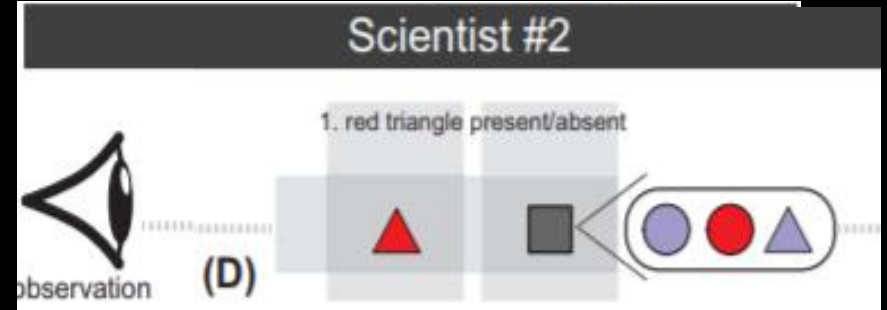
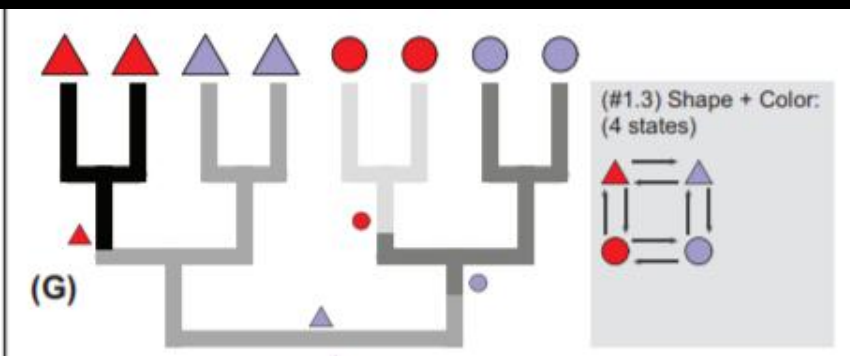
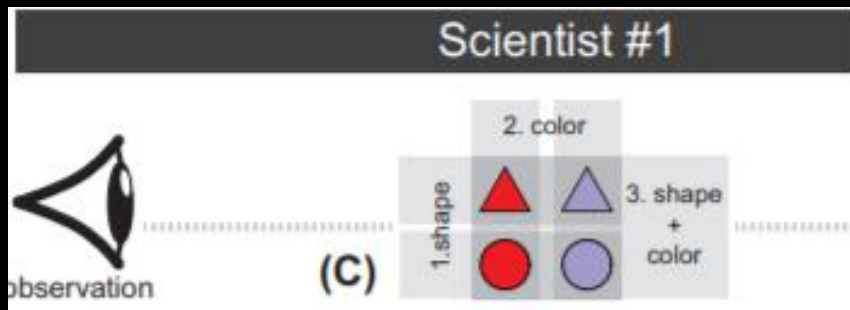
Two-Scientist Paradox (Tarasov, 2018)



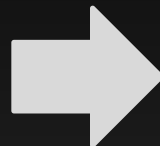
Scientist #1

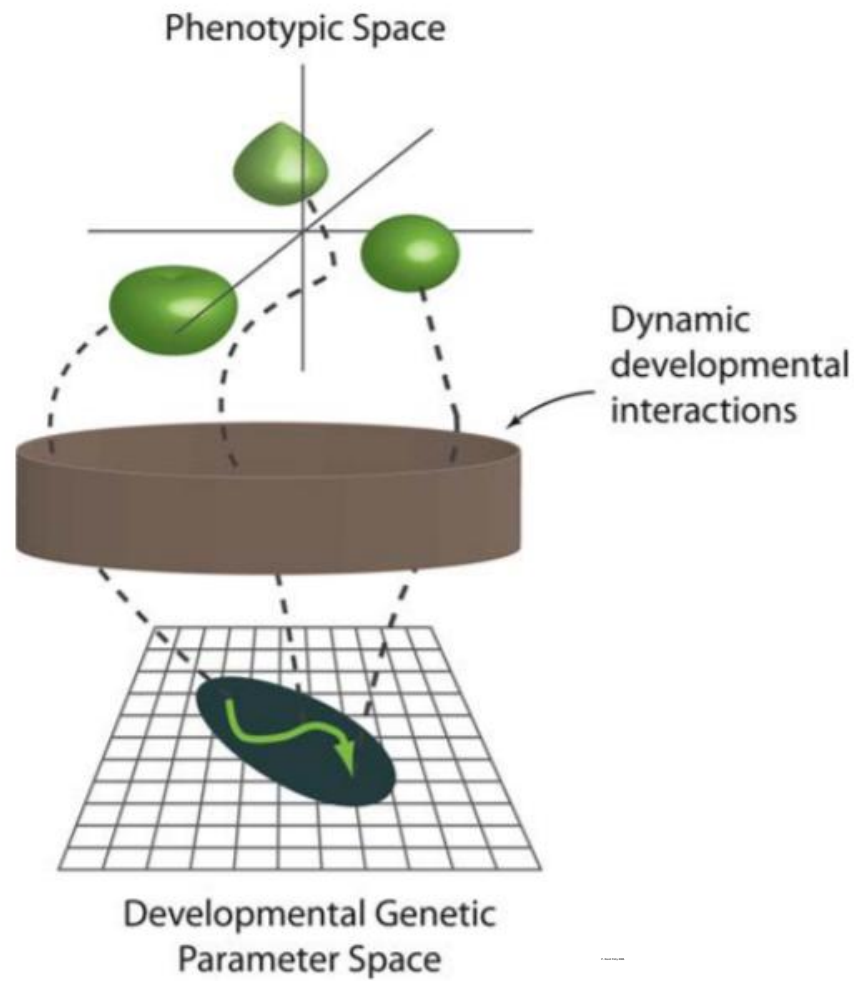
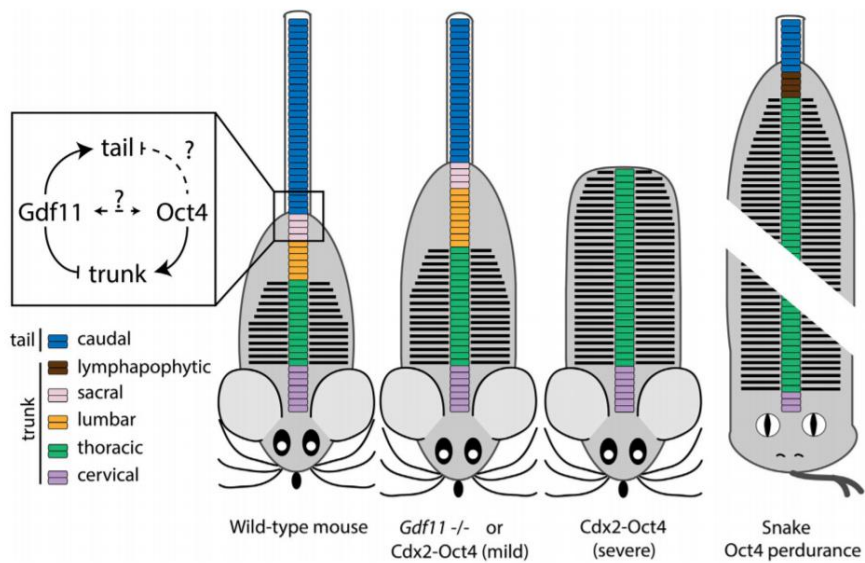
Scientist #2





Scientist #2 supports
Hidden State Model





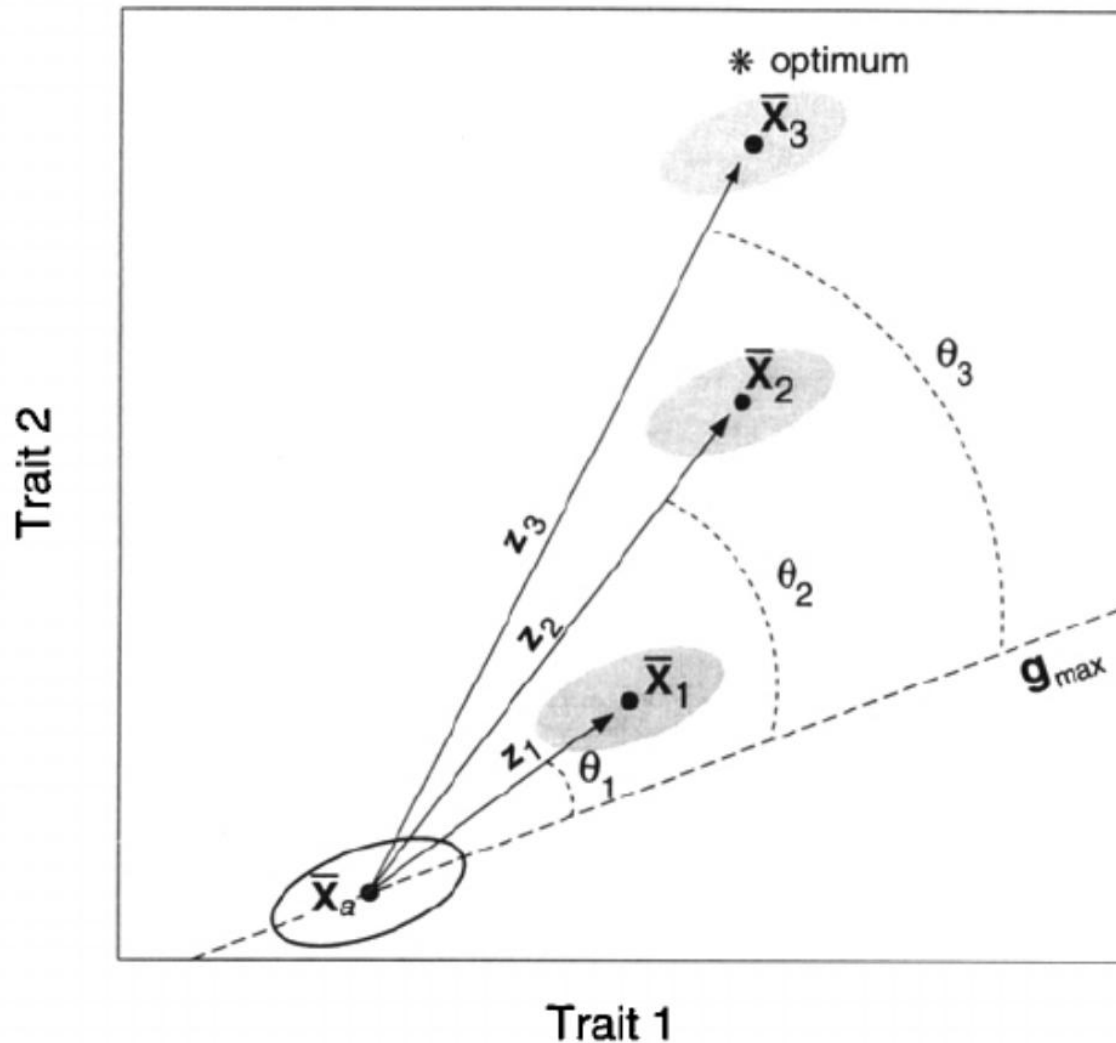
**Goal: Traits that reflect the underlying
developmental/parameter space**

ADAPTIVE RADIATION ALONG GENETIC LINES OF LEAST RESISTANCE

DOLPH SCHLUTER

*Department of Zoology and Centre for Biodiversity Research, University of British Columbia,
Vancouver, British Columbia V6T 1Z4, Canada*

E-mail: schluter@zoology.ubc.ca

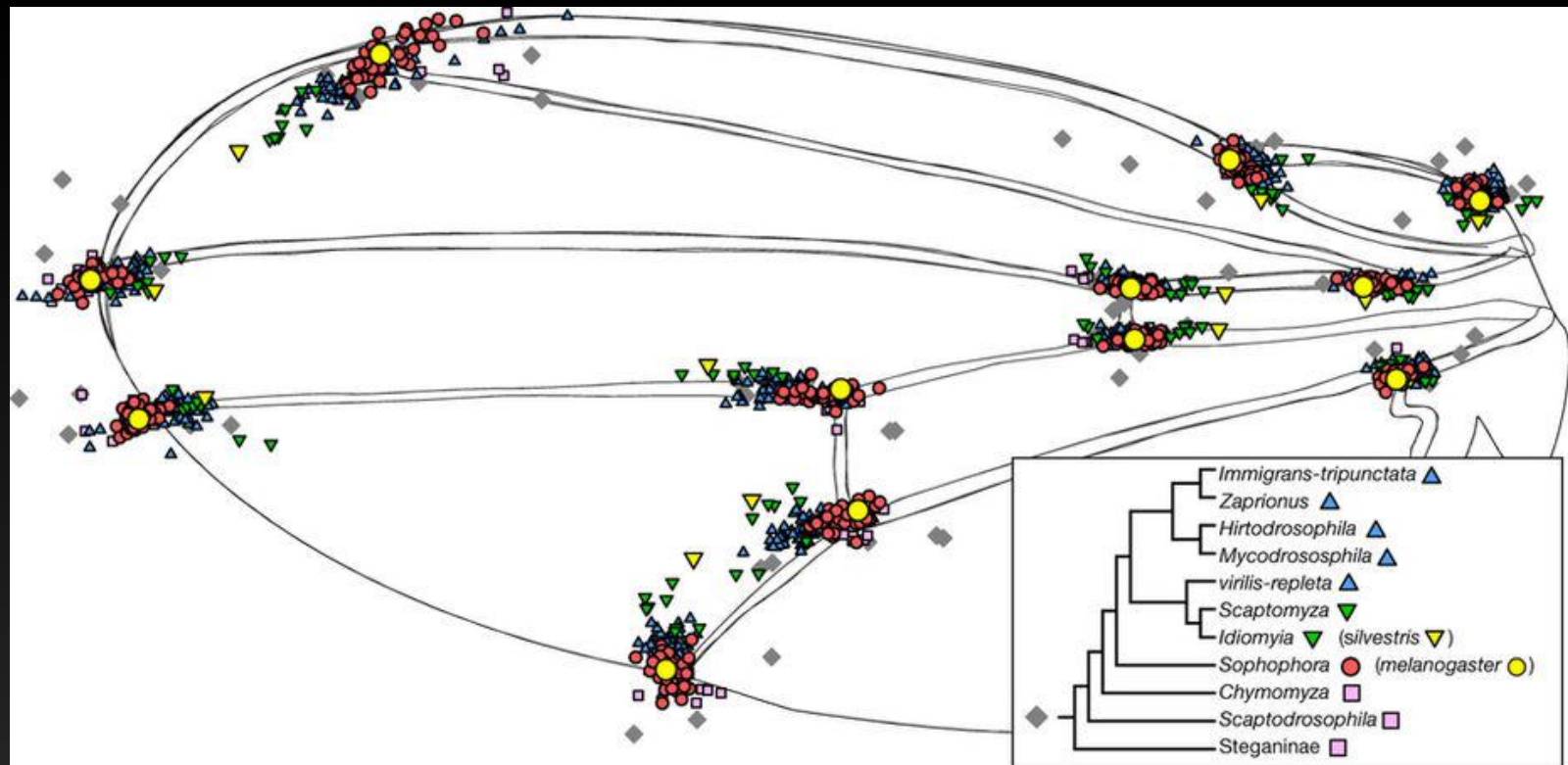


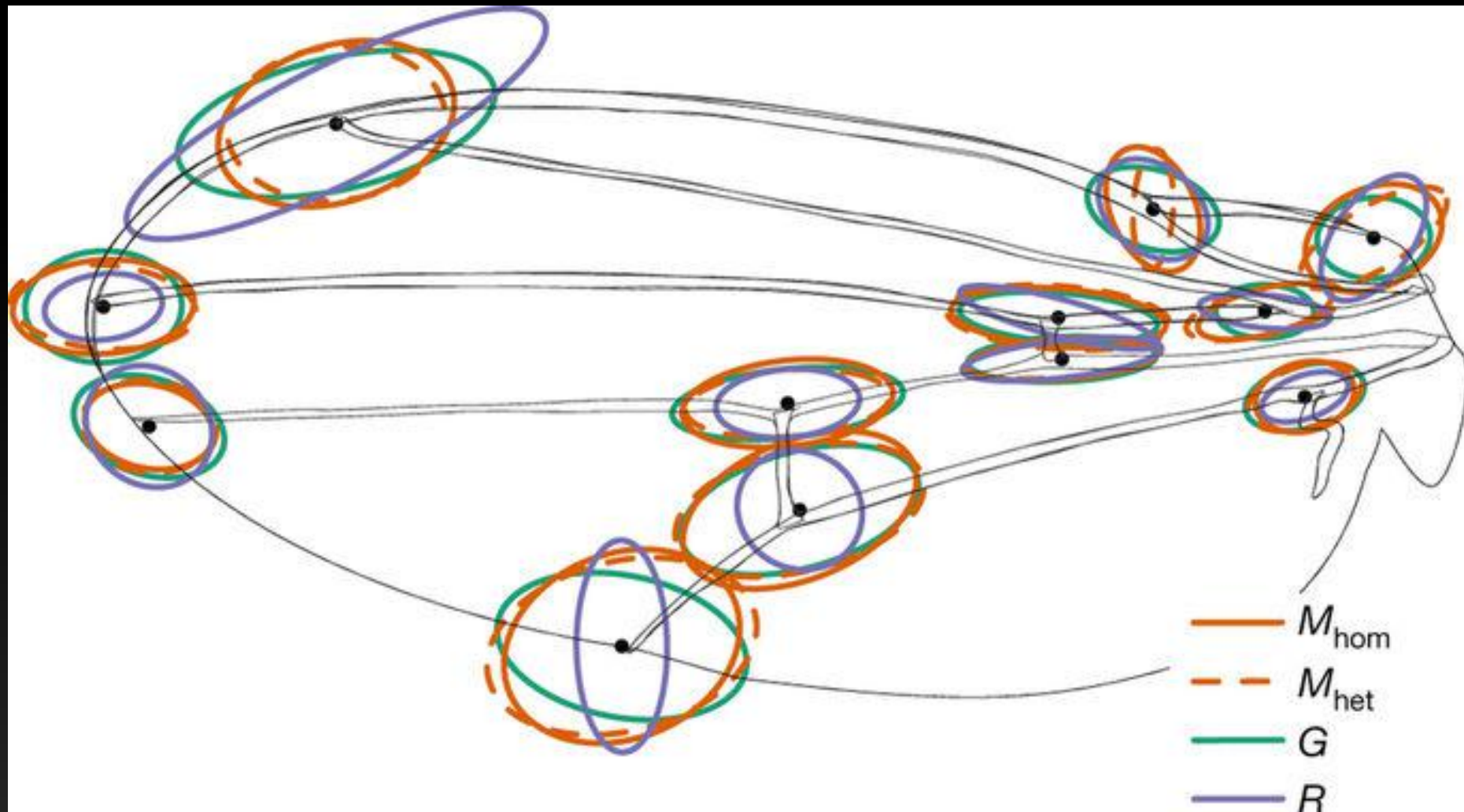
$R \sim G?$

**Does the direction of
evolutionary
divergence align with
the genetic variation
within species?**

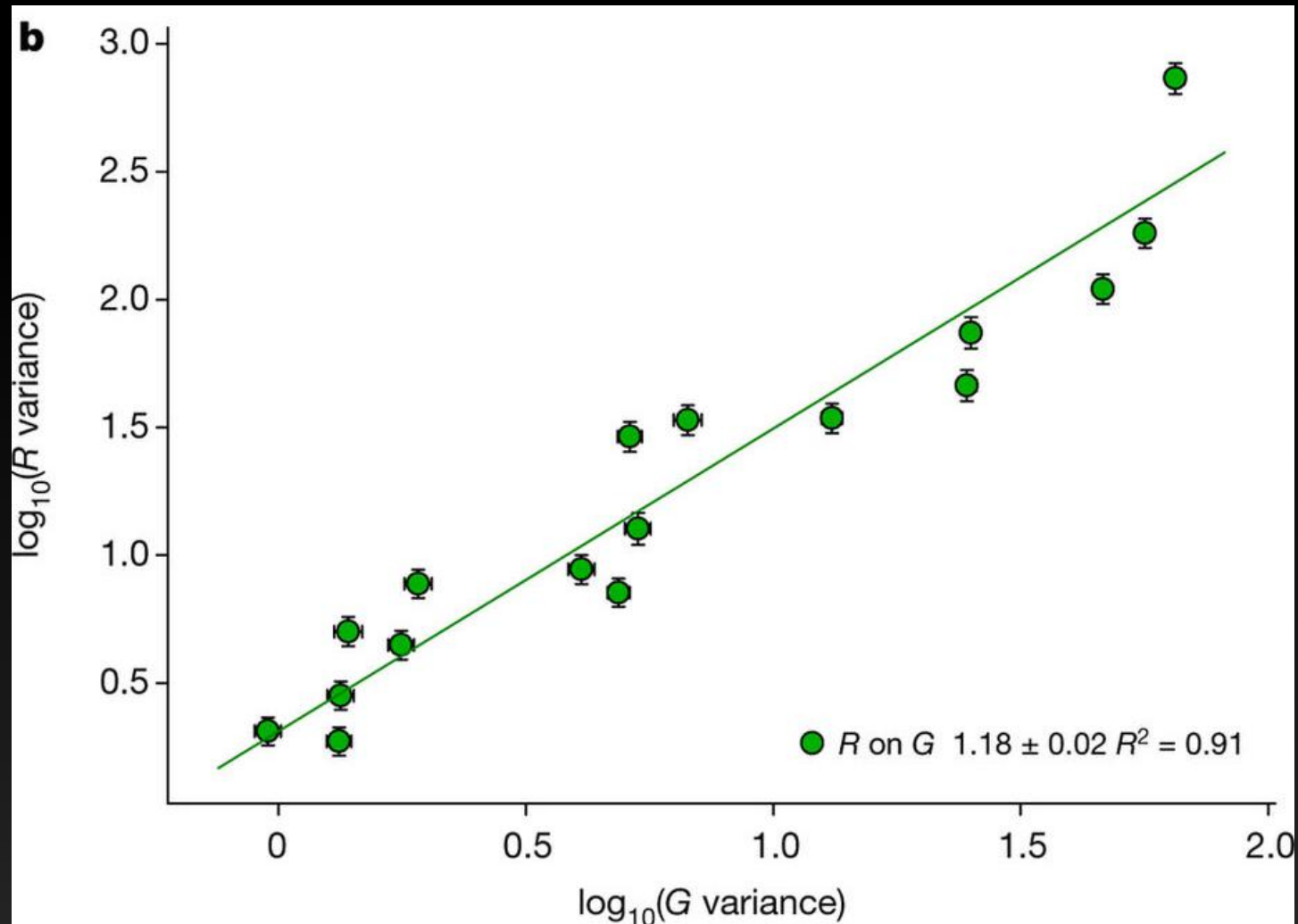
Mutation predicts 40 million years of fly wing evolution

David Houle¹, Geir H. Bolstad^{1,2}, Kim van der Linde^{1,3} & Thomas F. Hansen⁴





Strong relationship between genetic variation in shape dimensions within species (x-axis) and divergence between species (y-axis)

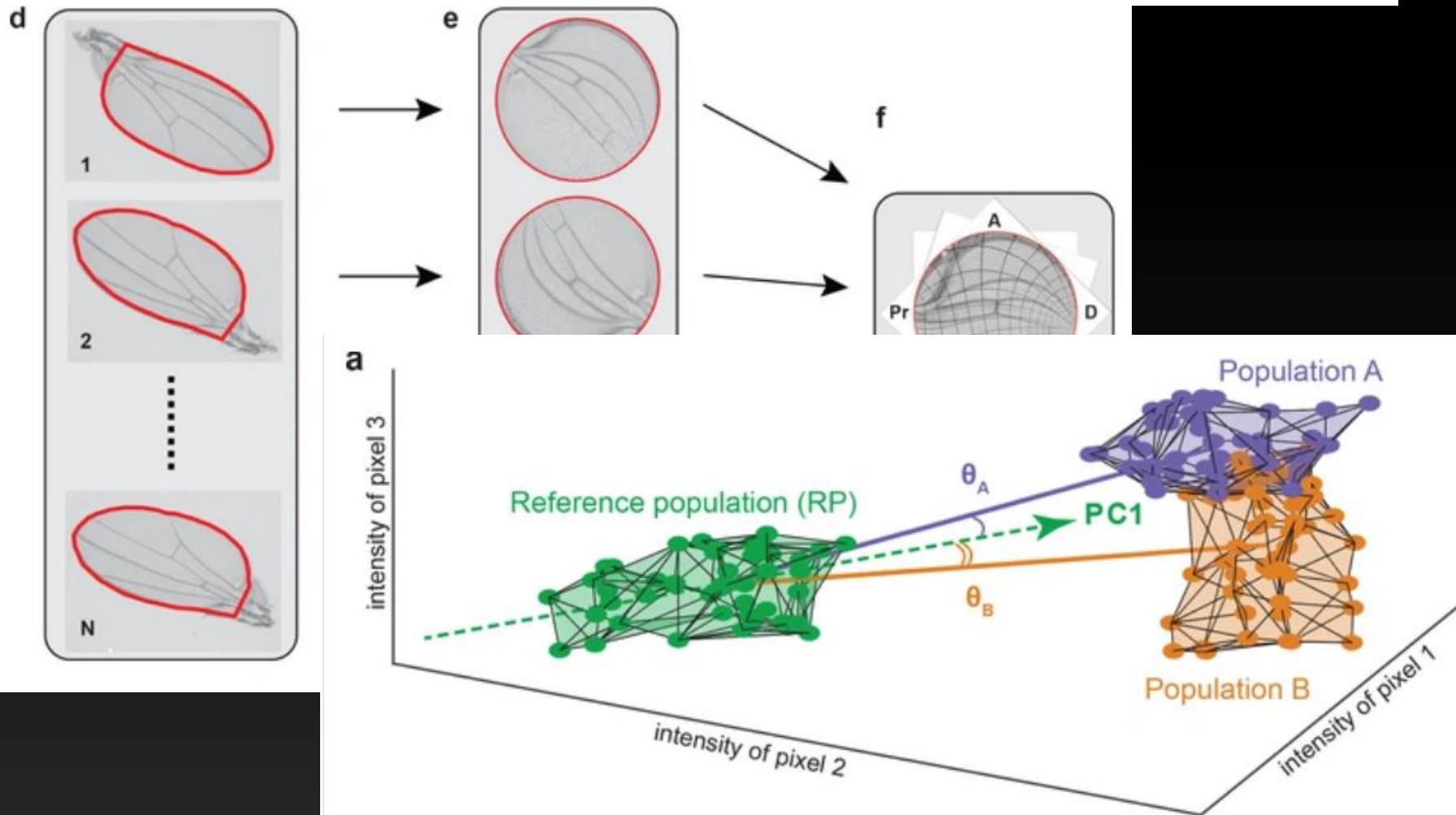


Global constraints within the developmental program of the *Drosophila* wing



Vasyl Alba, James E Carthew, Richard W Carthew, Madhav Mani

Department of Engineering Sciences and Applied Mathematics, Northwestern University, United States; NSF-Simons Center for Quantitative Biology, Northwestern University, United States; Department of Molecular Biosciences, Northwestern University, United States



Allometry

ontogenetic

growth in individuals

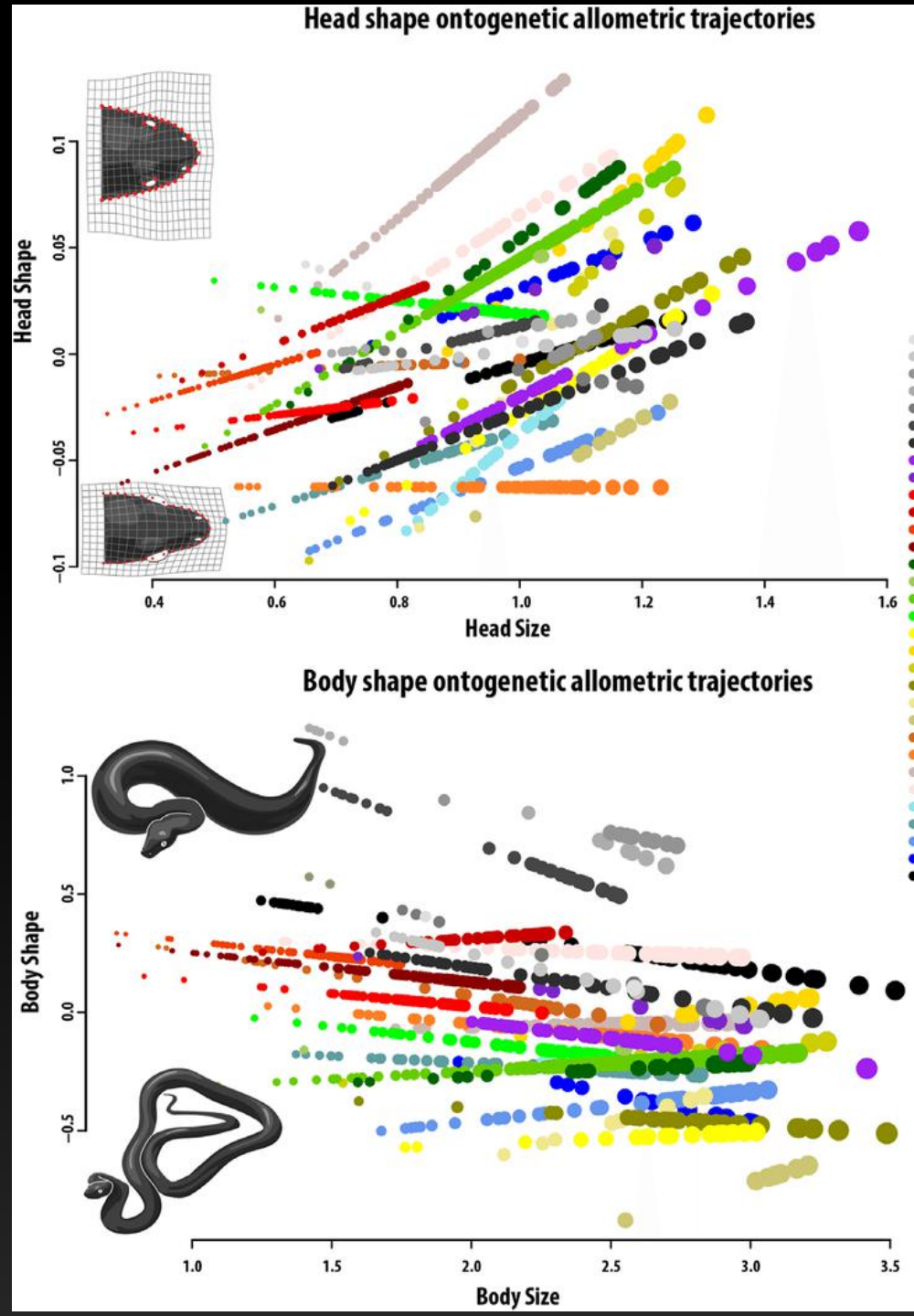
static

population variation w/size

evolutionary

species divergence w/size

PC1 usually covaries with size...even
after size standardization



What knowledge/tools/ideas/questions from trait biology can help develop ML?

Traits that align with phylogeny are likely to be meaningful, not noise

Transformation series of traits -- “the evolutionary architecture” -- often reflect the underlying genetic-developmental processes

Convergent (or converging) states among distantly related species can provide evidence for the effect of environment.

What else?