"Advanced" Topics in Morphological Evolution
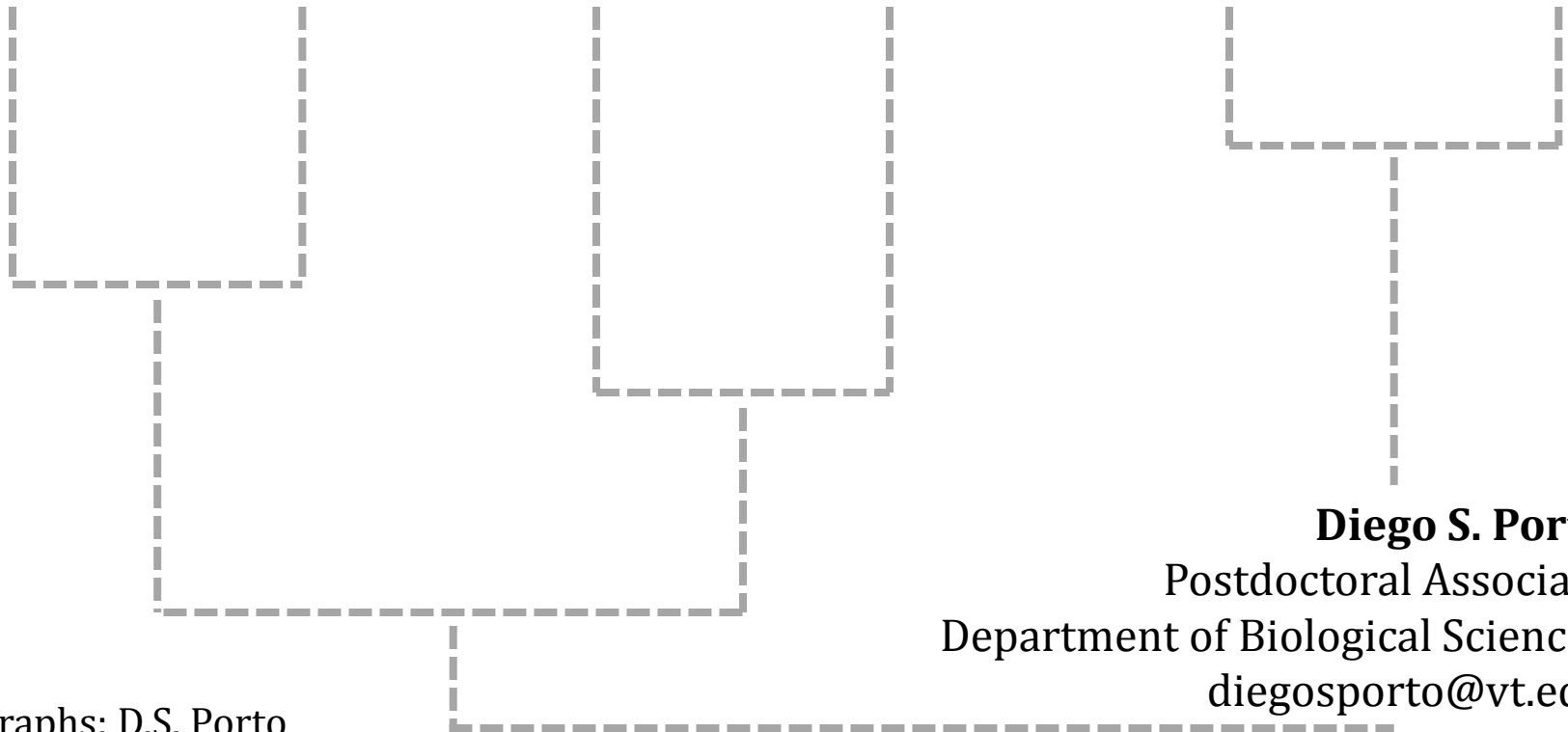
Diego S. Porto
Postdoctoral Associate
Department of Biological Sciences
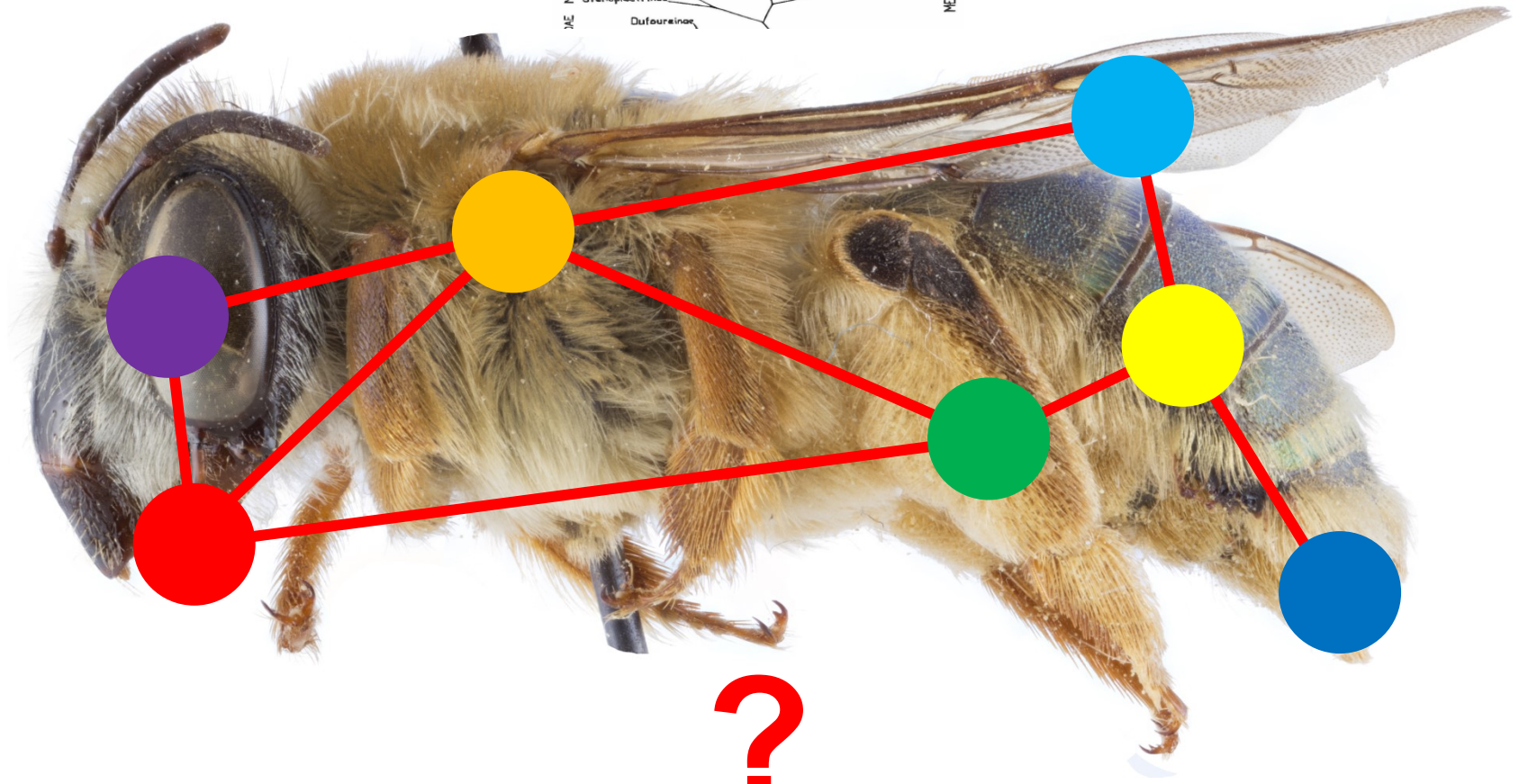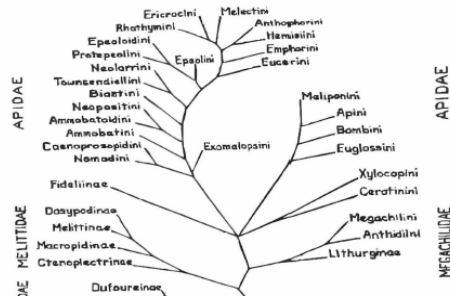diegosporto@vt.edu

Photographs: D.S. Porto

What my interests are?

Photograph: E.A.B. Almeida

Oxaea flavescens Klug, 1807

What my interests are?

Photograph: E.A.B. Almeida

Oxaea flavescens Klug, 1807

Overview

1. Morphological data in Bayesian analyses.

2. Extensions and additions to Mk model.

3. Exploiting the Bayesian "machinery" to answer questions about morphological evolution.

Morphological data in Bayesian analyses
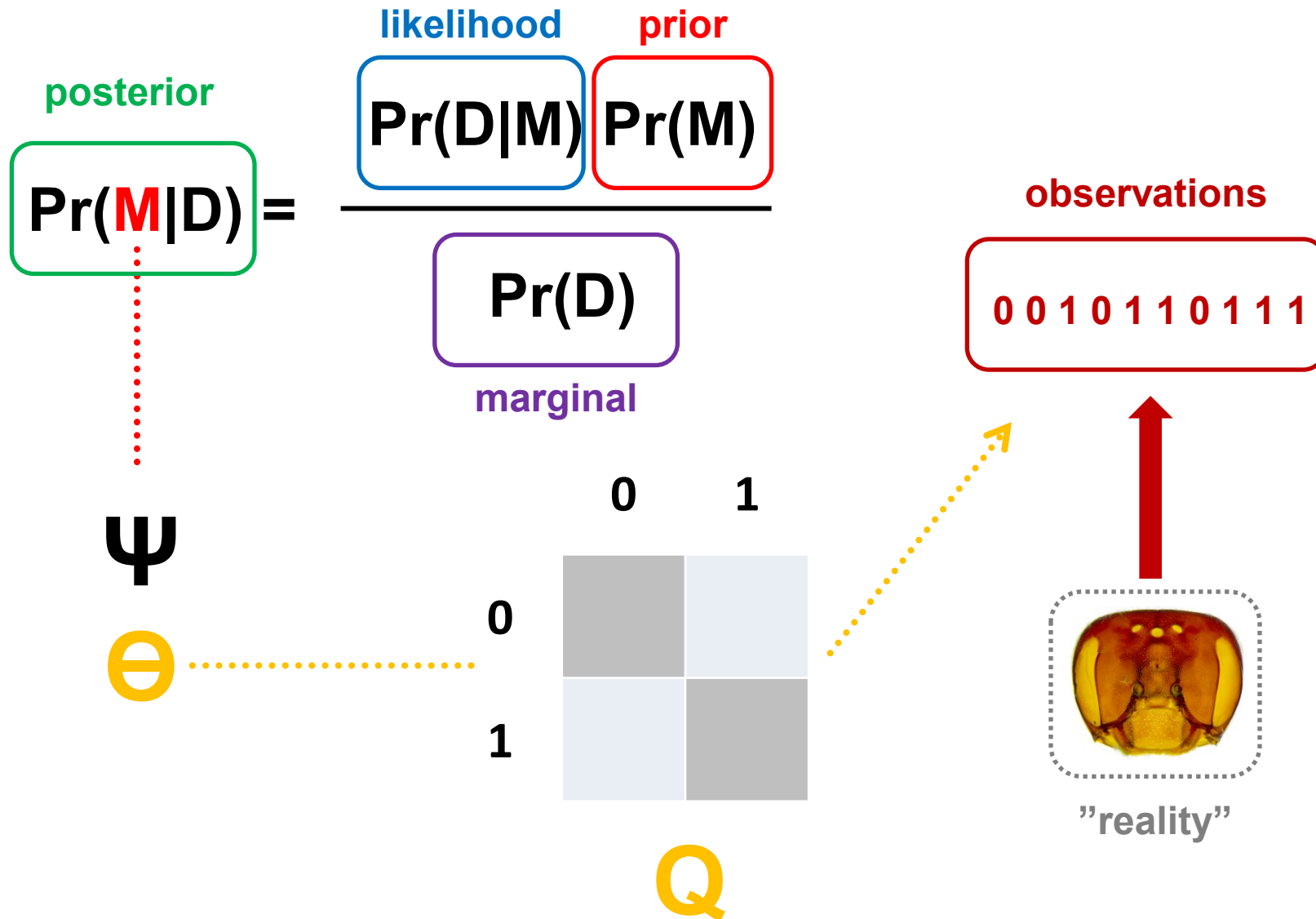
Particulars of morphological data

DNA:

1. Only 4 bases

2. Meaning across sites

3. Homology: alignment

4. Discrete

Morphology:

1. N states

2. No meaning across chars

3. Homology: expertise

4.1. Discrete:

- binary/multistate

- absence/presence

- qualities

4.2. Continuous:

- raw/ratios

The "pieces" of a Bayesian analysis

posterior

likelihood   **prior**

$$Pr(M|D) = \frac{Pr(D|M)\,Pr(M)}{Pr(D)}$$

marginal

observations

0 0 1 0 1 1 0 1 1 1

Ψ

Θ

	0	1
0		
1		

Q

"reality"

Photograph: D.S. Porto

Bayesian Inference X Parsimony

Bayesian Inference:

1. Expected # substitutions

2. Shared branch lengths

3. Invariable sites

4. Posterior sample of trees

5. Posterior probability

Parsimony:

1. Minimum # changes

2. Individual branch lengths

3. Only variable characters

4. Most parsimonious tree

5. Bootstrap/Bremer

Mk model: instantaneous ratematrix (Q)

JC69

	A	C	G	T
A	-3β	β	β	β
C	β	-3β	β	β
G	β	β	-3β	β
T	β	β	β	-3β

Mk (Lewis 2001)

α

	0	1	...	k
0	1 - k	1	...	1
1	1	1 - k	...	1
⋮	⋮	⋮	1 - k	⋮
k	1	1	...	1 - k

Mk model: assumptions

Assumptions of JC69:

1. All substitutions equally likely

2. Base frequencies equal

3. Every site has equal probability of substitution

4. Process is constant through time

5. Sites are independent of each other

6. Substitution is Markovian (memoryless)

7. All sites have the same evolutionary history

Mk model: assumptions

Assumptions of Mk:

1. All changes equally likely

2. State frequencies equal

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

8. Only informative chars are included (usually).

Extensions to Mk model

Mk model: some tweaks we can make
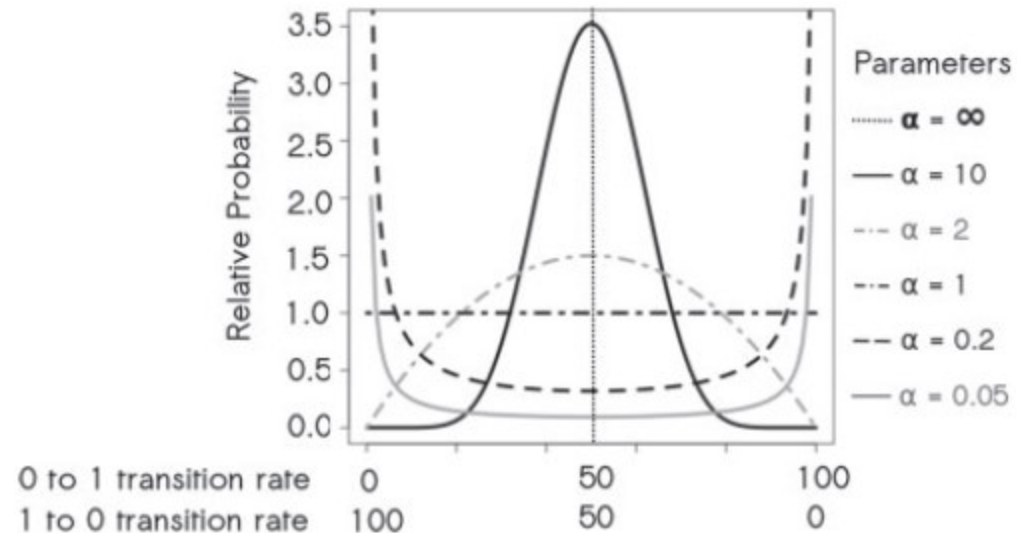
Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

Mk model: some tweaks we can make

Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

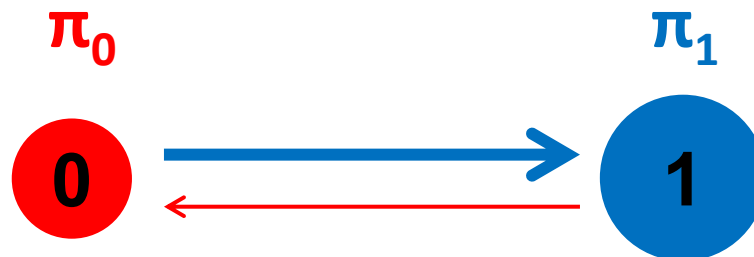# Tweak I: unequal char-state frequencies

ratematrix (Q)

	0	1
0	$-\mu\pi_0$	$\mu\pi_1$
1	$\mu\pi_0$	$-\mu\pi_1$

Beta distribution

Wright et al. (2016): Syst. Biol. 65: 602-611

Mk model: some tweaks we can make
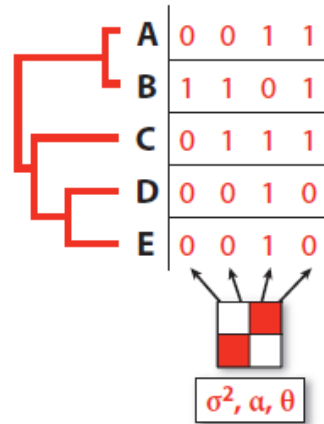
Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

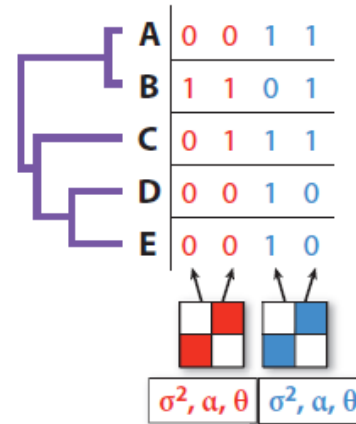# Tweak II: among-character rate var. (ACRV)

ACRV:

1. Shared or individual

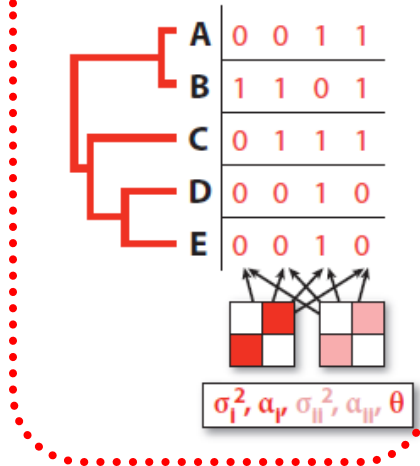2. More categories
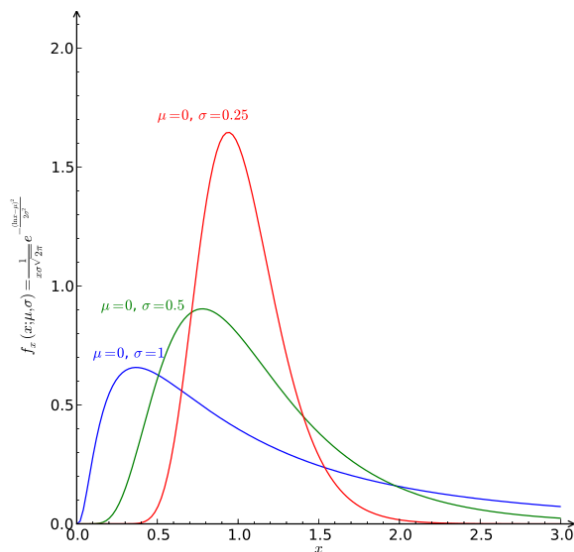
3. Lognormal distribution

a No heterogeneity

A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

σ^2, α, θ

b Partitioning by character

A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

σ^2, α, θ σ^2, α, θ

c Discrete gamma

A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

$\sigma_I^2, \alpha_I, \sigma_{II}^2, \alpha_{II}, \theta$

O'Maera (2012): Annu. Rev. Ecol. Evol. Syst. 43: 267-285

Lognormal distribution

$f_x(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$

$\mu=0, \sigma=0.25$

$\mu=0, \sigma=0.5$

$\mu=0, \sigma=1$

Gamma distribution

$k = 1, \theta = 2.0$
$k = 2, \theta = 2.0$
$k = 3, \theta = 2.0$
$k = 5, \theta = 1.0$
$k = 9, \theta = 0.5$
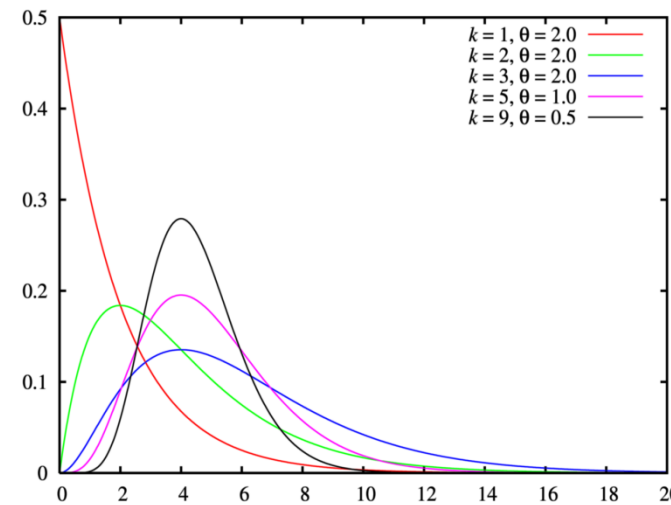
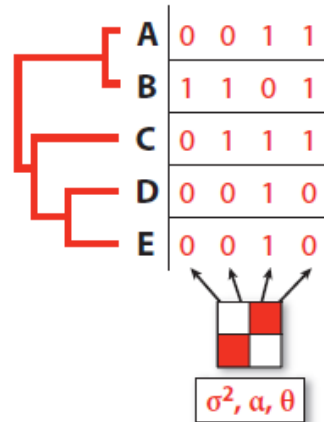# Tweak II: among-partition rate var. (APRV)

APRV:

1. Rate multipliers

2. Unlinked branch lengths

a No heterogeneity

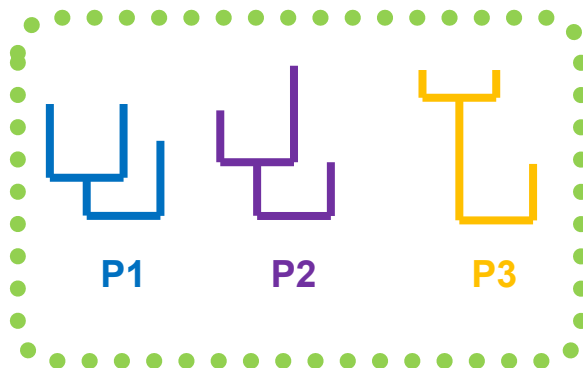|				
A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

σ^2, α, θ

b Partitioning by character

A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

σ^2, α, θ σ^2, α, θ

c Discrete gamma

A	0	0	1	1
B	1	1	0	1
C	0	1	1	1
D	0	0	1	0
E	0	0	1	0

$\sigma_I^2, \alpha_I, \sigma_{II}^2, \alpha_{II}, \theta$
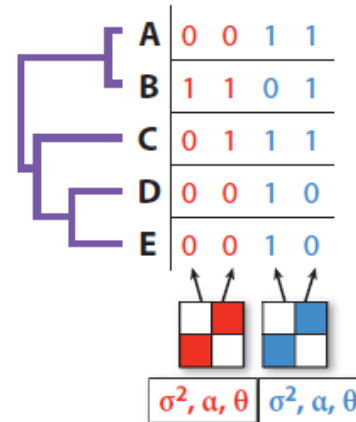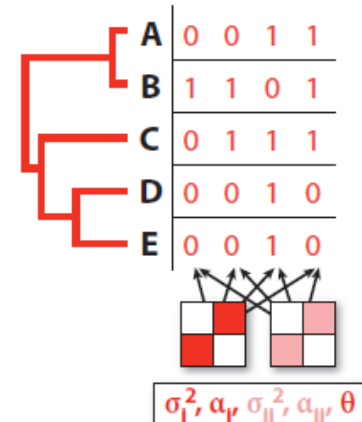
O'Maera (2012): Annu. Rev. Ecol. Evol. Syst. 43: 267-285

unlinked brlens

P1   P2   P3

branch length proportions

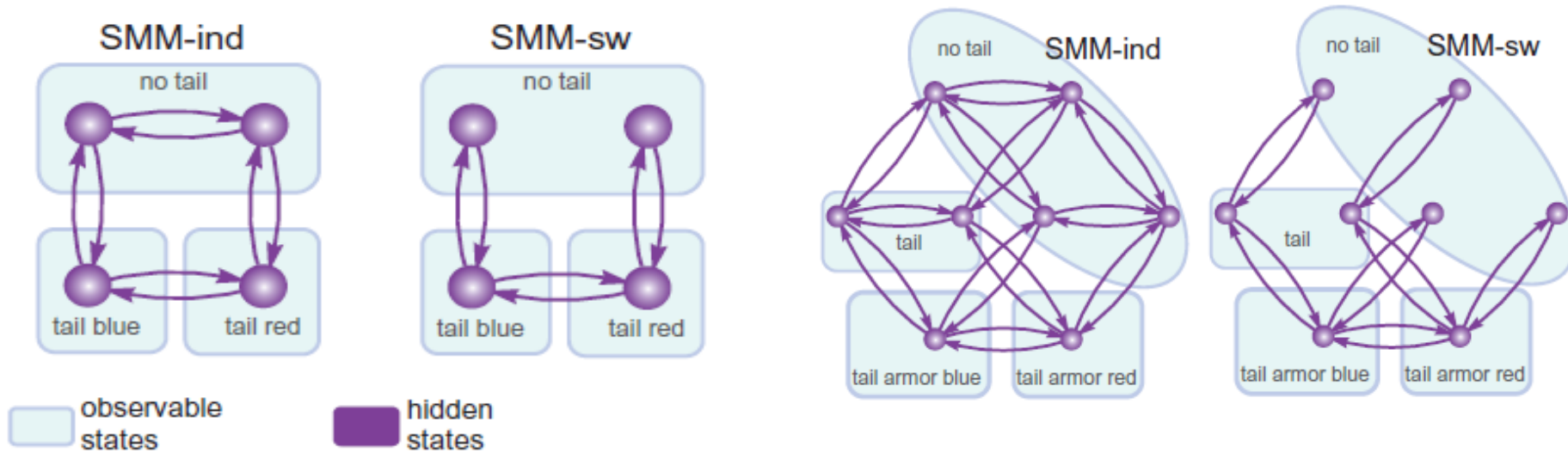**rate multipliers**

P1   P2   P3
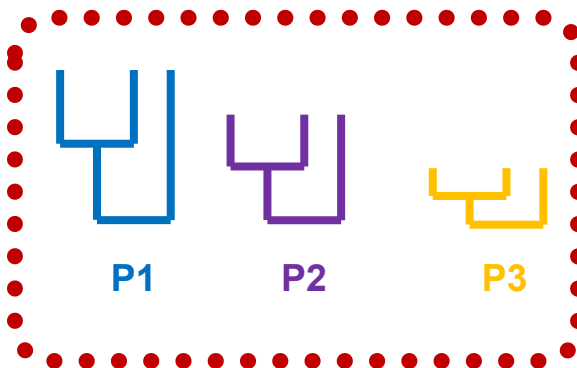
scale all branches

Mk model: some tweaks we can make

Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

Tweak III: other Markovian Models

Hidden Markov Model (HMM) + Structured Markov Model (SMM)

2 characters

TC	00	01	10	11
00	-	1	1	0
01	1	-	0	1
10	1	0	-	1
11	0	1	1	-

TC	00	01	10	11
00	-	0	1	0
01	0	-	0	1
10	1	0	-	1
11	0	1	1	-

3 characters

SMM:
1. Correlations
2. Hierarchy

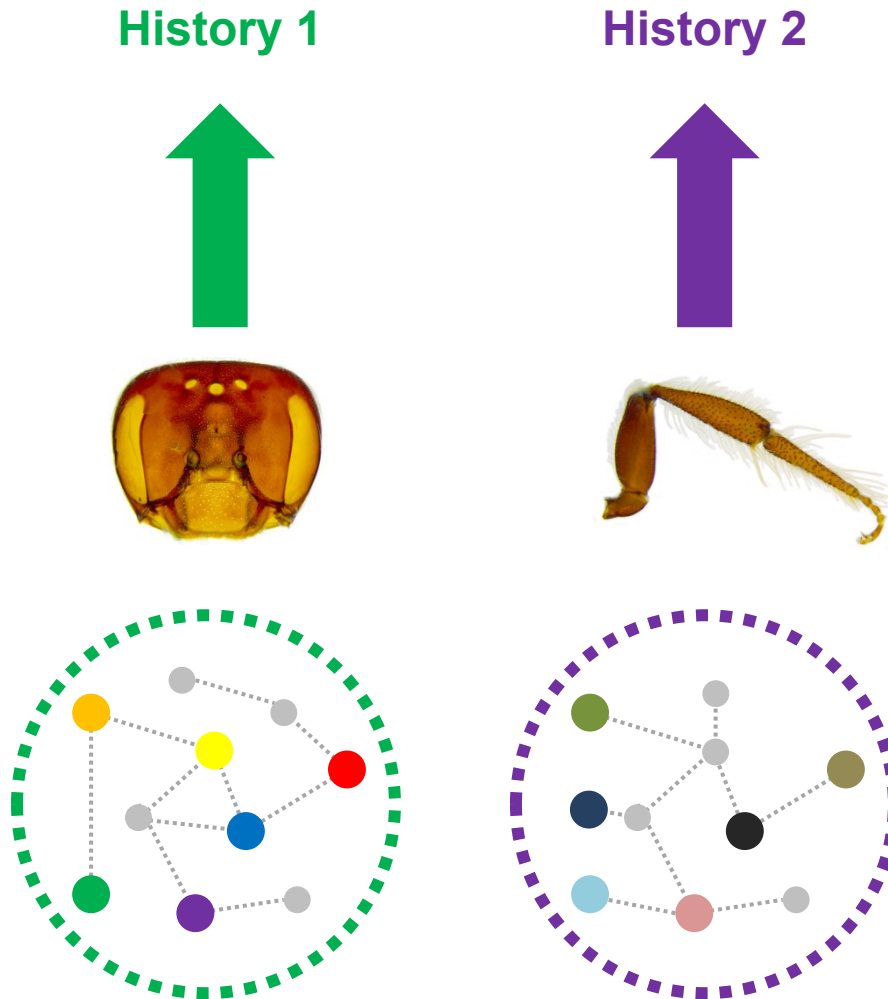HMM:
1. Visible layer
2. Hidden layer

Mk model: some tweaks we can make

Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

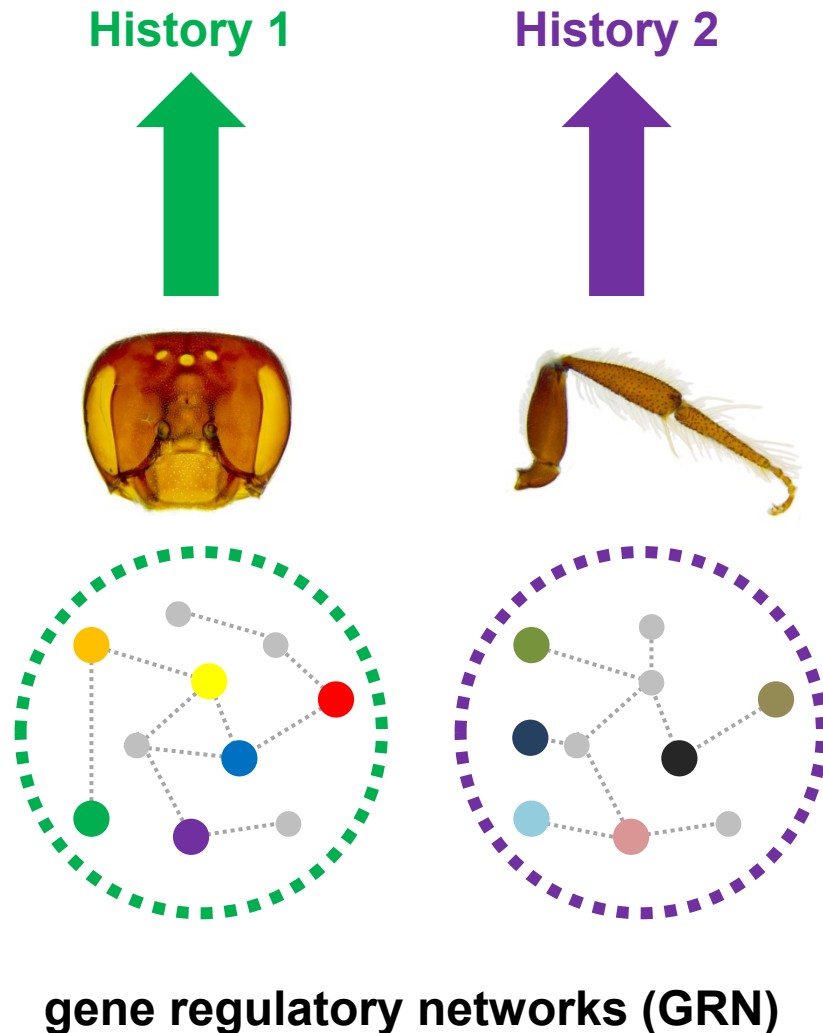7. All chars have the same evolutionary history

Tweak IV: renewal process (and others)

Memory in Trait Macroevolution[*]

Emma E. Goldberg[1,†,‡] and Jasmine Foo[2,†]

1. Department of Ecology, Evolution and Behavior, University of Minnesota, Saint Paul, Minnesota 55108;   2. Department of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

Goldberg & Foo (2020): Am. Nat. 195: 300-314

1. Renewal process

2. Threshold models

3. Random walk models

Mk model: some tweaks we can make

Assumptions of Mk:

1. All changes equally likely (exchangeabilities)

2. State frequencies equal (equilibrium freqs.)

3. Every character has equal probability of change

4. Process is constant through time

5. Characters are independent of each other

6. Change is Markovian (memoryless)

7. All chars have the same evolutionary history

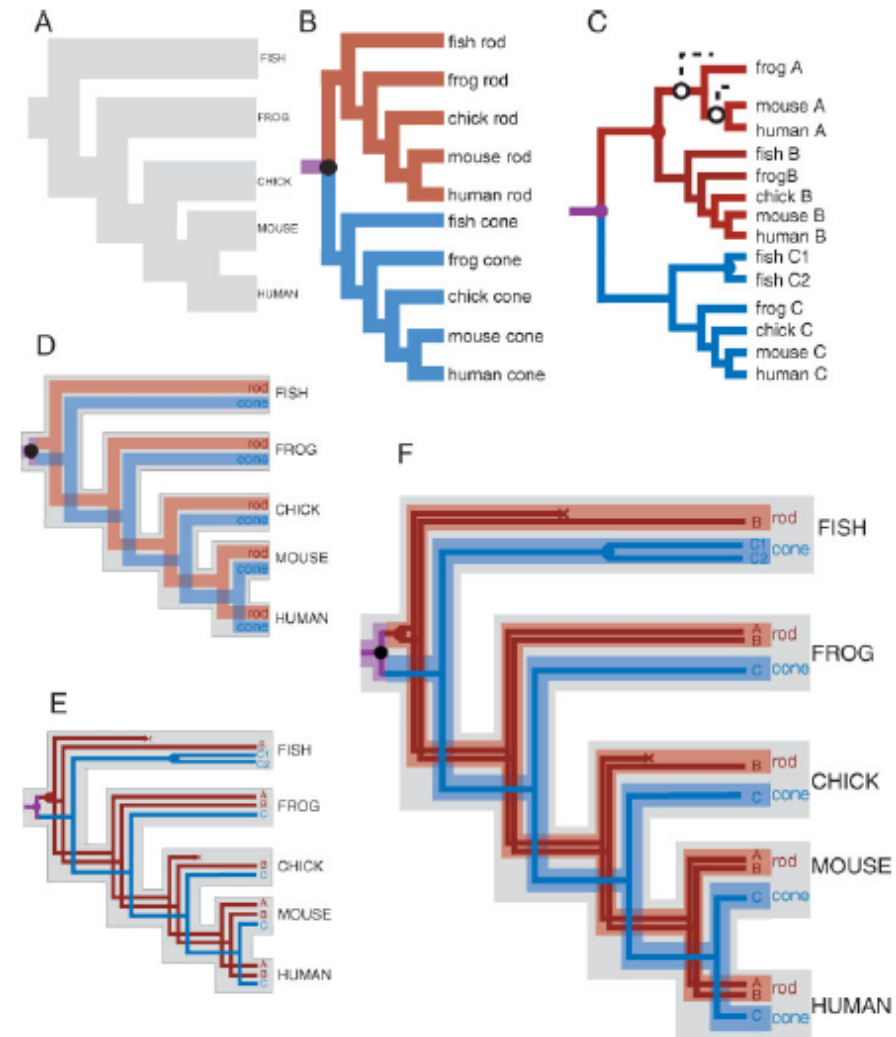# Exploiting the Bayesian "machinery"

Exploiting the Bayesian "machinery"

History 1

History 2

gene regulatory networks (GRN)

Photographs: D.S. Porto

Exploiting the Bayesian "machinery"

History 1

History 2

evolutionary modules

gene regulatory networks (GRN)

Serb & Oakley (2005): Bioessays 27:1158-1166

Exploiting the Bayesian "machinery"
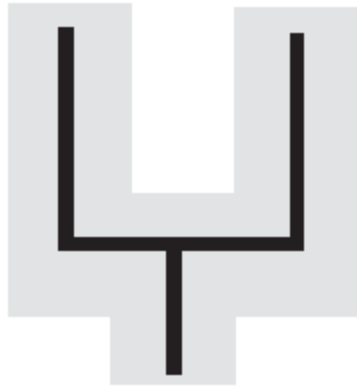
Orthologous duplication

Incomplete tracking
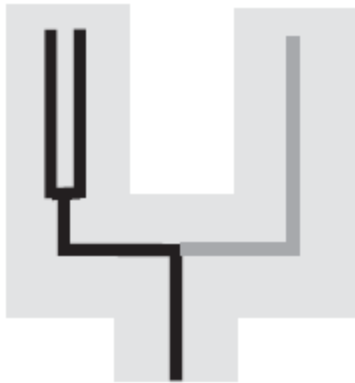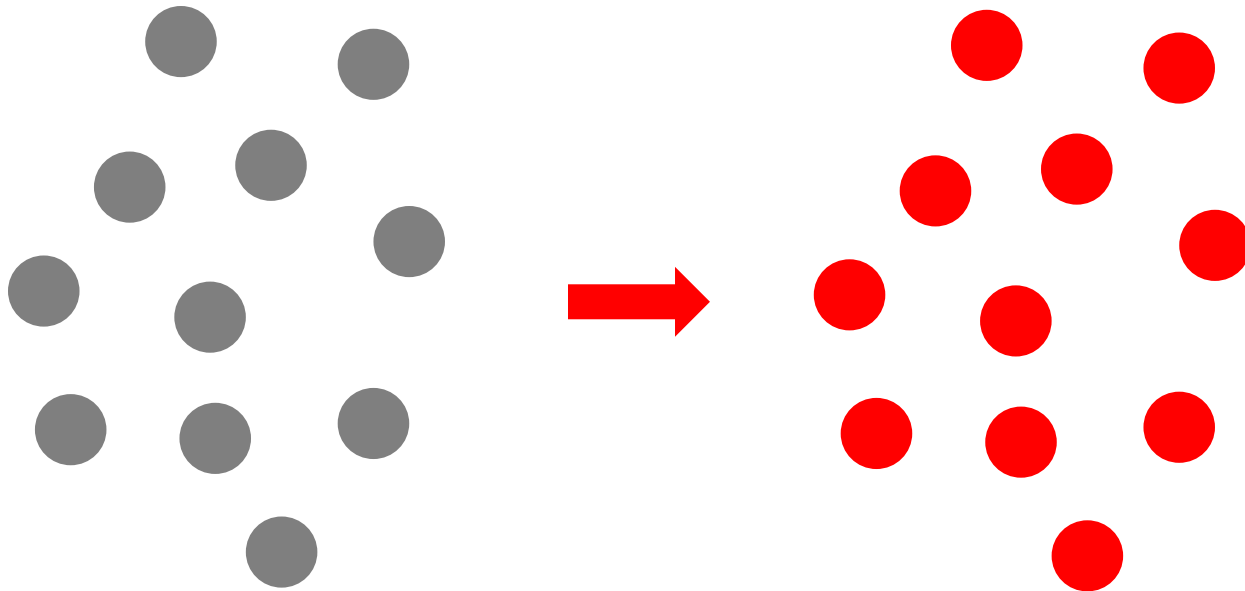
Paralogous duplication

Loss

Serb & Oakley (2005): Bioessays 27:1158-1166
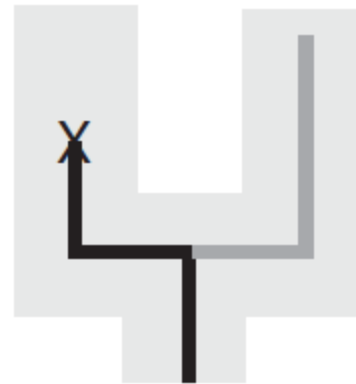
Bayesian Machinery: Information Theory

Information Theory: Entropy (Shannon 1948)
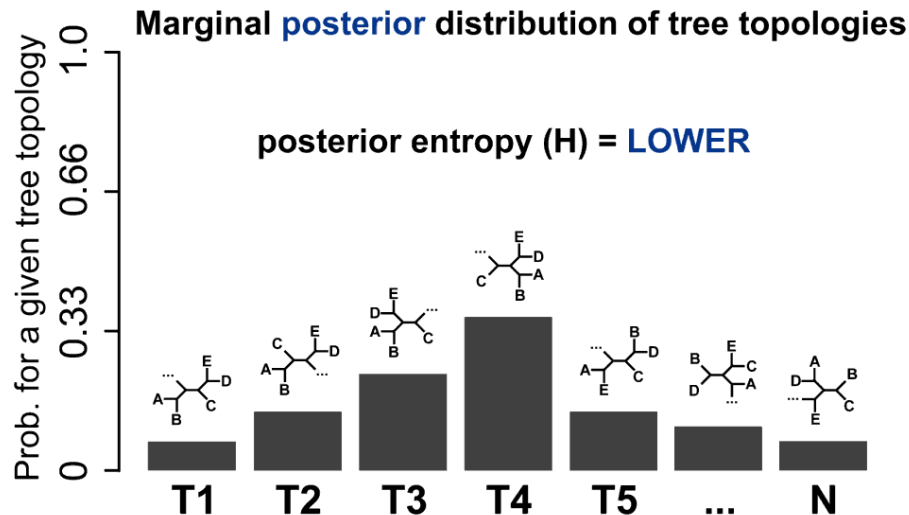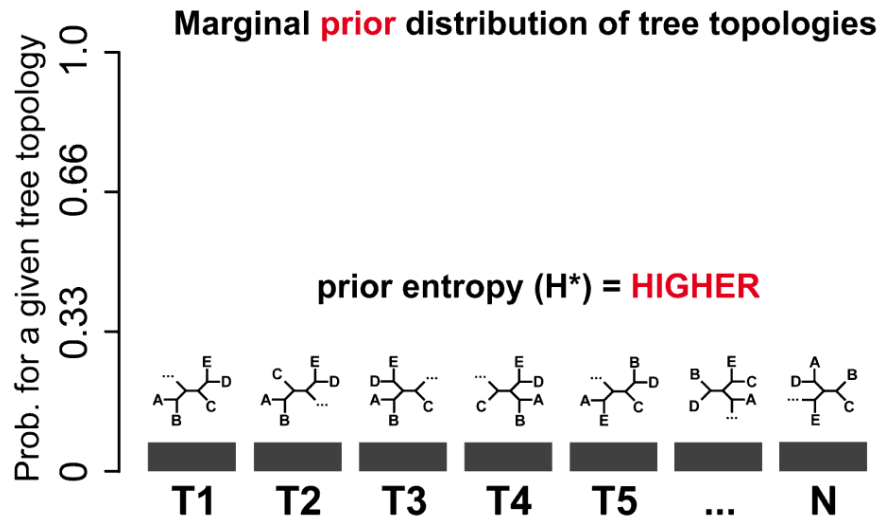

PRIOR
1. more uncertainty
2. higher entropy
3. equiprobable outcomes

POSTERIOR
1. less uncertainty
2. lower entropy
3. unequal probability

Bayesian Phylogenetic Information (BPI)

Photograph: D.S. Porto

Phylogenetic dissonance (= conflict)
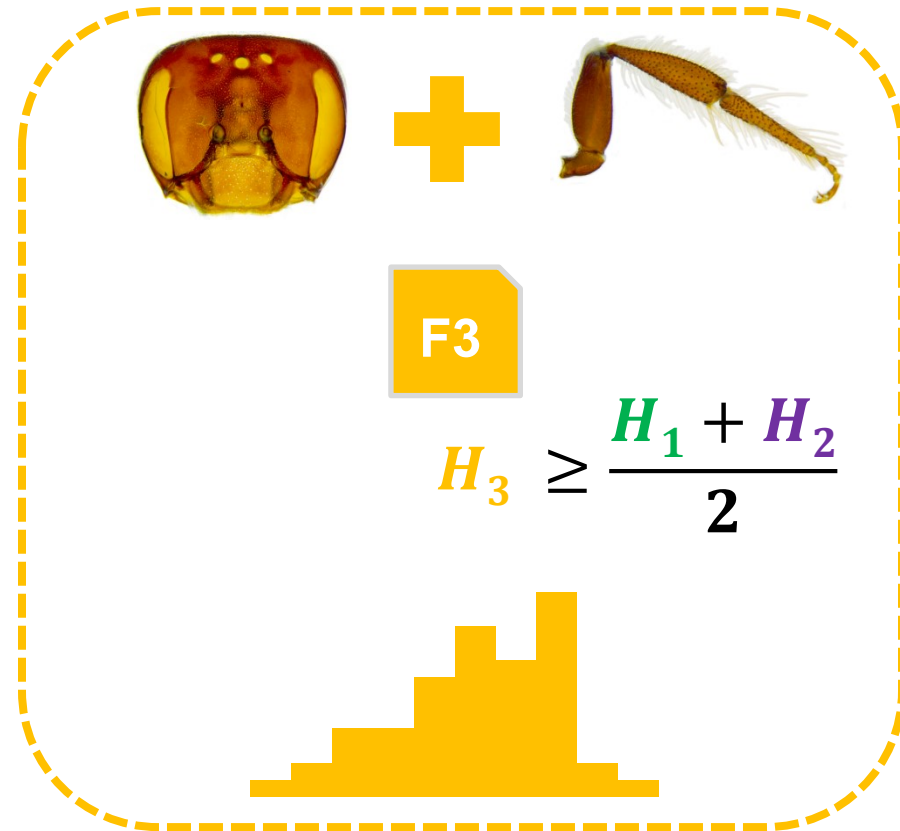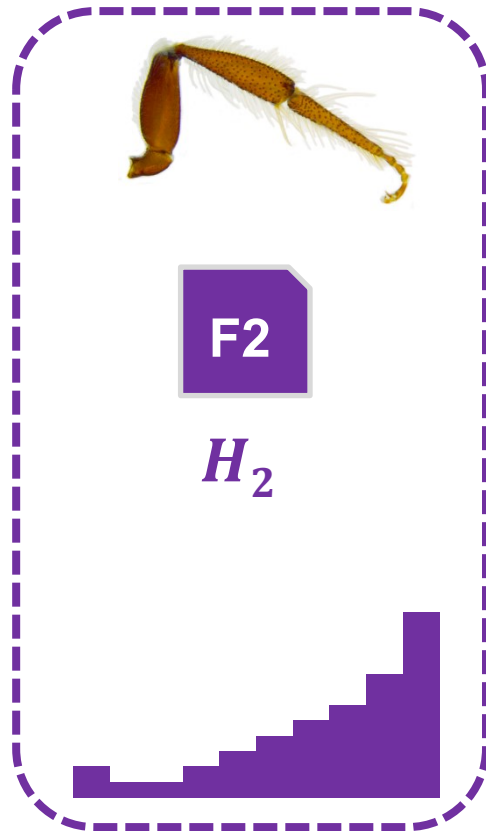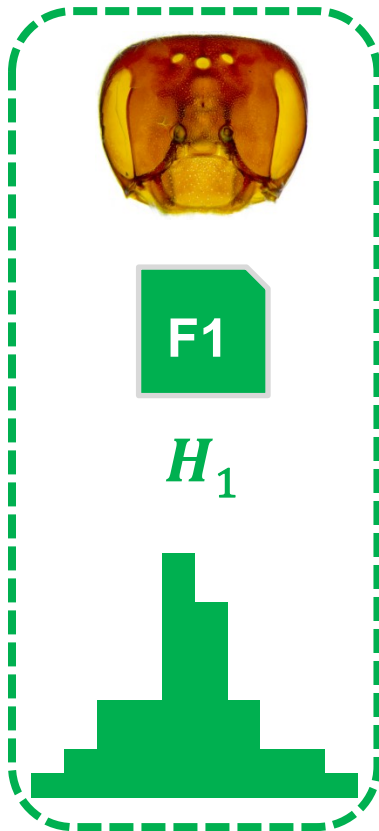
Applications: some examples

Example I: information content

Porto et al. (2020): Syst. Biol.

Example II: conflict among partitions

N = 42 chars. I = 75.86 % **HD**

N = 52 chars. I = 77.13 % **MP**

N = 57 chars. I = 77.65 % **MS**

N = 16 chars. I = 64.86 % **WG**

N = 49 chars. I = 78.04 % **LG**

N = 11 chars. I = 58.87 % **MT**

N = 55 chars. I = 71.69 % **GN**

N = 181 chars. I = 92.26 % **EXT**

N = 101 chars. I = 82.05 % **INT**
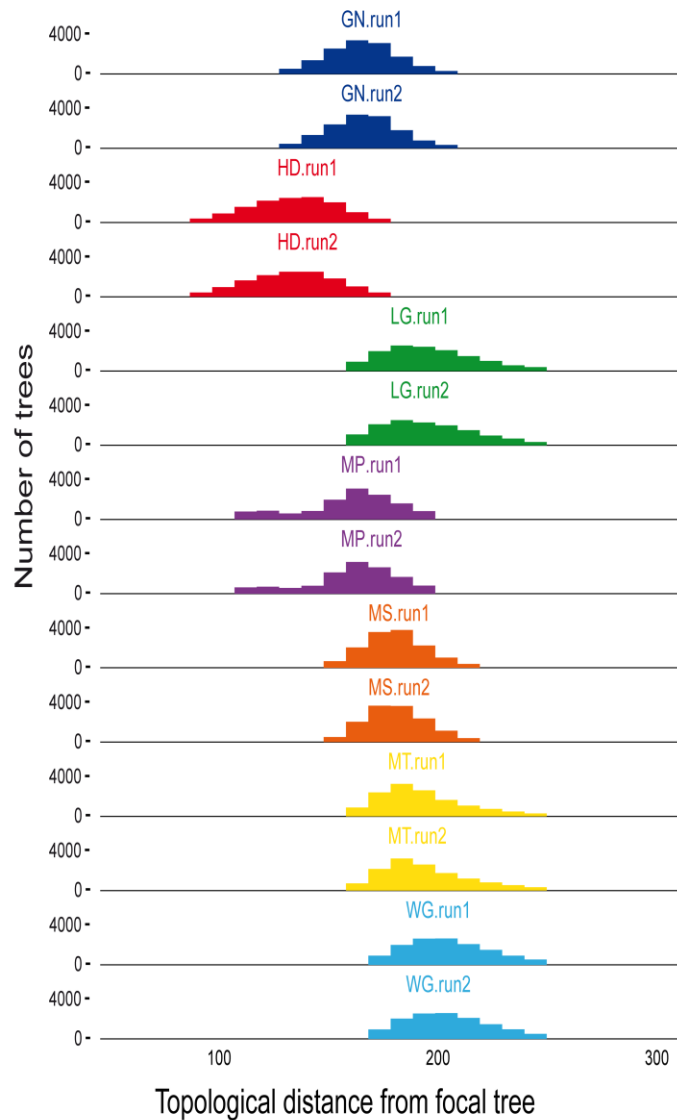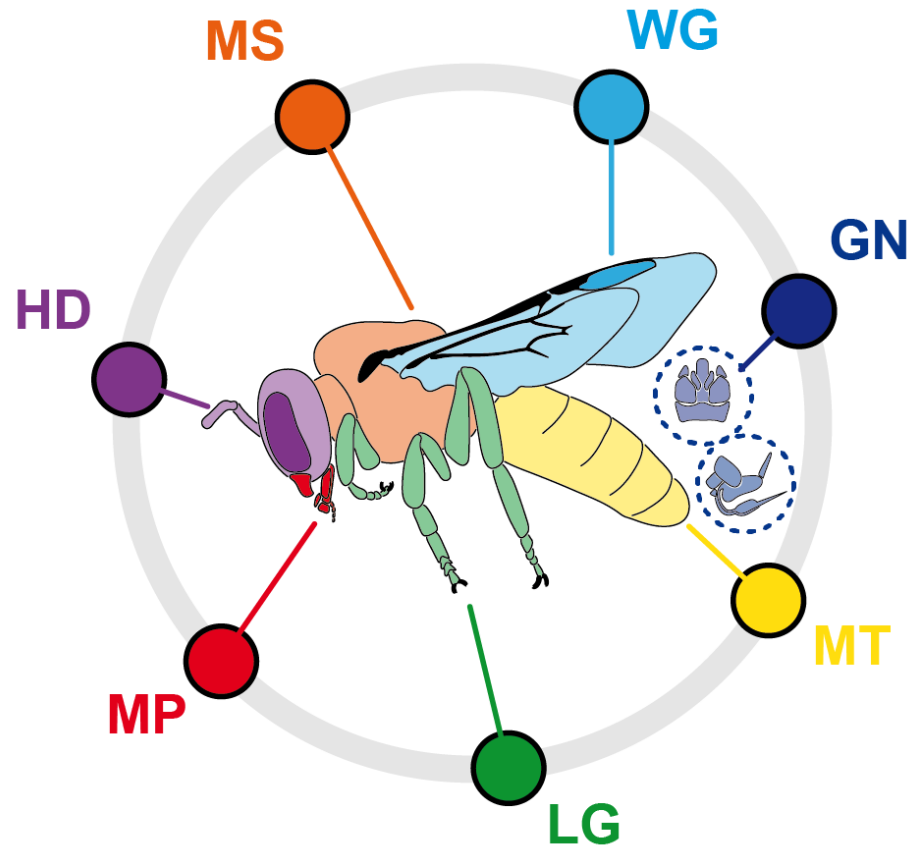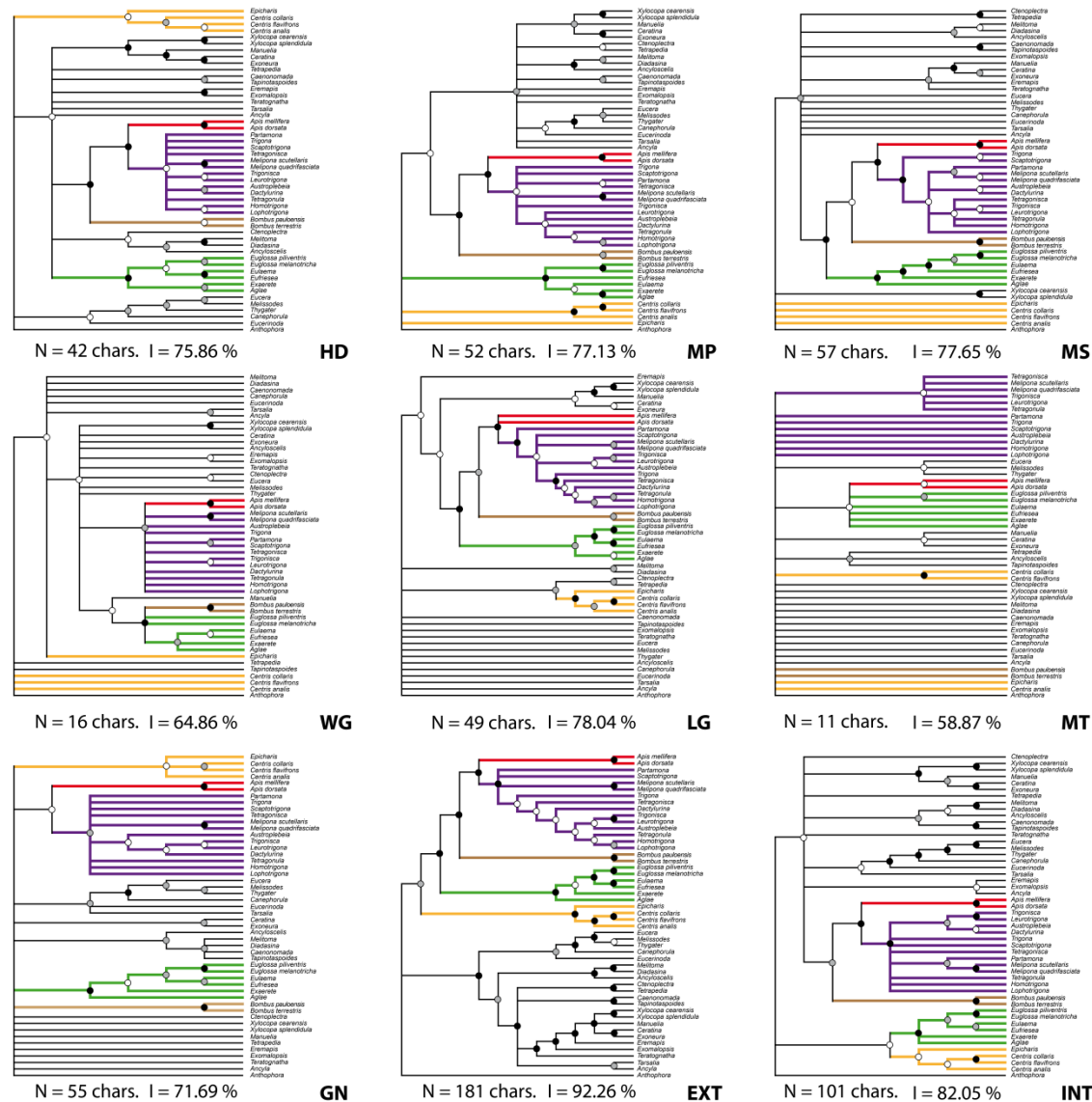
"structure trees"

Porto et al. (2020): Syst. Biol.

Example II: conflict among partitions

species tree

N = 42 chars. I = 75.86 %   **HD**

N = 52 chars. I = 77.13 %   **MP**

N = 57 chars. I = 77.65 %   **MS**

N = 16 chars. I = 64.86 %   **WG**

N = 49 chars. I = 78.04 %   **LG**

N = 11 chars. I = 58.87 %   **MT**

N = 55 chars. I = 71.69 %   **GN**

N = 181 chars. I = 92.26 %   **EXT**

N = 101 chars. I = 82.05 %   **INT**

"structure trees"

Legend:

- ● pp > 0.90
- ◍ 0.9 > pp > 0.75
- ○ 0.75 > pp

N = # of characters
I = Information

Phylogenetic Dissonance (D)

D* = 1.59 %	D* = 0.87 %
D* = 0.92 %	D* = 2.63 %
D* = 1.36 %	D* = 4.31 %
D* = 1.49 %	D** = 14.61 %

D*: dissonance between runs of same partition; D**: dissonance among different partitions.

Porto et al. (2020): Syst. Biol.

Example III: address evolutionary hypotheses

Photographs: D.S. Porto