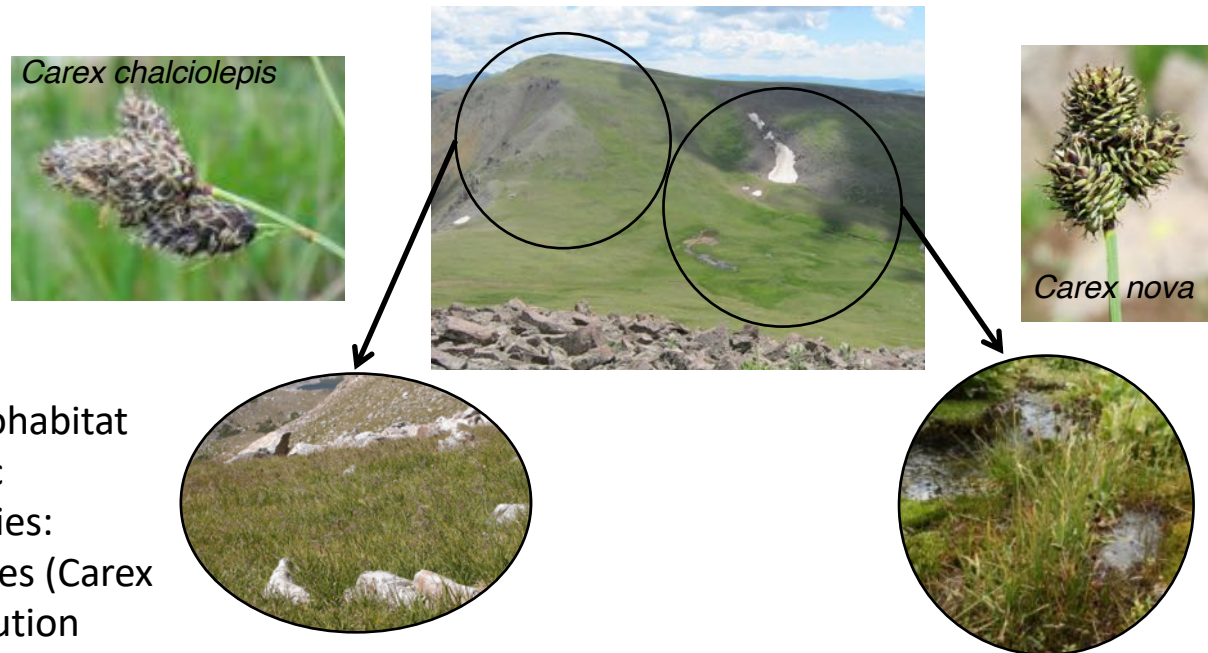# Refined hypotheses based on taxon-specific traits in comparative phylogeography

- statistical tests of discordant phylogeographic structure that is predicted from differences in taxon-specific traits

Carex chalciolepis

Carex nova

Massatti & Knowles LL (2014) Microhabitat differences impact phylogeographic concordance of co-distributed species: genomic evidence in montane sedges (Carex L.) from the Rocky Mountains. Evolution 68:2833-2846.

- Glaciated areas act as barriers, but only in wetland specialist

Communities may be characterized by species-specific responses to climate change

Advantages of iDDC:

- Flexible (expand to multiple species)
- Complex history
- Test of processes

Challenges:

- Computationally intensive
- Model verifications

- Is the model capable of generating the observed data: the likelihood of the empirical data can be compared with the likelihoods of other retained simulations (a p-value of 0 means all the simulations had a better likelihood than the observed data)

- Coefficient of variation of each parameter explained by each PLSs of the summary statistics was computed as an indicator for the power of the estimation

- Accuracy of parameter estimation in the most supported model evaluated using 1000 PODs generated from prior distributions of the parameters

How do we know if we used the "right" model?

In practice we can never completely model the evolutionary process, all we can hope for is that we have captured the important features.

(i.e., IT IS KEY that you are knowledgeable about your system!)

Biological insights:
(i) hypotheses that capture processes structuring genetic variation, and
(ii) model-based approaches to evaluate statistical support for alternative hypotheses

Does microhabitat affect responses to climate change

Massatti & Knowles (2014, 2016)
Evolution, Mol. Ecol.

Biological insights:
(i) hypotheses that capture processes structuring genetic variation, and
(ii) model-based approaches to evaluate statistical support for alternative hypotheses

Does microhabitat affect responses to climate change

Massatti & Knowles (2014, 2016)
Evolution, Mol. Ecol.

Role of habitat stability in structuring genetic variation

He et al (2013) *Evolution*

Present versus past distributions as drivers of divergence
Knowles & Massatti (2017) *Ecography*
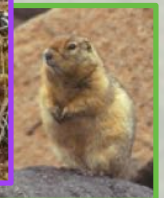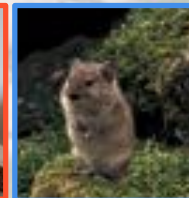
Extent of distributional shifts or rate of climatic change as determinants of concordant patterns of genetic structure

Knowles et al. (2016) *J. Biogeogr.*
He et al. (2017) *Mol Ecol.*

How best to extract information from genetic data about the underlying evolutionary history?

$\Rightarrow$ use of summary statistic (sacrifices information content for simplification and ease)

- observed quantities are compared to expectations

Summary statistics of genetic variation will have different values depending upon the biogeographic and demographic processes generating the genetic data

Summaries of genetic variation

Population subdivision

Tajima's D > 0

θ, pi

Mismatch distribution
(Rogers & Harpending 1992)

Population growth

Tajima's D < 0

θ, pi

Mismatch distribution
(Rogers & Harpending 1992)

How best to extract information from genetic data about the underlying evolutionary history?

$\Rightarrow$ use of summary statistic (sacrifices information content for simplification and ease)

•   observed quantities are compared to expectations

$\Rightarrow$ calculate full likelihood of the sequence data (computationally demanding, and may not work for complex models, but makes full use of the data)

Understanding the effects of rapid climate change on species diversity:

Did the frequent and repeated shifts in species distribution in response to the Pleistocene glacial cycles promote or inhibit divergence?
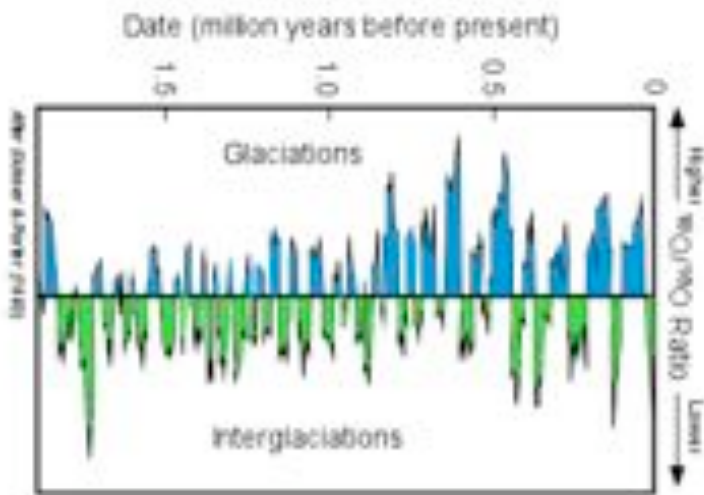
What is the timing of speciation?
• Pleistocene versus pre-Pleistocene?
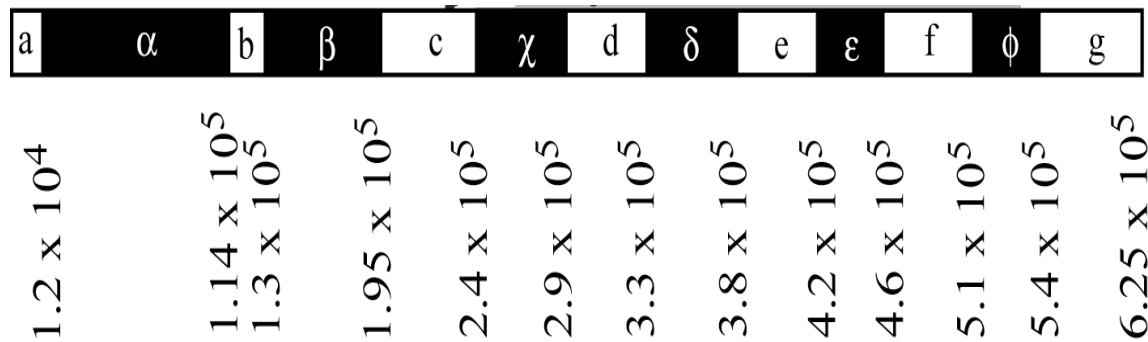• Glacial versus inter-glacial?

Timing of divergence?
- Pleistocene versus pre-Pleistocene?
- Glacial versus inter-glacial?
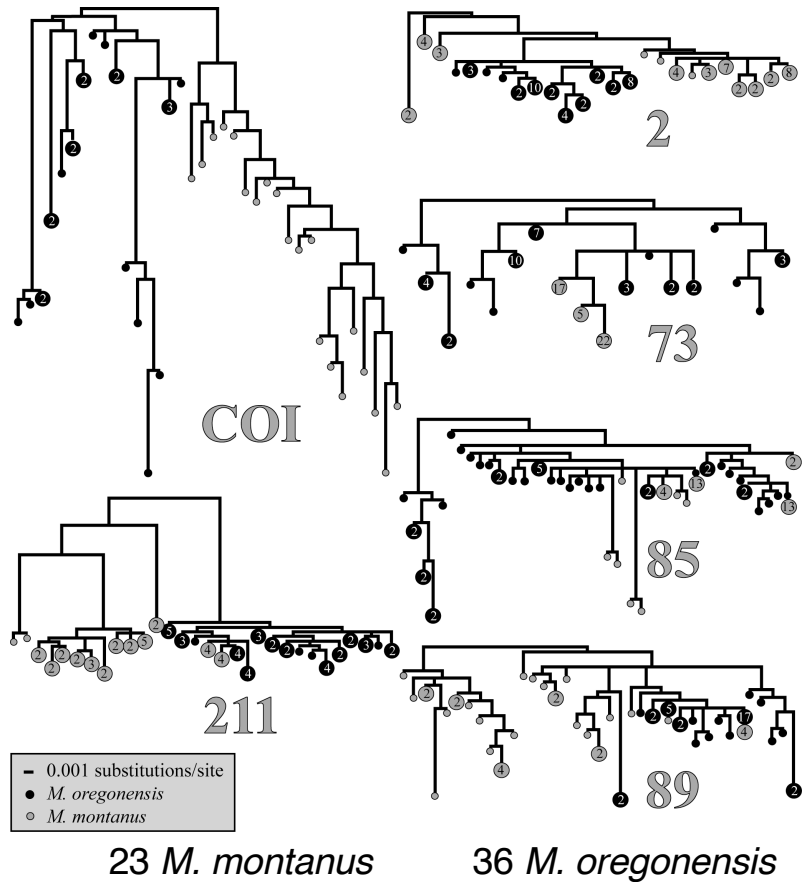

T
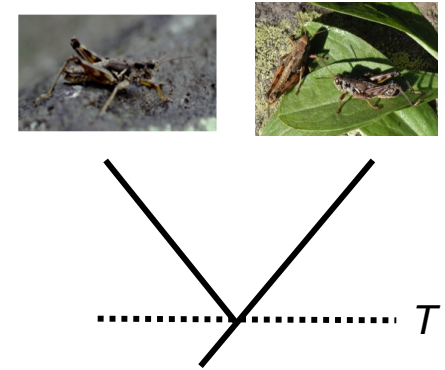
Insufficient resolution of
past genetic approaches

estimate from average
mtDNA genetic distance:
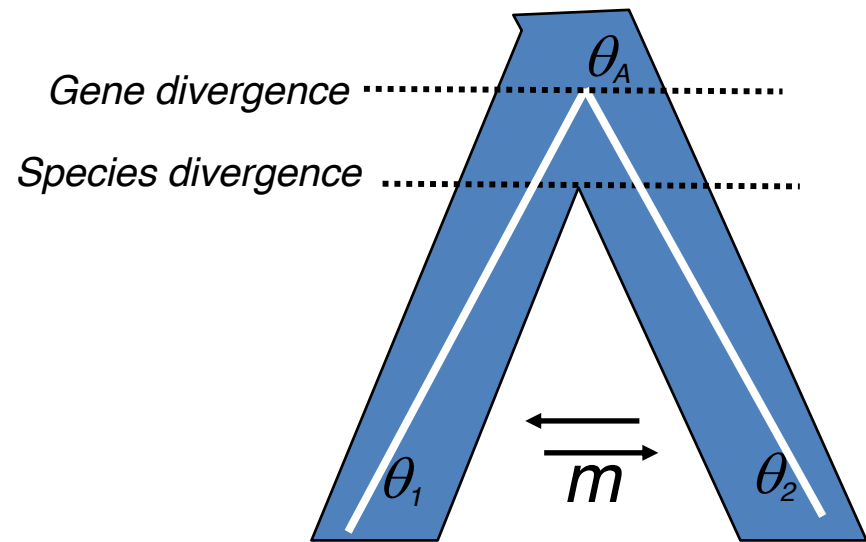
4.9×10^5 to 2.0×10^6 years ago

| a | α | b | β | c | χ | d | δ | e | ε | f | φ | g |

1.2×10^4

1.14×10^5
1.3×10^5

1.95×10^5

2.4×10^5

2.9×10^5

3.3×10^5

3.8×10^5

4.2×10^5

4.6×10^5

5.1×10^5

5.4×10^5

6.25×10^5

■ glacial period
□ interglacial period

Knowles 2001, Evolution

- Use multilocus data and a coalescent framework to estimate the timing of divergence

COI

2

73

85

211

89

- 0.001 substitutions/site
- *M. oregonensis*
- *M. montanus*

23 *M. montanus*   36 *M. oregonensis*

(Bayesian program IM)

Gene divergence $\cdots\cdots\cdots$ θ_A $\cdots\cdots\cdots$

Species divergence $\cdots\cdots\cdots\cdots\cdots\cdots$

θ_1 $\quad \overline{m}$ $\quad \theta_2$

Timing of divergence?

Carstens & Knowles 2007, Mol. Ecol. 16:619-27.

Precise estimate of T suggests species diverged during a glacial period

Single locus, coalescent approach

estimate from average mtDNA genetic distance:

4.9×10^5 to 2.0×10^6

Multilocus, coalescent approach

| a | α | b | β | c | χ | d | δ | e | ε | f | φ | g |

1.2×10^4 1.14×10^5 1.3×10^5 1.95×10^5 2.4×10^5 2.9×10^5 3.3×10^5 3.8×10^5 4.2×10^5 4.6×10^5 5.1×10^5 5.4×10^5 6.25×10^5

■ glacial period
☐ interglacial period
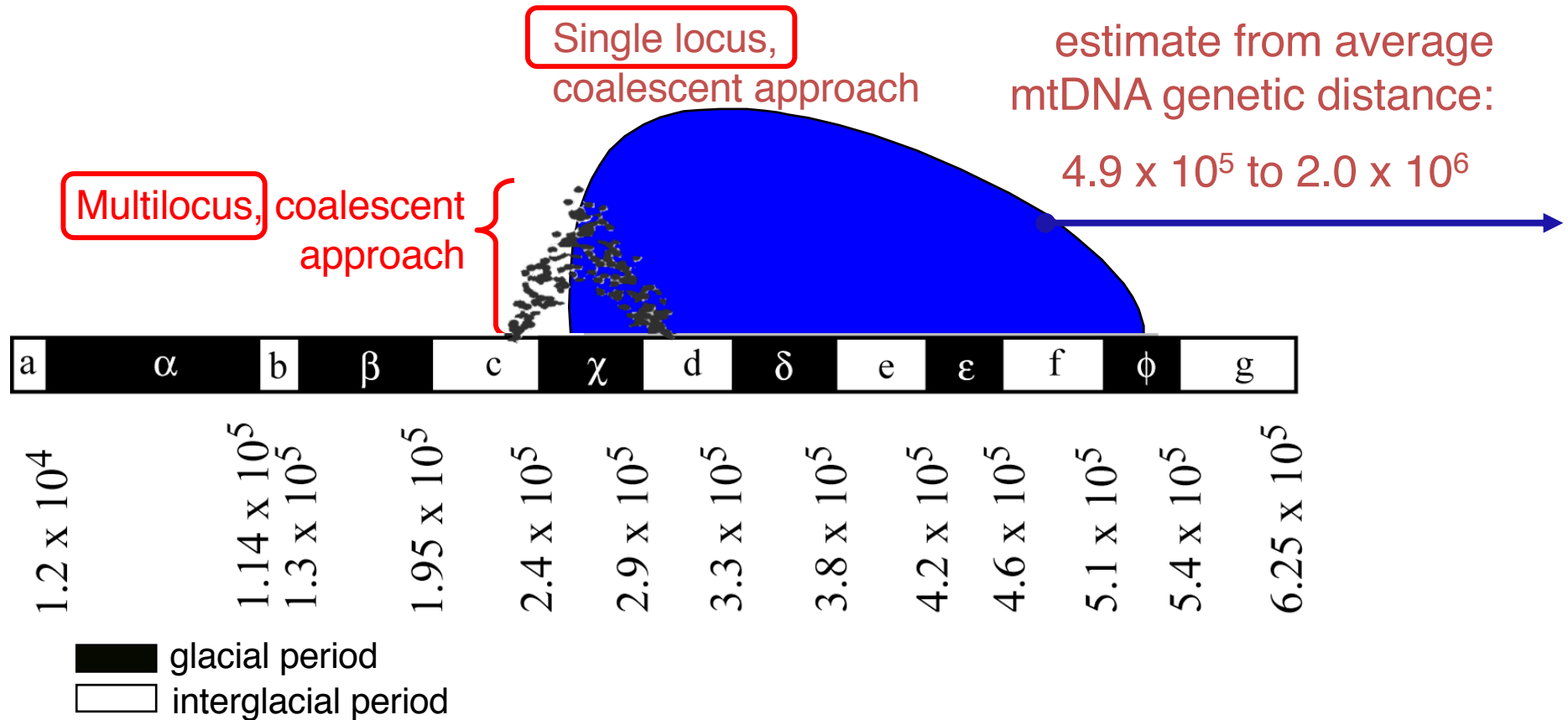
*same mutation rate used in the different approaches

Carstens & Knowles 2007, Mol. Ecol. 16:619-27.

Verified the accuracy of the speciation model given the data

(estimates may be compromised when the complexity of the
model exceeds the information content of the genetic data)

• Simulate genetic data under divergence models estimated for the
empirical grasshopper data and ask if recover the parameter *t*
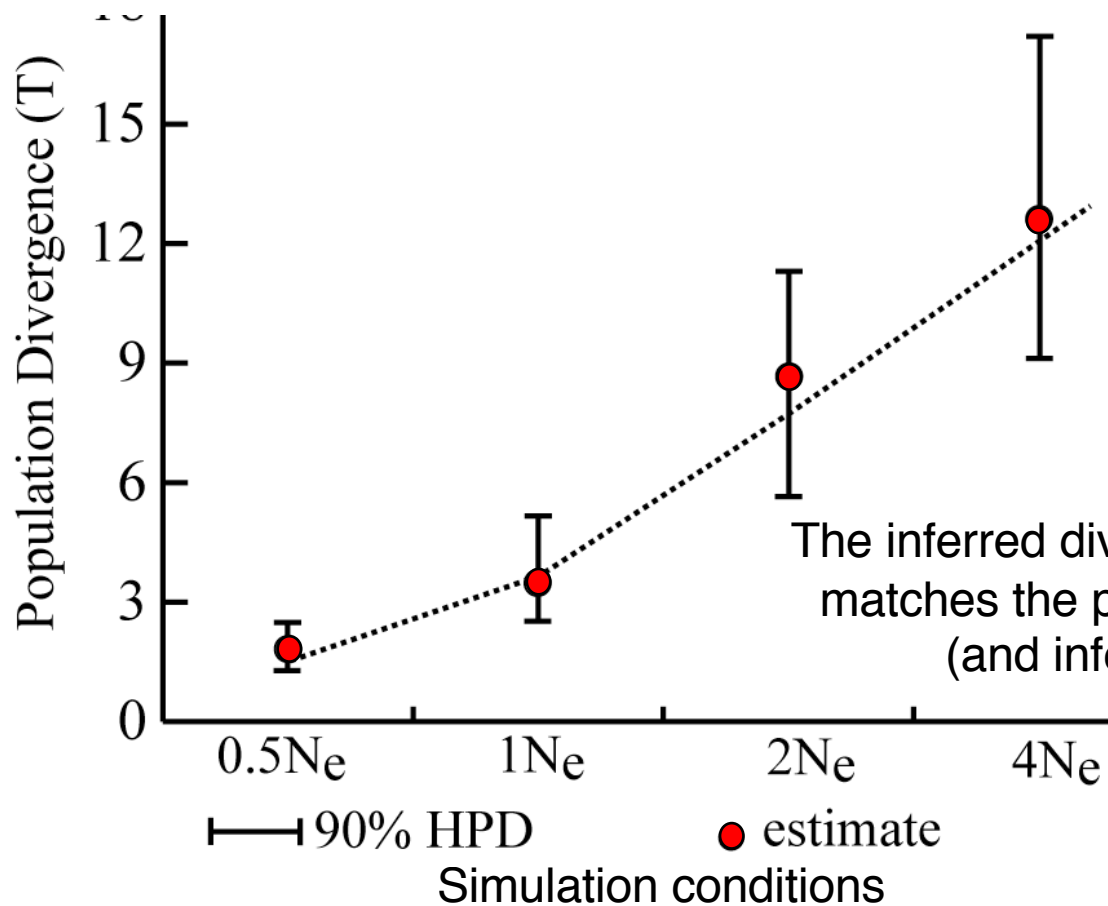
The inferred divergence time of speciation generally
matches the parameter used to simulate the data
(and inferred from the empirical data)

Simulation conditions
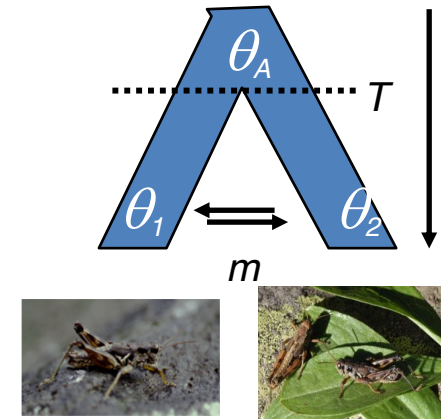
Inference in comparative phylogeography

• Impact of how we use similarity of the association between genes and geography across species to test evolutionary hypotheses

• Importance of considering refined-hypotheses based on taxon-specific traits

Refined hypotheses based on taxon-specific traits in comparative phylogeography

Soil inhabitants

Aegean archipelago
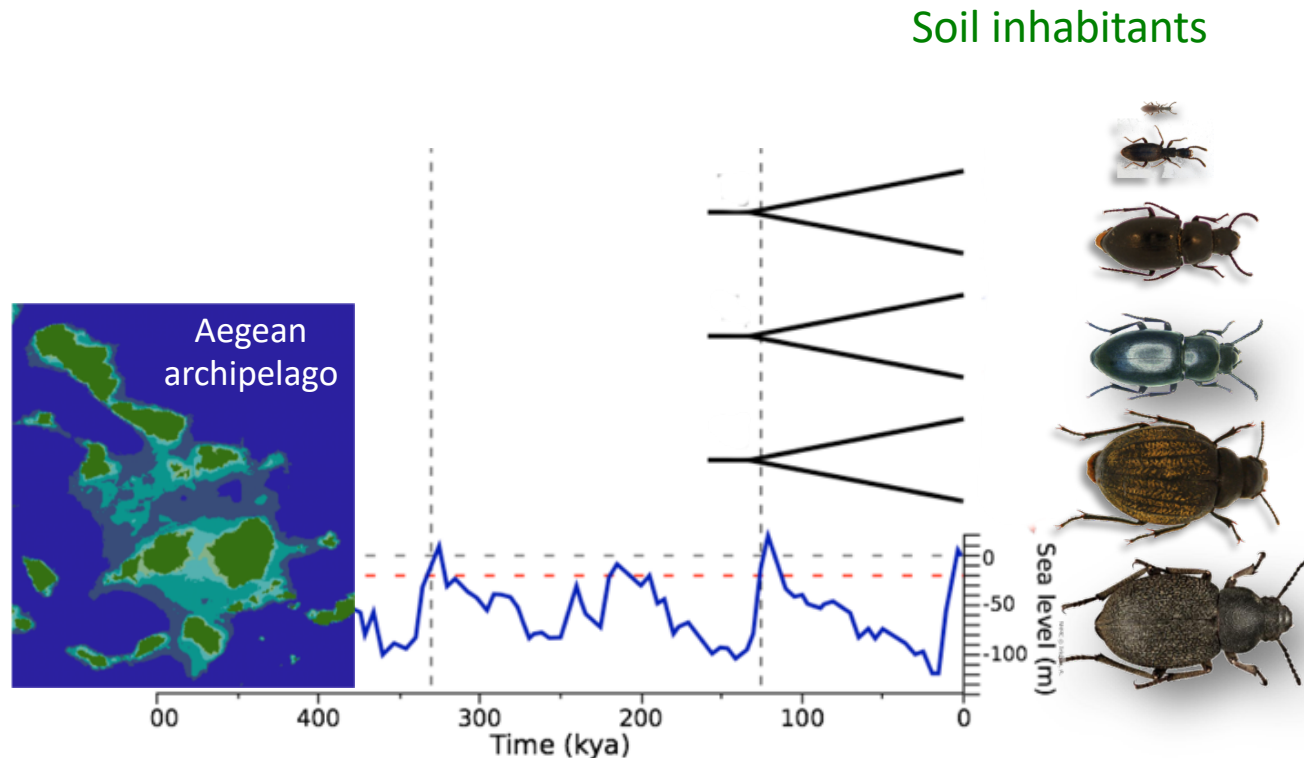
- key to avoid misleading inference

- bias toward tests of the effects of abiotic factors if rely on similarity in genetic structure across taxa for hypothesis testing

Genes and Geography Across Species

Neutral loci
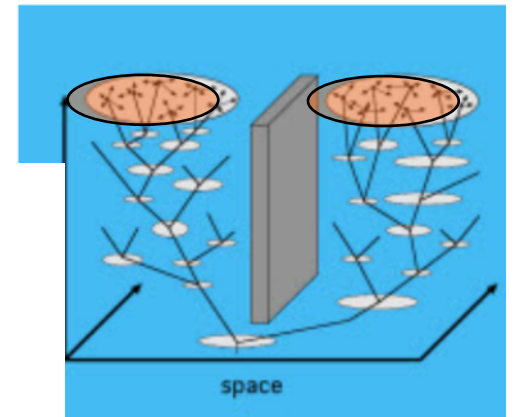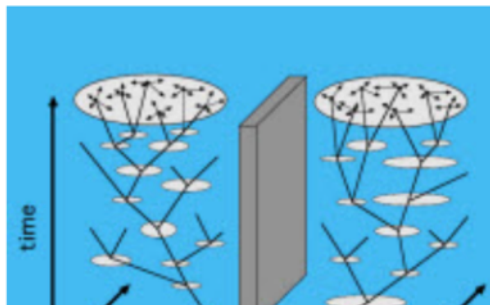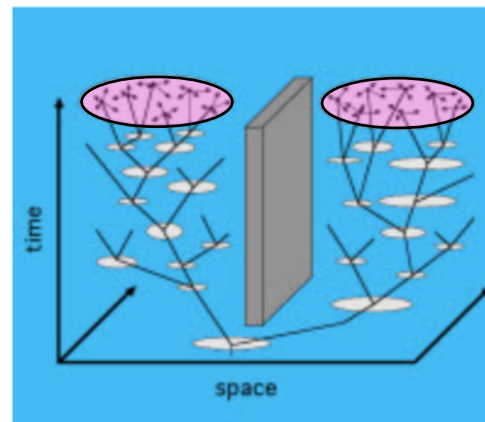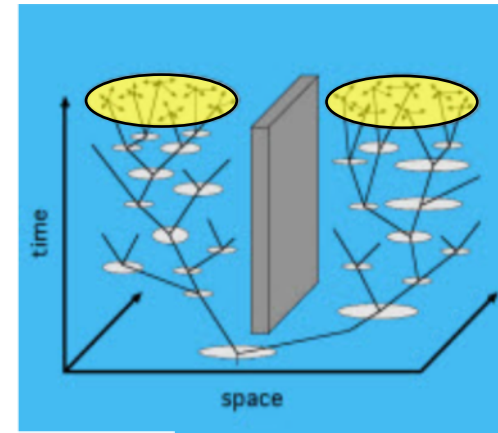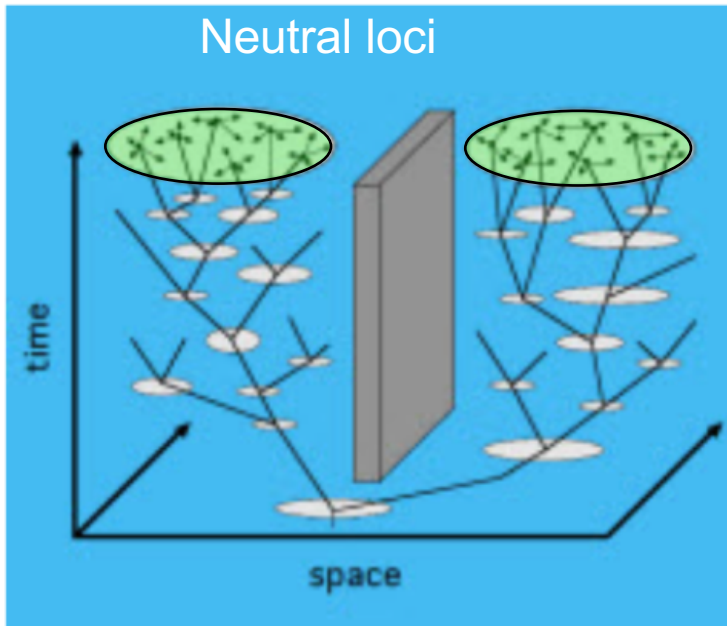
similarity of the association between genes and geography across species – CONCORDANCE – is typically used to test evolutionary hypotheses
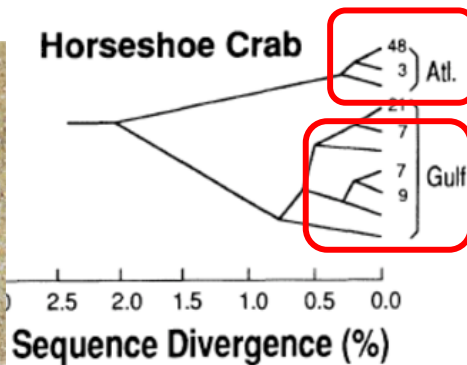
Concordance used in descriptive studies

Black Sea Bass — Atl. (17), Gulf

Seaside Sparrow — Atl. (17), Gulf (14)

Horseshoe Crab — Atl. (48, 3), Gulf (7, 7, 9)

American Oyster — Atl., Gulf

Sequence Divergence (%)
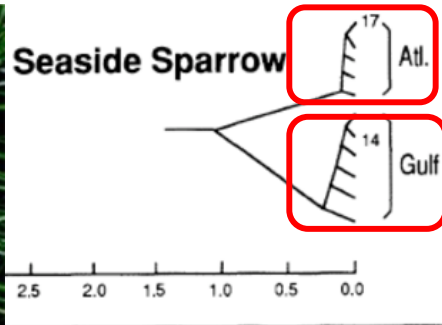
Avise 1992

Concordance used in statistical phylogeography

Species 1   Species 2   Species 3   Species 4
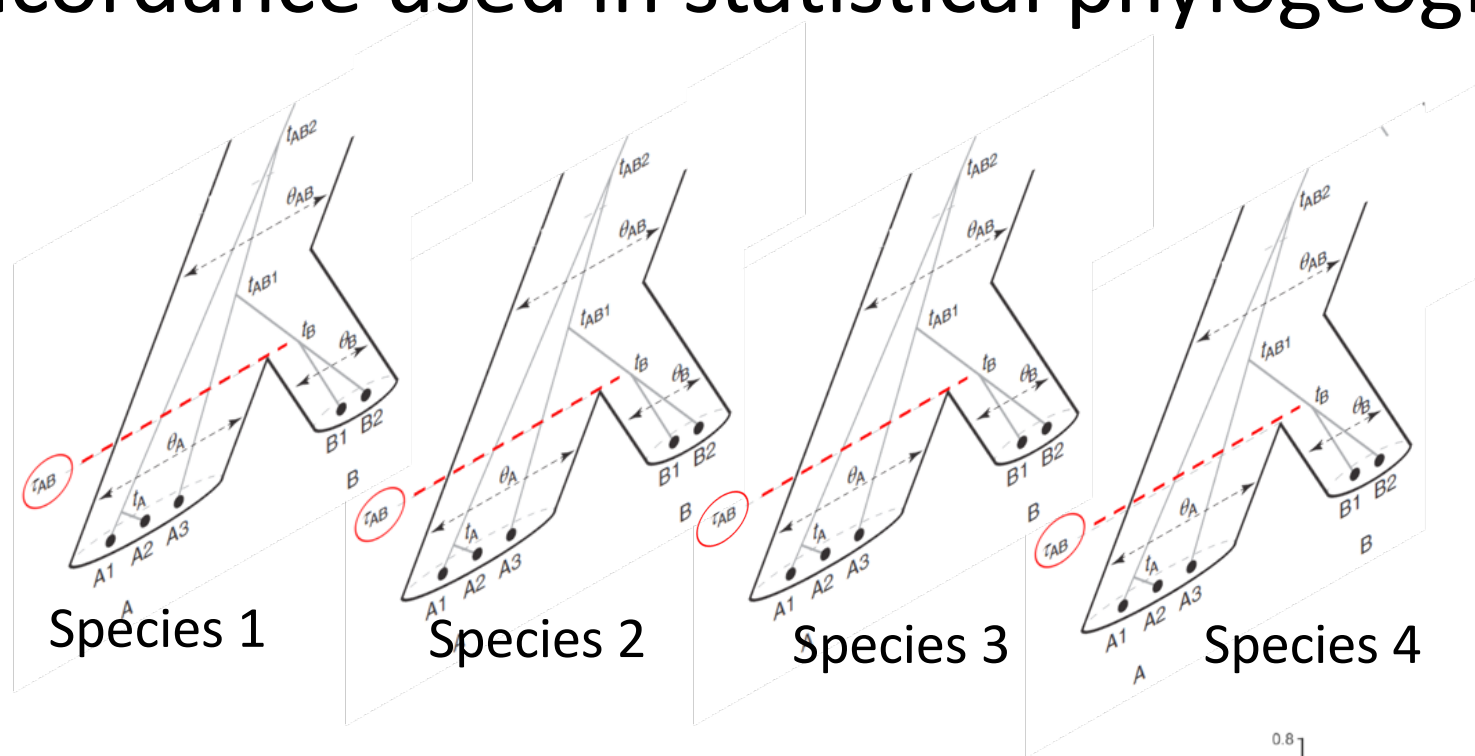
Statistically evaluate a parameterized model of co-divergence among species using hierarchical Approximate Bayesian Computation (hABC)
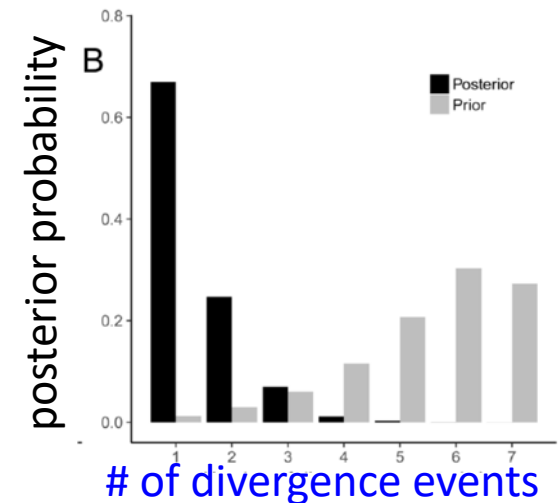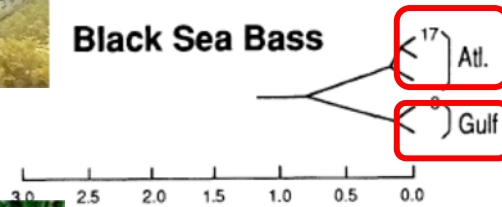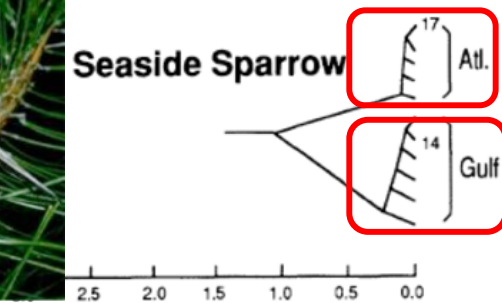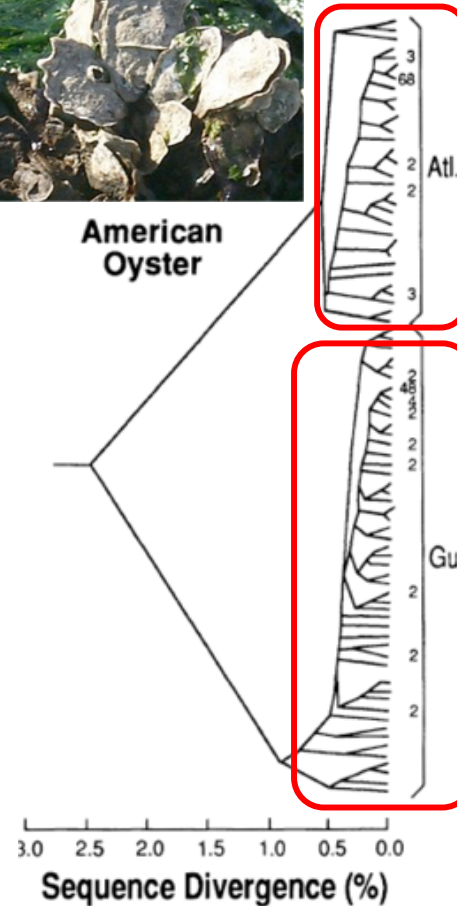
Concordance to test hypotheses

Black Sea Bass
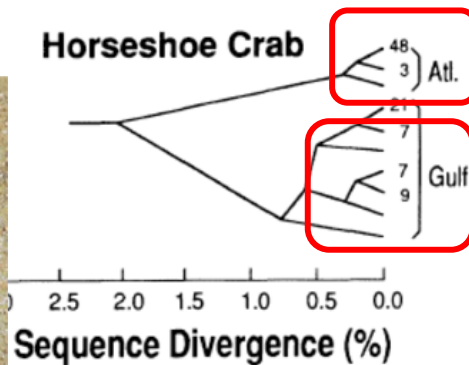
Seaside Sparrow

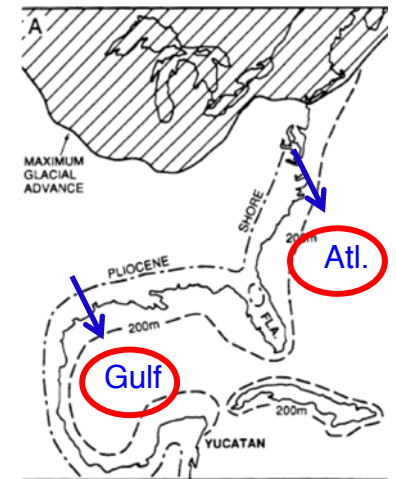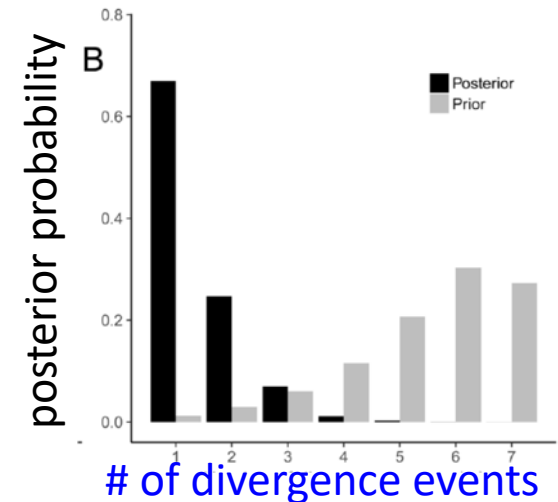Horseshoe Crab

Sequence Divergence (%)
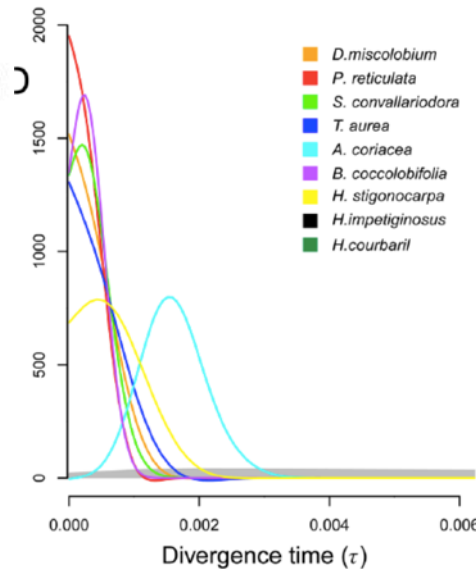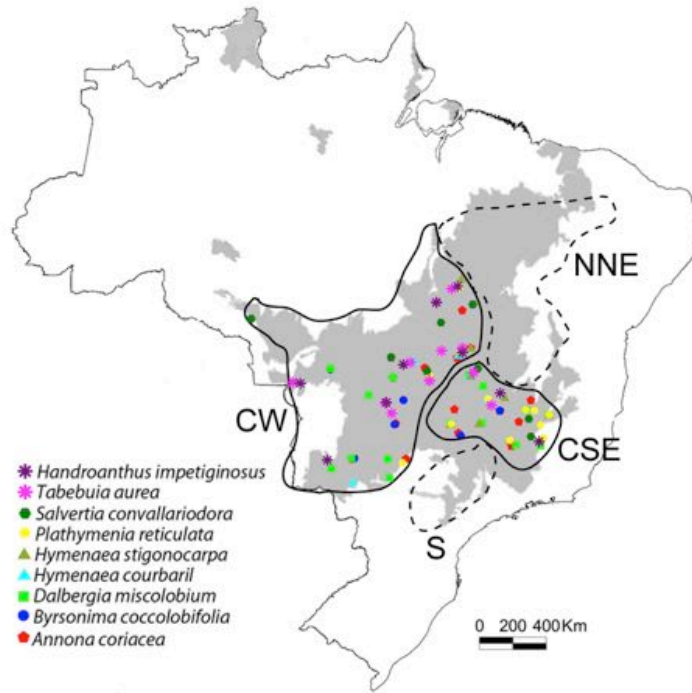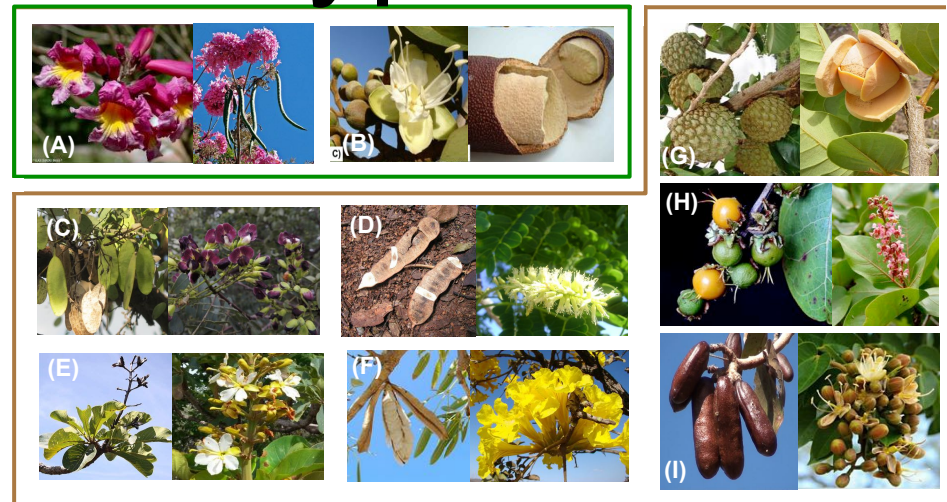
American Oyster

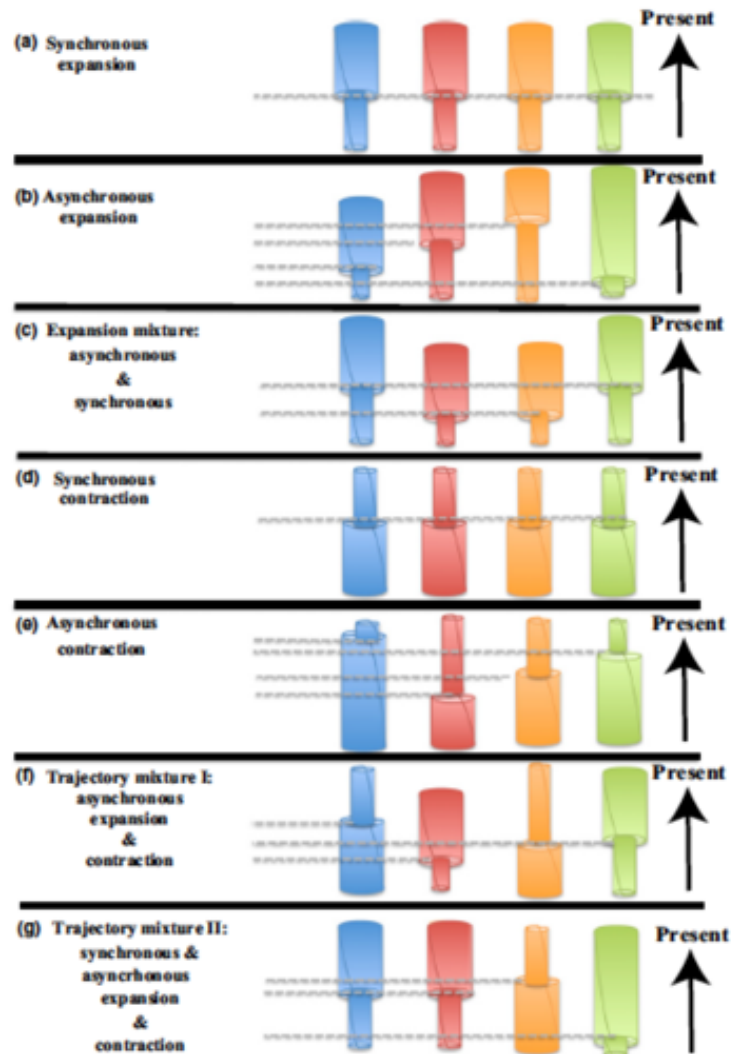Sequence Divergence (%)

Biogeographic barrier

Avise 1992

Concordance to test hypotheses

Estimate degree of co-divergence among species to evaluate hypothesized barrier associated with floristic provinces in Cerrado

concordance is a parameter in model that is estimated from genetic data across multiple species

Resende et al. (in prep)

Concordance to test hypotheses

(a) Synchronous expansion — Present

(b) Asynchronous expansion — Present

(c) Expansion mixture: asynchronous & synchronous — Present

(d) Synchronous contraction — Present

(e) Asynchronous contraction — Present

(f) Trajectory mixture I: asynchronous expansion & contraction — Present

(g) Trajectory mixture II: synchronous & asynchronous expansion & contraction — Present

Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates
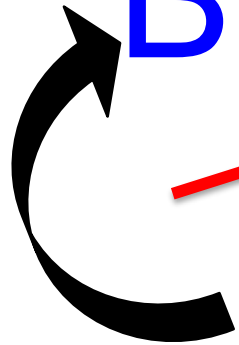
Burbrink et al. 2016

Elaboration of statistical methods for testing concordance, including tests of co-expansion

See also Oaks 2020 SystBiol.

ecoevolity,   https://doi.org/10.1093/sysbio/syy063

Genes and Geography across species

Biotic component?

CONCORDANCE

for testing hypotheses
about evolutionary history

• potential for misleading inference by not considering
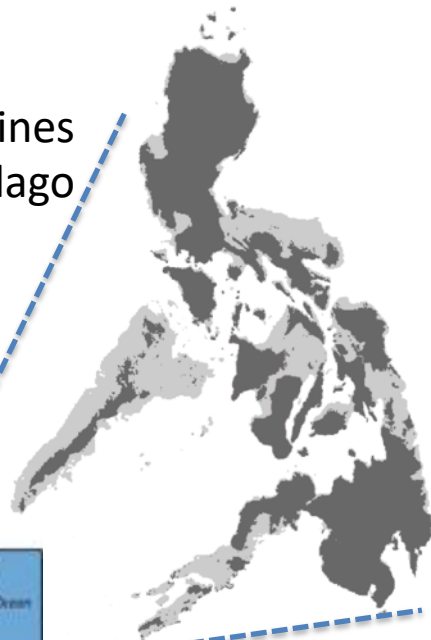both biotic and abiotic components
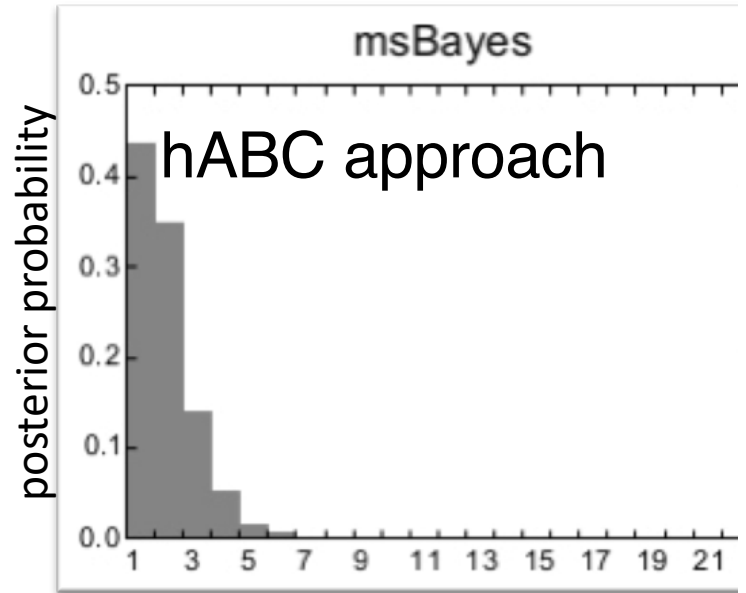
Concordance criteria for hypothesis testing

Hypothesis of simultaneous divergence to test whether sea-level oscillations during the Pleistocene caused diversification

Oaks et al. (2012) *Evolution*

Phillipines
Archipelago

hABC approach

posterior probability

Number of divergence events

- Inferred the distribution of divergence times among 22 pairs of co-distributed vertebrate taxa

changes in connectivity/isolation of islands with sea-level changes (light versus dark grey outlines)

Sea level (m)

Time (kya)

Concordance criteria for hypothesis testing

Hypothesis of simultaneous divergence to
test whether sea-level oscillations during
the Pleistocene caused diversification

Performed a suite of simulation-based power analyses
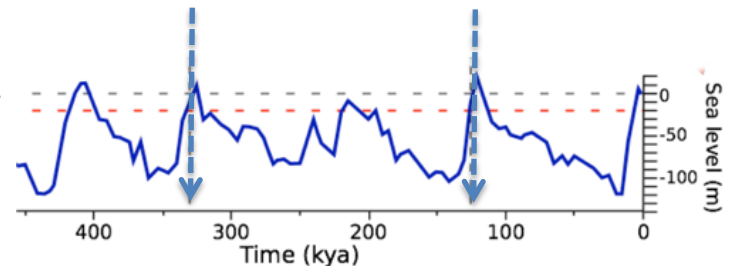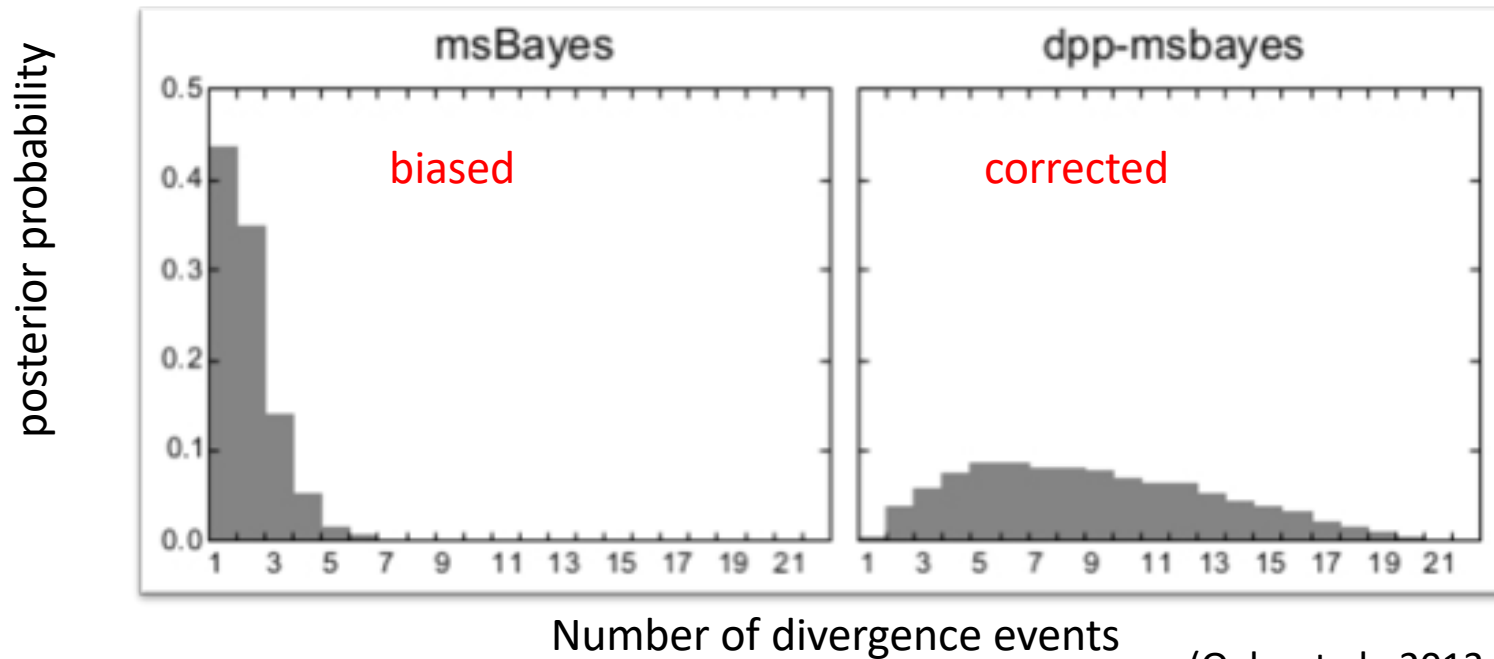

(Oaks et al., 2013; Oaks, 2014)

Should this be interpreted as a rejection of the "species pump" model of
diversification in which sea-level changes drive divergence?

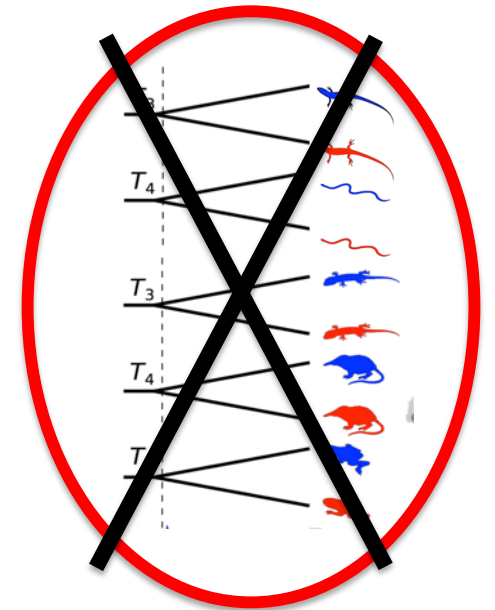# Hypothesis of phylogeographic concordance Is TOO Generic

Hypothesis of simultaneous divergence to test whether sea-level oscillations during the Pleistocene caused diversification

Phillipines Archipelago

(Oaks et al., 2013; Oaks, 2014)

Should this be interpreted as a rejection of the "species pump" model of diversification in which sea-level changes drive divergence?

~~Generic~~ Refined   hypothesis of phylogeographic concordance
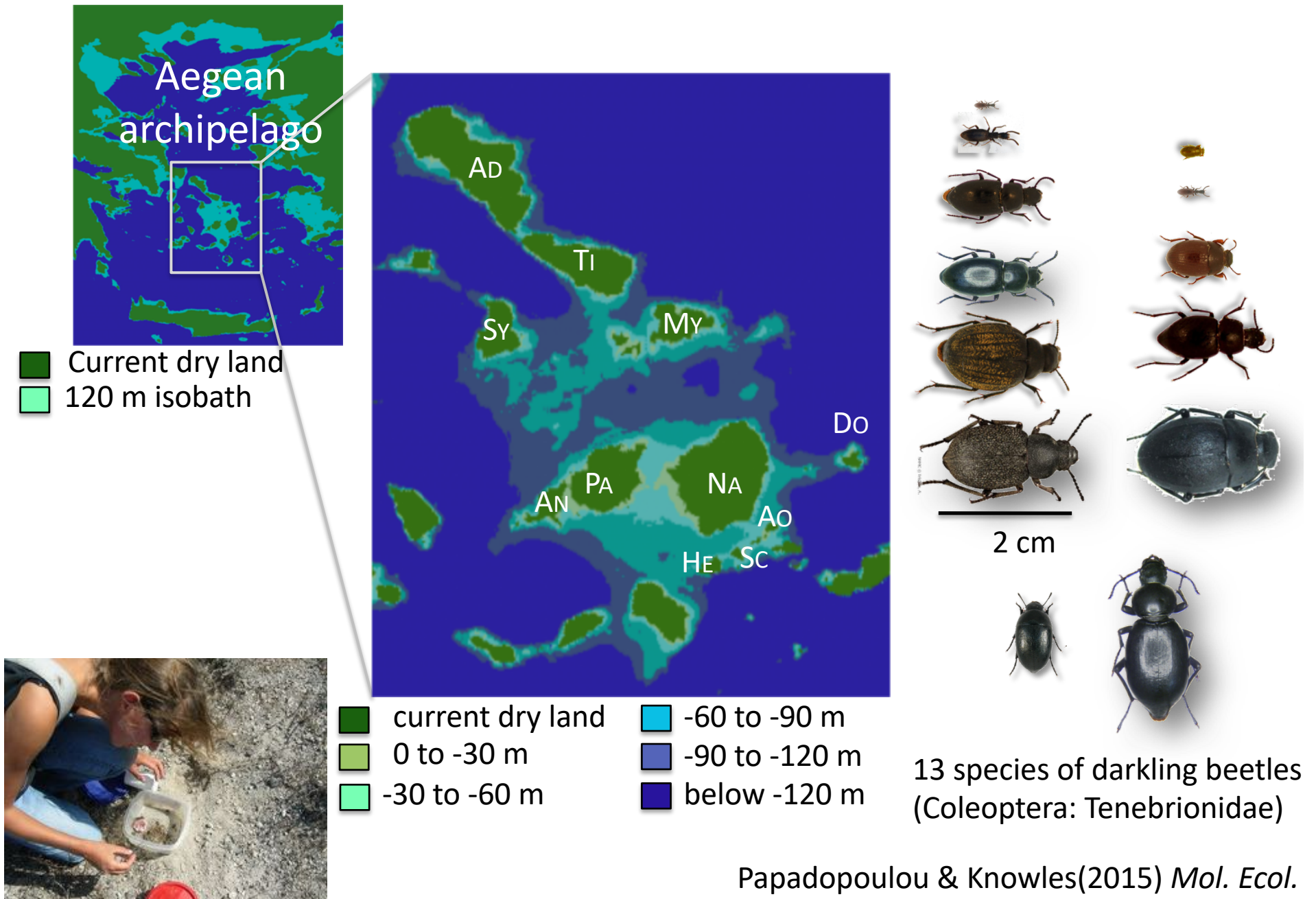
• a study design that considers taxon attributes

Hypothesis of simultaneous divergence to
test whether sea-level oscillations during
the Pleistocene caused diversification

Phillipines
Archipelago

(Oaks et al., 2013; Oaks, 2014)

Refined models of phylogeographic concordance to test the "species pump" model

Aegean archipelago

- ■ Current dry land
- ■ 120 m isobath

AD
TI
SY
MY
DO
PA
NA
AN
AO
HE
SC

current dry land | -60 to -90 m
0 to -30 m | -90 to -120 m
-30 to -60 m | below -120 m

2 cm

13 species of darkling beetles
(Coleoptera: Tenebrionidae)

Papadopoulou & Knowles(2015) *Mol. Ecol.*

• taxa differ in their soil associations

Aegean archipelago

Sandy habitats
"Soil" habitats
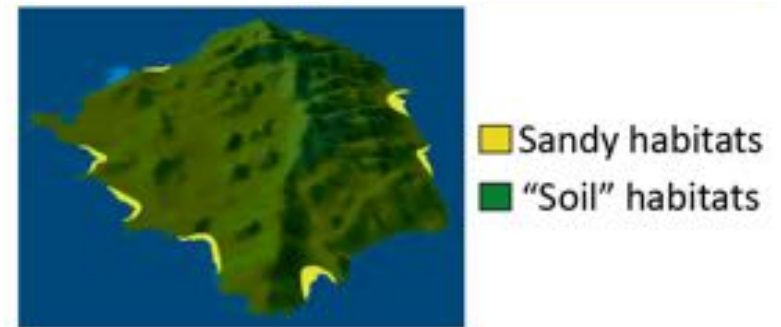
Ephemerality of sand habitats may supersede effects of sea-level connections!

Soil – stable habitat

2 cm

Sand – disturbed habitat

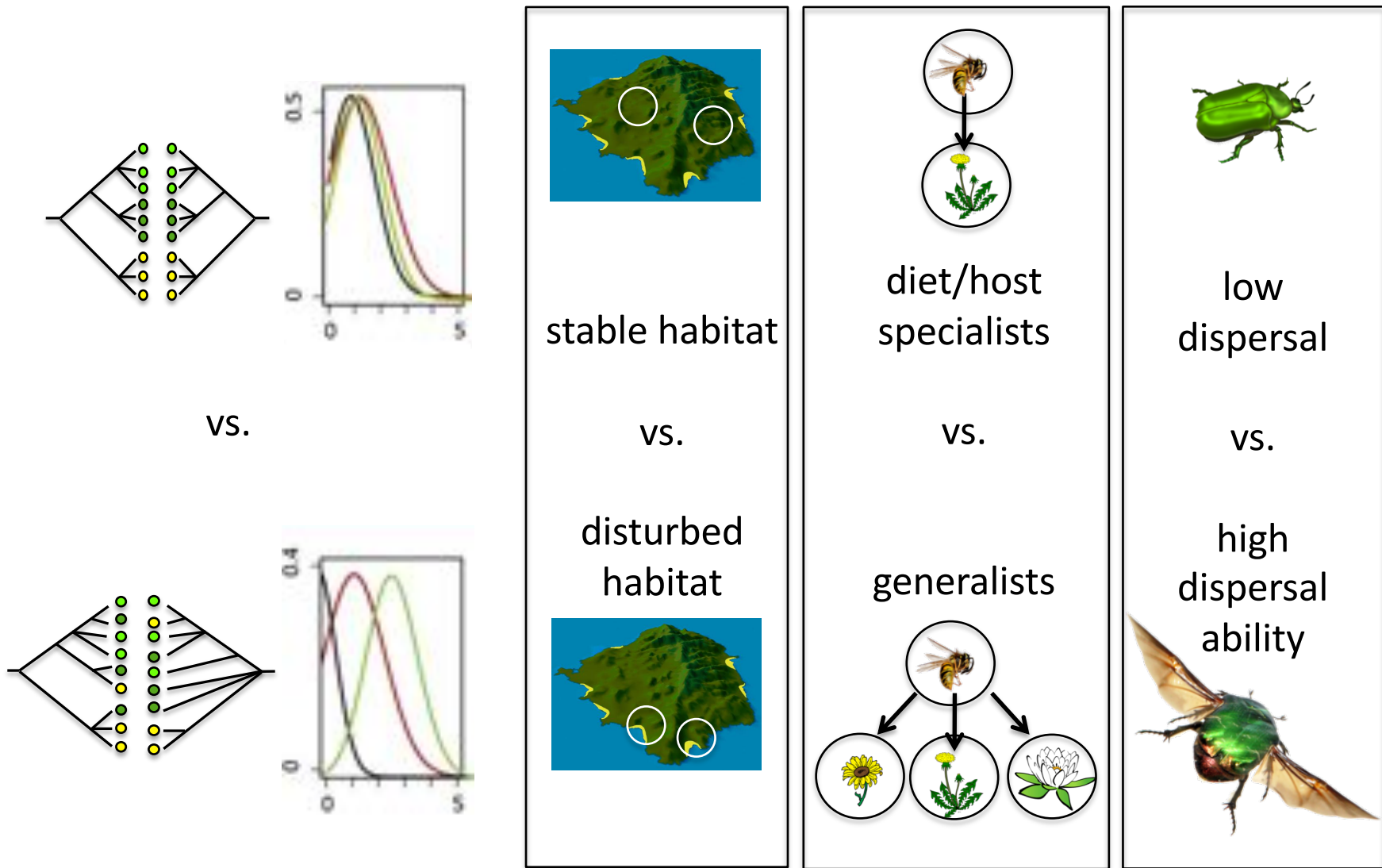Generalists on both soil & sand

• uniform trophic ecology & inherent dispersal abilities

Papadopoulou & Knowles(2015) *Mol. Ecol.*

Which species-trait to consider for refined hypothesis testing?

What ecological traits may be used to refine hypotheses about predicted phylogeographic concordance vs. discordance?

vs.

stable habitat

vs.

disturbed habitat

diet/host specialists

vs.

generalists

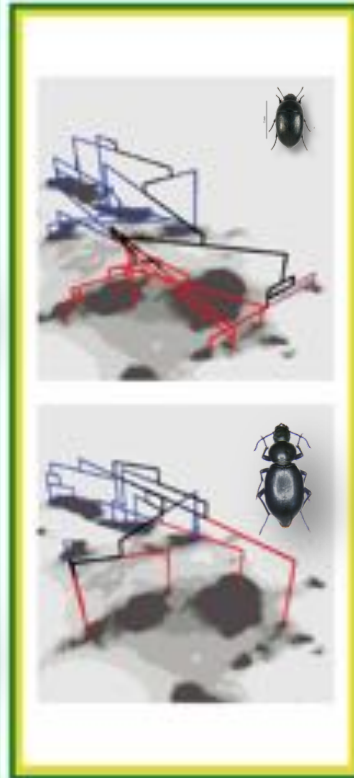low dispersal

vs.

high dispersal ability

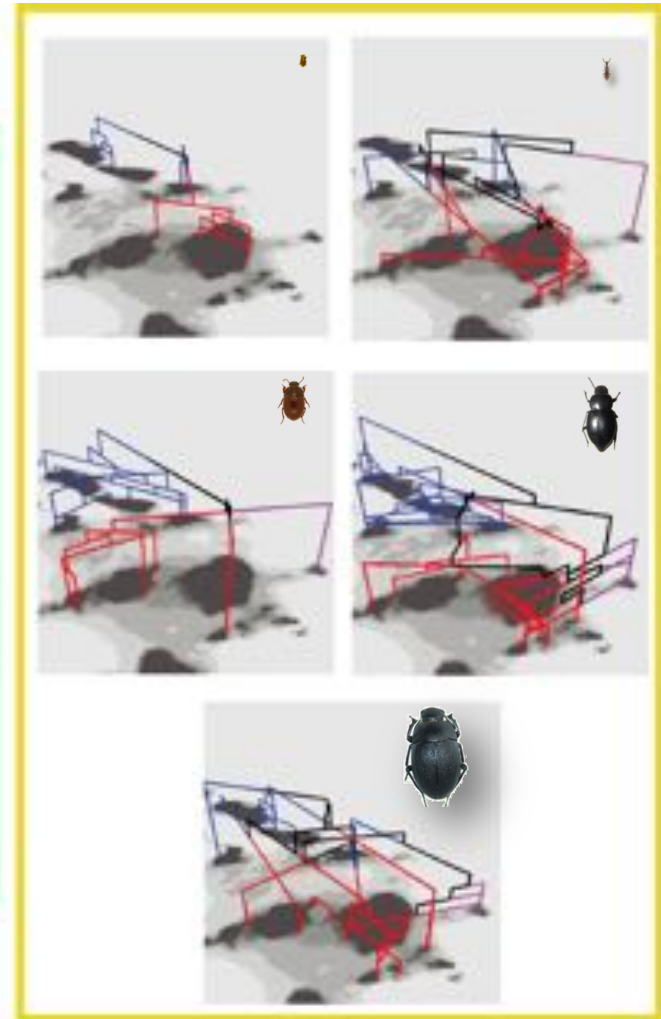# Different degrees of structure of mtDNA gene trees suggestive of differences in habitat stability

Soil – stable habitat
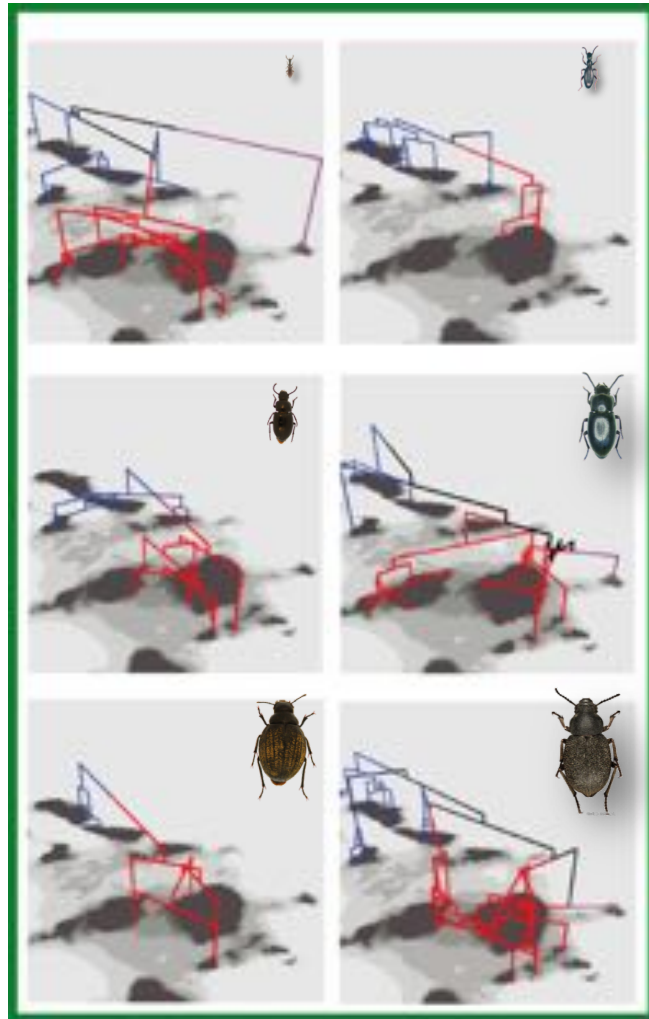
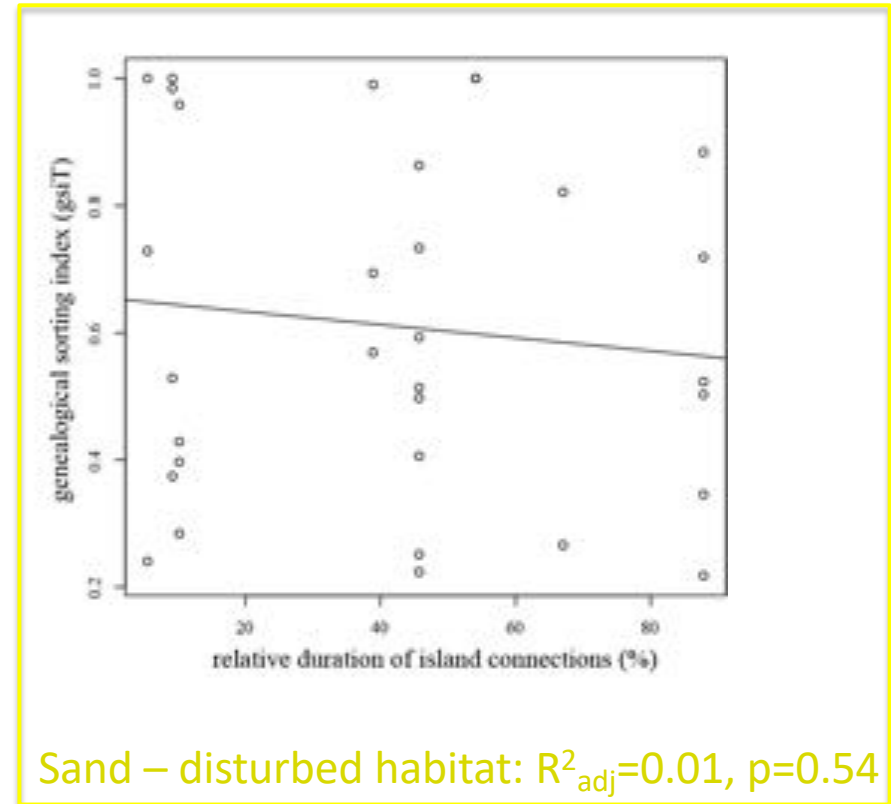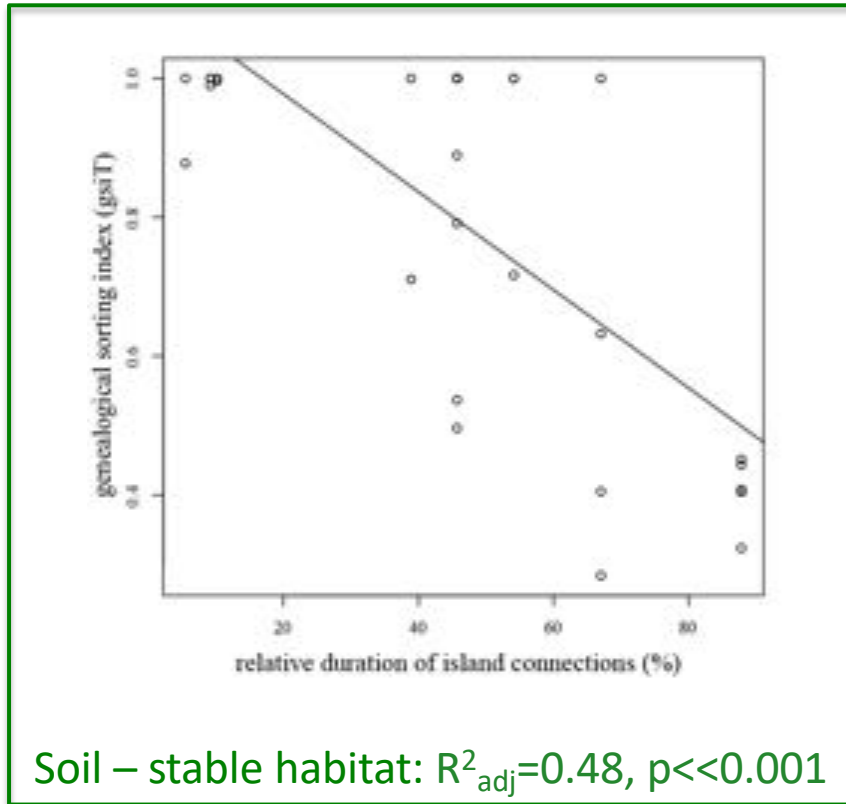generalists both stable and disturbed habitats

Sand – disturbed habitat

Northern Islands bathymetrically separated by 95m trench from   Southern islands

Degree of lineage sorting correlated with duration of island connections?

Soil – stable habitat: $R^2_{adj}=0.48$, p<<0.001

Sand – disturbed habitat: $R^2_{adj}=0.01$, p=0.54

Model comparisons in subsequent analyses also identified the relative duration of island connection in combination with habitat type as the best predictors of genealogical sorting (in contrast to other explanatory variables such as body size or island size) based on AICs

Papadopoulou & Knowles (2015) *Mol. Ecol.* 24: 4252-4268

Refined hypothesis for tests of concordance that focus on stable-habitat taxa

Test of simultaneous divergence
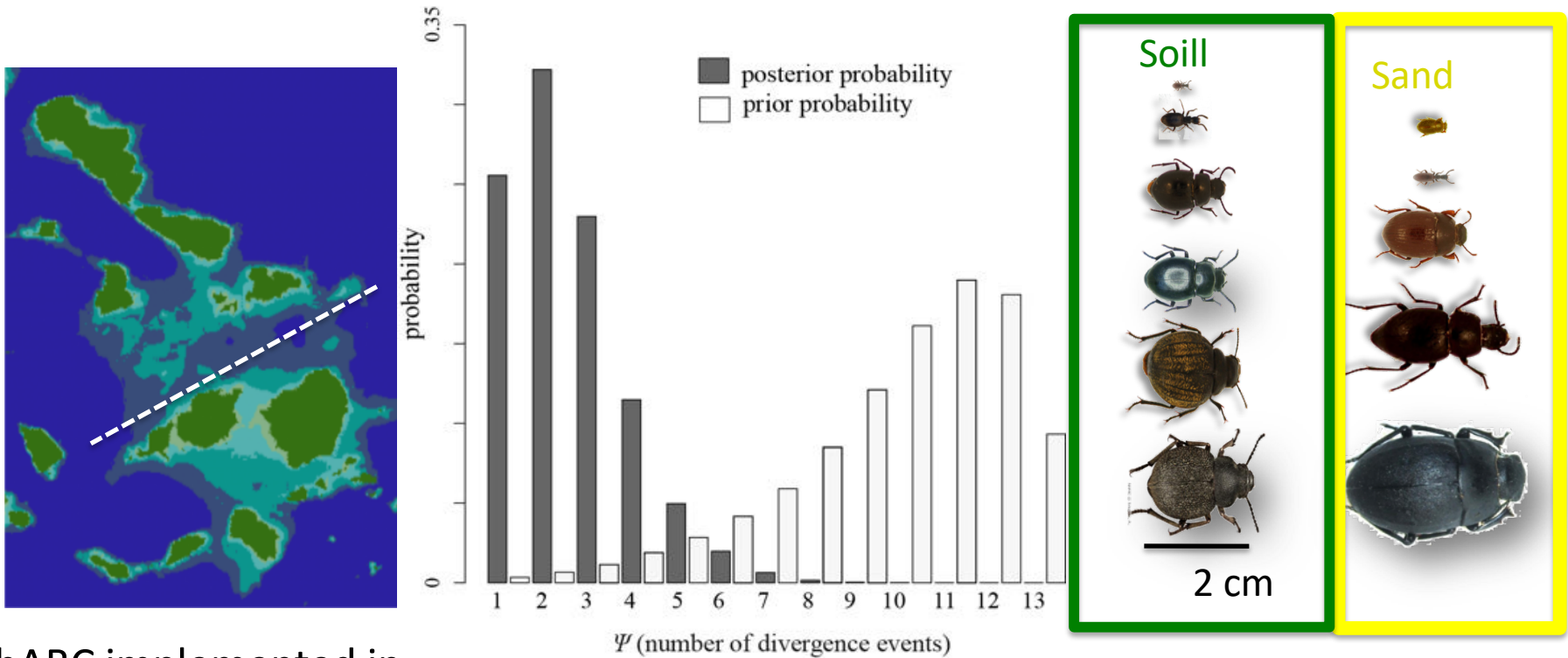

Soil – stable habitat

2 cm
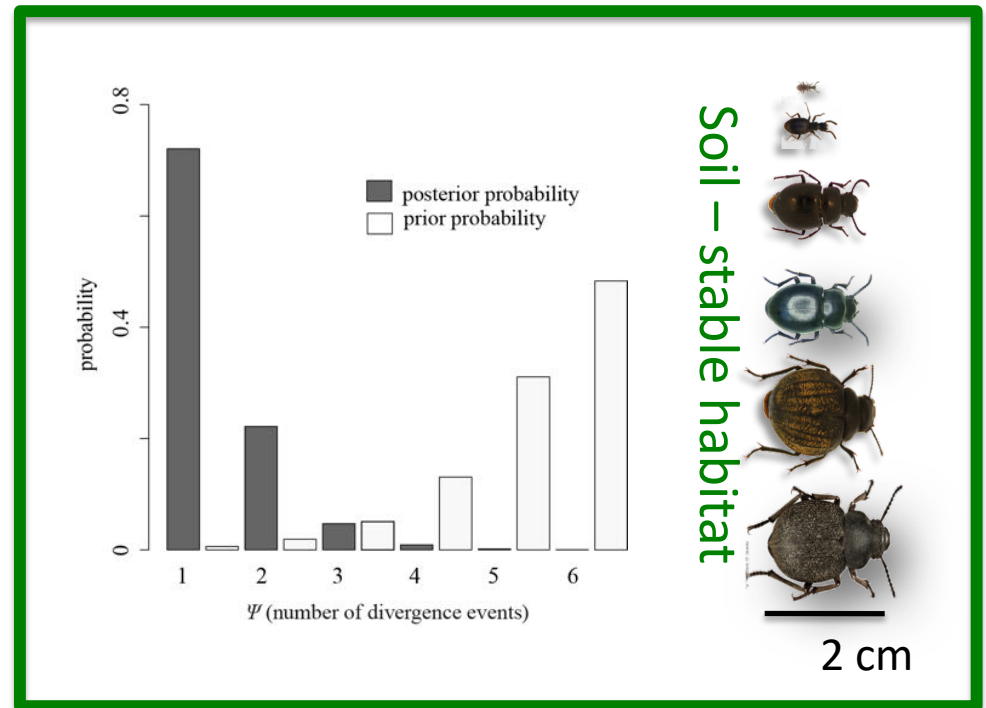
hABC: hierarchical Approximate Bayesian Computation;
Implemented in dpp-msbayes (Oaks, 2014)

By focusing on ecologically equivalent taxa, test of concordance
supported the species pump model of divergence

Papadopoulou & Knowles (2015) *Mol. Ecol.* 24: 4252-4268

Generic hypotheses of global phylogeographic concordance

No evidence for simultaneous divergence

hABC implemented in
dpp-msbayes (Oaks, 2014)

Papadopoulou & Knowles (2015) *Mol. Ecol*. 24: 4252-4268

Generic hypotheses of global phylogeographic concordance

No evidence for simultaneous divergence

Soill

Sand

2 cm

probability

Ψ (number of divergence events)

posterior probability
prior probability

0.35

1 2 3 4 5 6 7 8 9 10 11 12 13

Ephemerality of sand habitats!
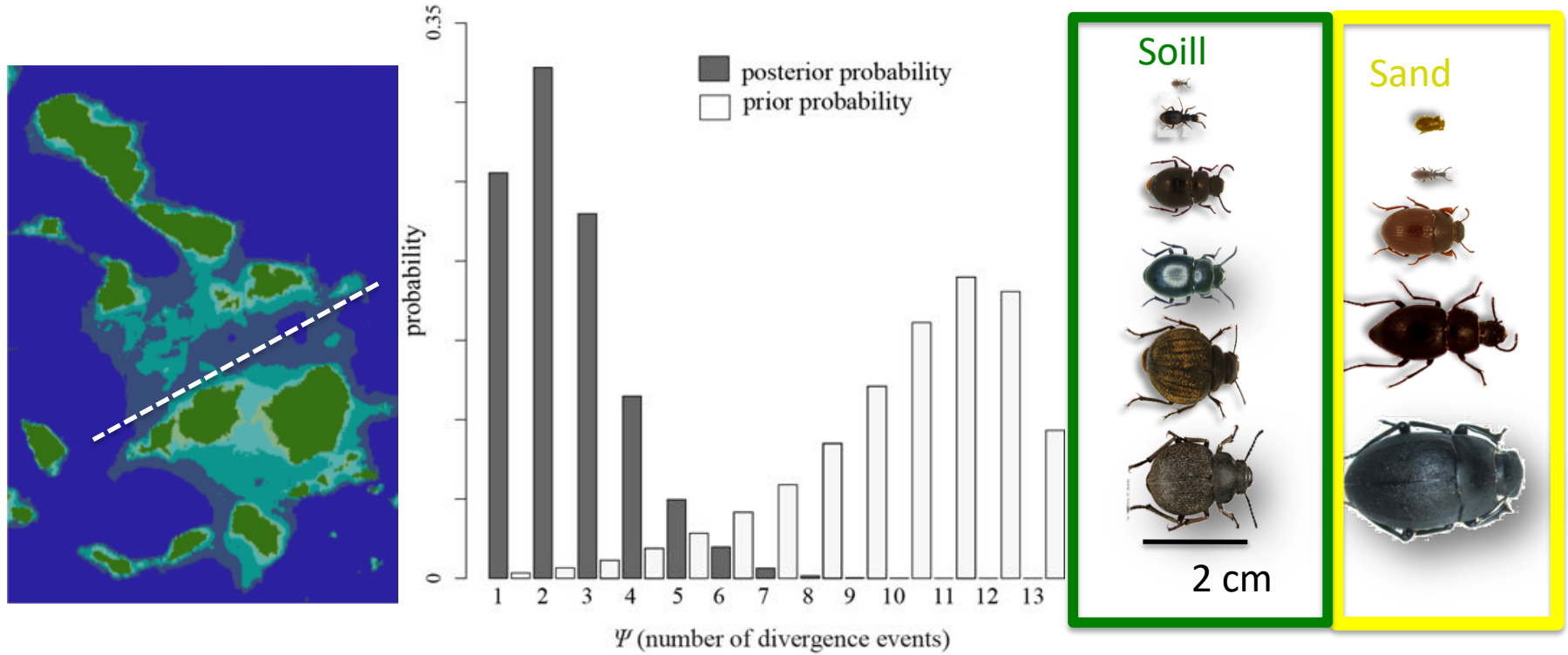
Sandy habitats
"Soil" habitats
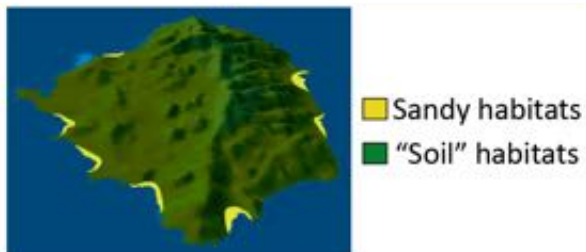
Lack of global concordance → rejection of species pump model of divergence ???

Papadopoulou & Knowles (2015) *Mol. Ecol.* 24: 4252-4268

Refined hypotheses based on taxon-specific traits in comparative phylogeography

Soil– stable habitat

Aegean archipelago

Time (kya)

Sea level (m)
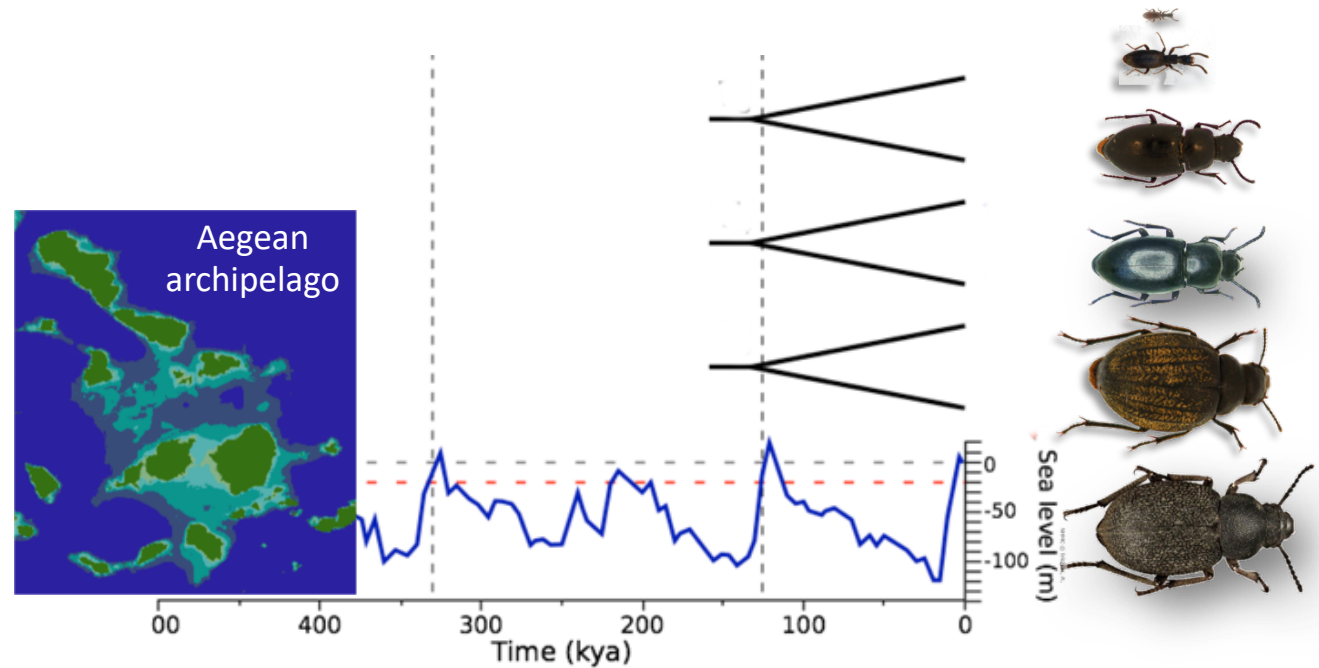
- refinement of the expectation for concordance is needed for concordance itself to be a meaningful metric

- reduced predictive power of generic hypotheses – their rejection leads to inconclusive statements that do not offer particularly meaningful insights

Papadopoulou & Knowles (2016) *PNAS*

- comparative phylogeographic methods are designed to quantify congruence, rather than gain insights from discordant patterns

- indirectly encourages users to emphasize idiosyncratic aspects of history!

dpp-msbayes

posterior probability

Number of divergence events
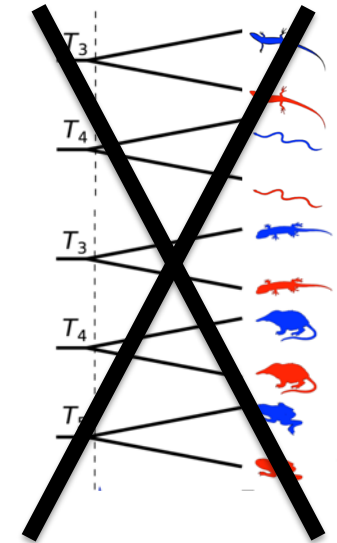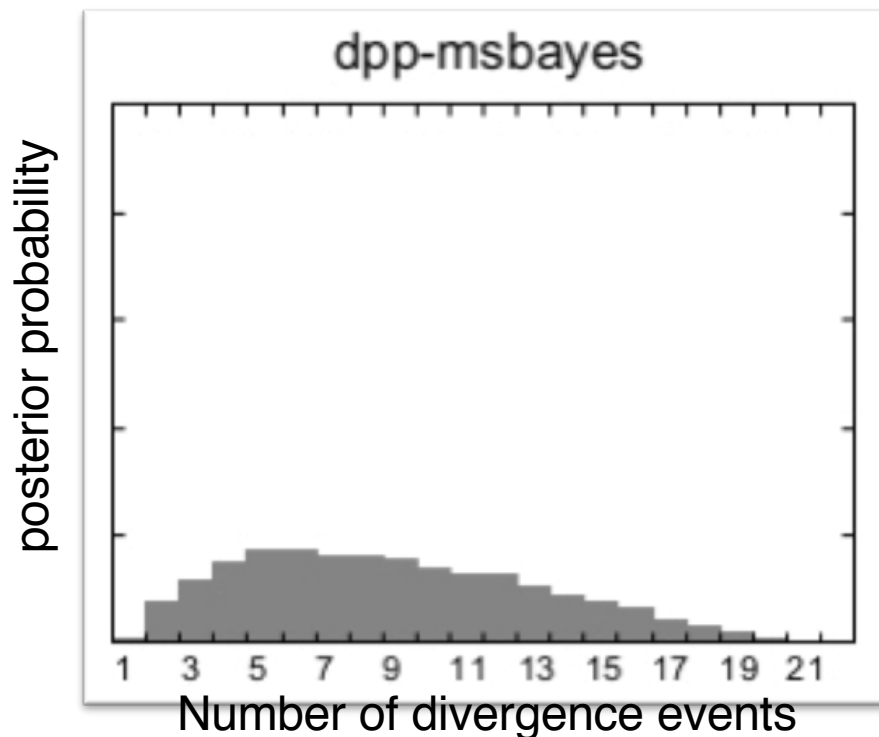
- ad hoc interpretations of discordance

- NEED development/application of methods for statistical evaluation of phylogeographic discord as an expectation of deterministic processes

Papadopoulou & Knowles (2016) *PNAS*

- Model formulation is a way of communicating our expert knowledge to statistical apparatus to test hypotheses

How do we know if we used the "right" model?

In practice we can never completely model the evolutionary process, all we can hope for is that we have captured the important features.

(i.e., IT IS KEY that you are knowledgeable about your system!)

"The purpose of models is not to fit the data but to sharpen the questions."

- *Samuel Karlin*