Workshop on Molecular Evolution

Marine Biological Laboratory

THE UNIVERSITY OF CHICAGO

Capping it all off!

2026 Course TAs

Joshua Justison (lead TA), University of Wisconsin

Analisa Milkey, University of Connecticut

Solomon McShea, University of California, San Francisco

Thao Nguyen, Iowa State University

2026 Course Faculty

Peter Beerli, Florida State University

Joseph Bielawski, Dalhousie University

Jeremy Brown, Louisiana State University

Belinda Chang, University of Toronto

Scott Edwards, Harvard University

Laura Eme, University Rhode Island

Mandev Gill, University of Georgia

Tracy Heath, Iowa State University

Lacey Knowles, University of Michigan

Laura Kubatko, Ohio State University

Emily Jane McTavish, University of California-Merced

Corrie Moreau, Cornell University

Claudia Solís-Lemus, University of Wisconsin-Madison

Megan Smith, Mississippi State University

David Swofford, Duke University

Rosana Zenil-Ferguson, University of Kentucky

2026 Course Assistant

Anne Smith

Evolutionary applications of genomic data

L. Lacey Knowles
Dept. of Ecology and Evolutionary Biology
University of Michigan

Evolutionary applications of genomic data:

- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Inferring species boundaries (aka species delimitation)
- Demographic inference

- All models are flawed..., but they are important because models are how we communicate our knowledge to a statistical apparatus

Evolutionary applications of genomic data

what I'll emphasize:

- Decisions/choices we make about model formulation
- Recognizing the subjectivity of model formulation itself when making inferences
- Decisions when analyzing empirical data

Evolutionary applications of model-based analyses:

(i) Inferring species boundaries (aka species delimitation)

(ii) Phylogenetic inference (and beyond the species tree)

(iii) Biogeographic study

(iv) Phylogeography

(v) Adaptive evolution

(vi) Speciation

Evolutionary applications of model-based analyses:

(i) Inferring species boundaries (aka species delimitation)

(ii) Phylogenetic inference (and beyond the species tree)

(iii) Biogeographic study

(iv) Phylogeography

(v) Adaptive evolution

(vi) Speciation

Laura S. Kubatko and
L. Lacey Knowles
Species Tree Inference

Model-based approaches for phylogeographic inference

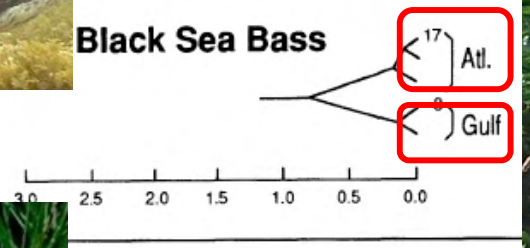
Discussion points:

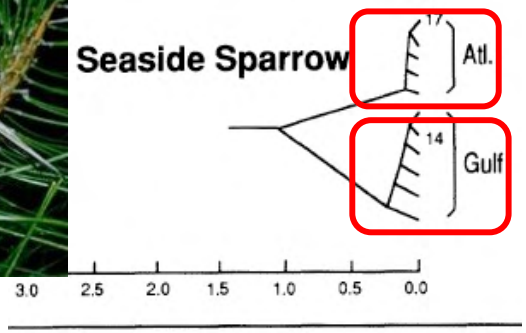
- Why models are important
- Generic versus informed models
- Species-specific expectations of genetic variation
(e.g.. trait-based hypotheses, spatially explicit coalescent models, etc.)
- Concordance versus discord among species in communities
(i.e.. lessons from comparative phylogeography)

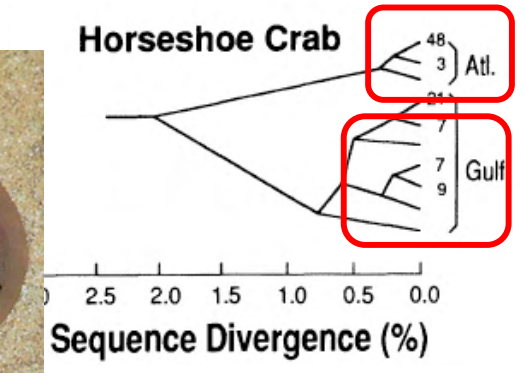
Why the transition from describing patterns of genetic variation to understanding process requires model-based approach

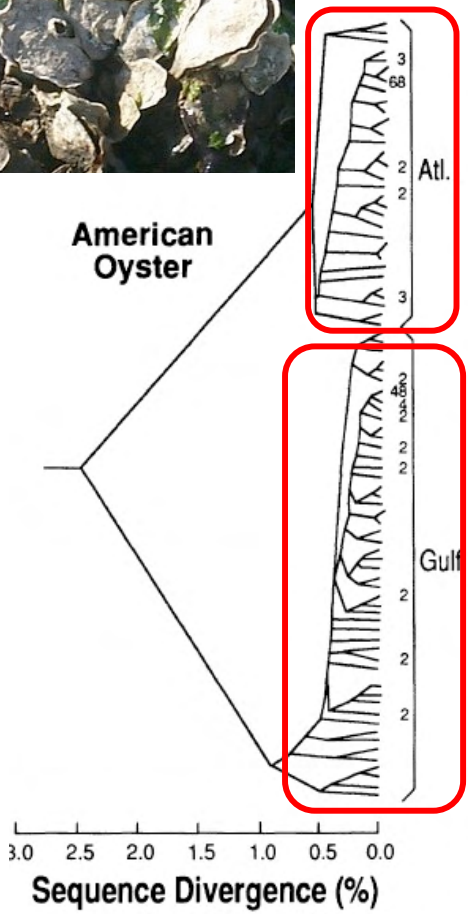
Classics in phylogeography

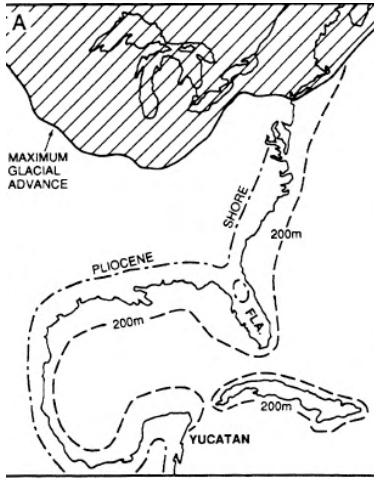
Concordance reflects a common vicariant history of population separation







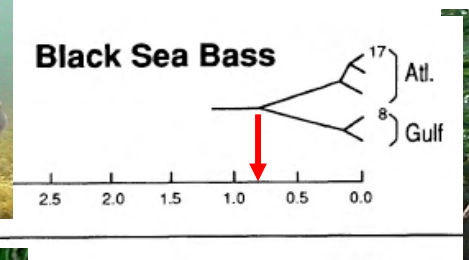


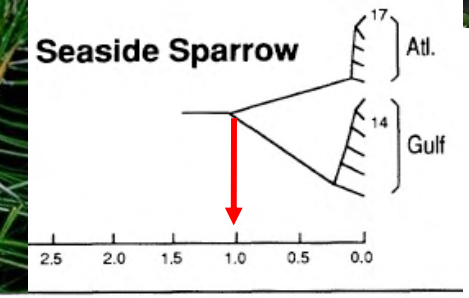


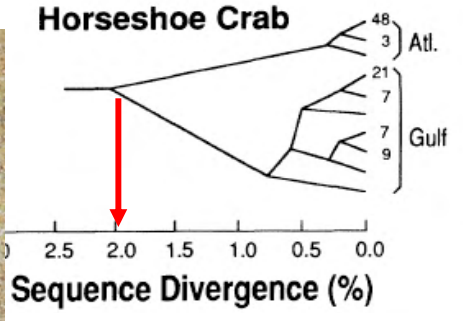
Avise 1992

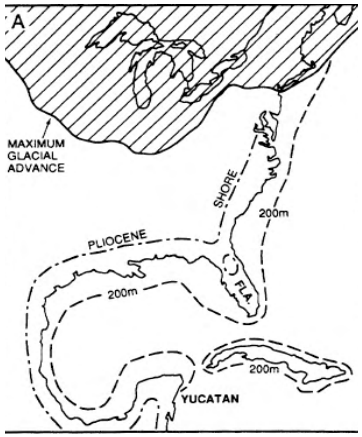
The data may be consistent with a shared response to a specific geologic event, despite differing gene tree depths among taxa? Or maybe not?

By looking only at the gene trees,
it isn't clear how the differences in gene tree depths should be interpreted!

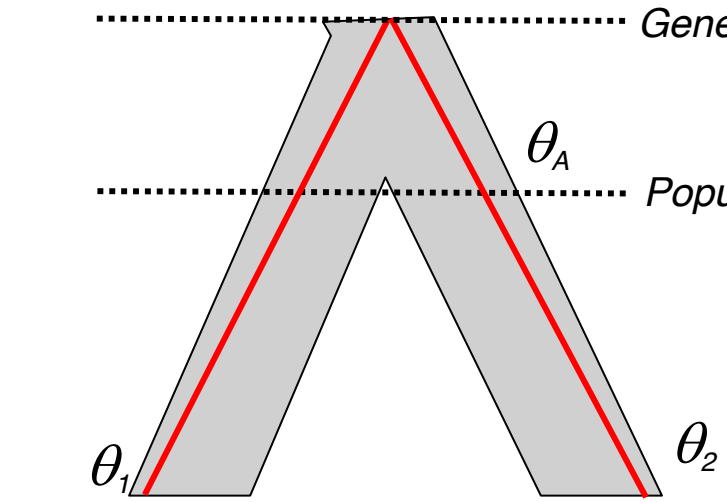




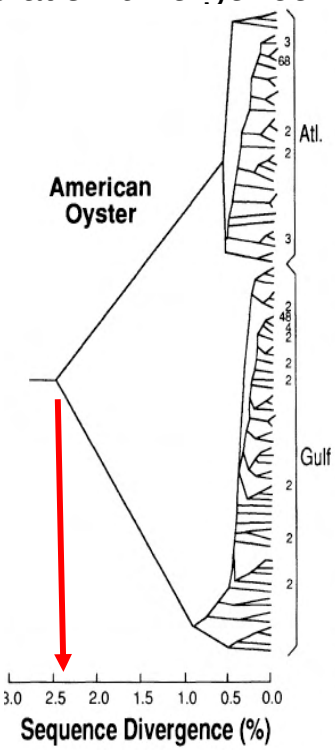


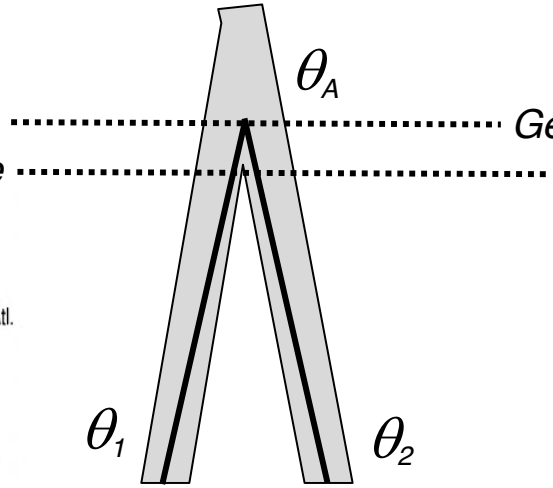


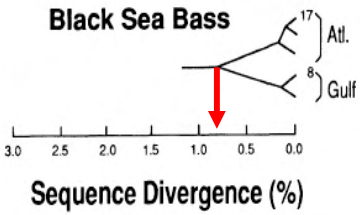
The difference between the timing of gene divergence and population divergence depends on the effective population size of descendant and ancestral populations



Present

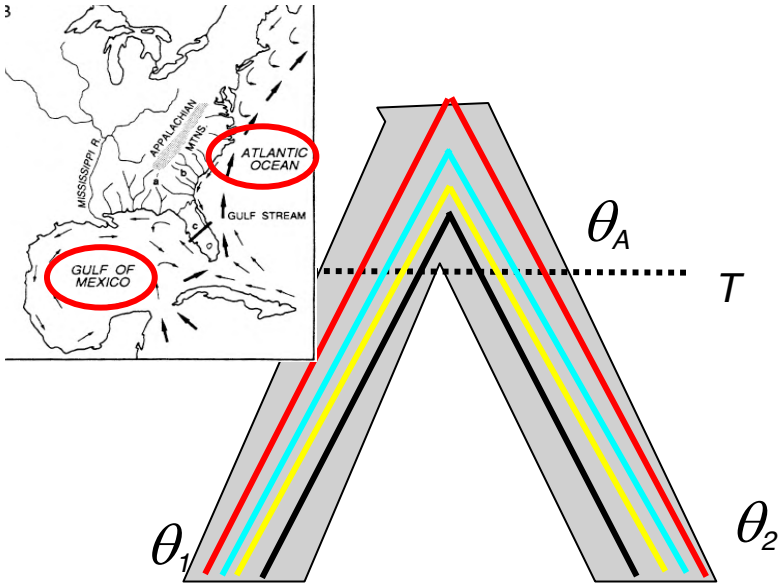






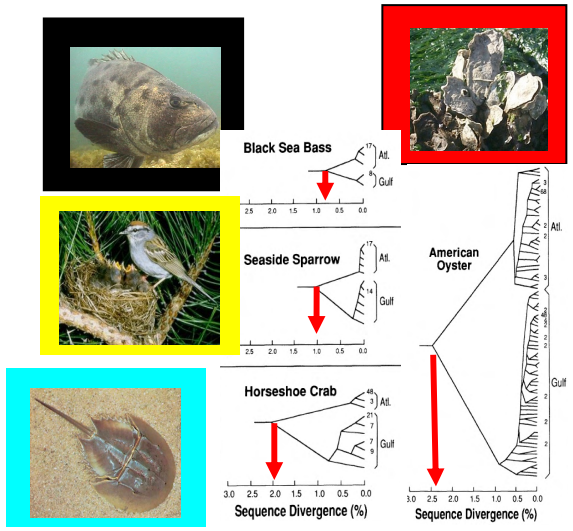
To test for shared vicariant history of the coastal community:

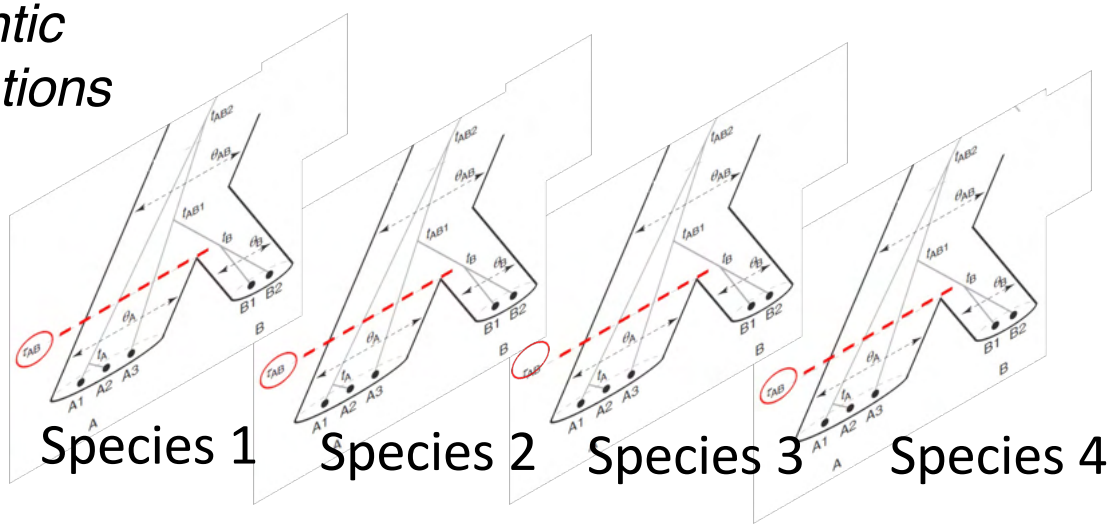
Need a model to assess statistically how much of a difference in the depths of the gene trees would still be consistent with the same timing of population divergence



Gulf populations

Atlantic populations





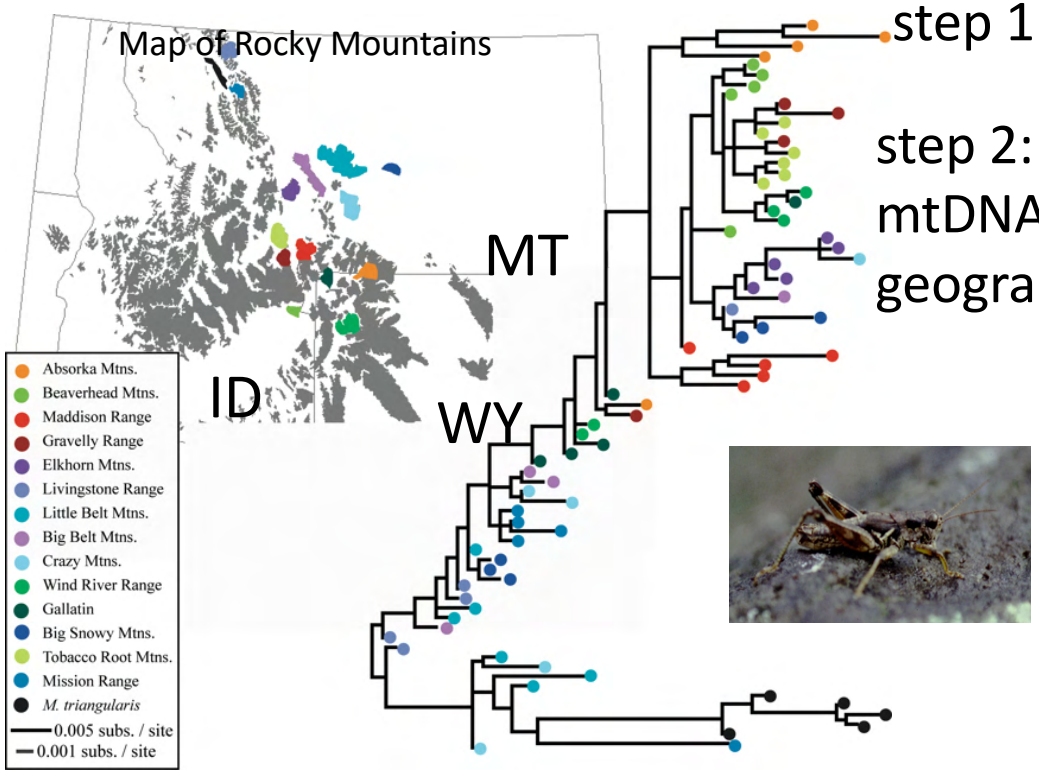
In the past, the central focus was on the ‘phylo’ component

PHYLOgeography

Use of gene trees predominated and genetic variation across populations described by:

step 1: reconstruct a gene tree

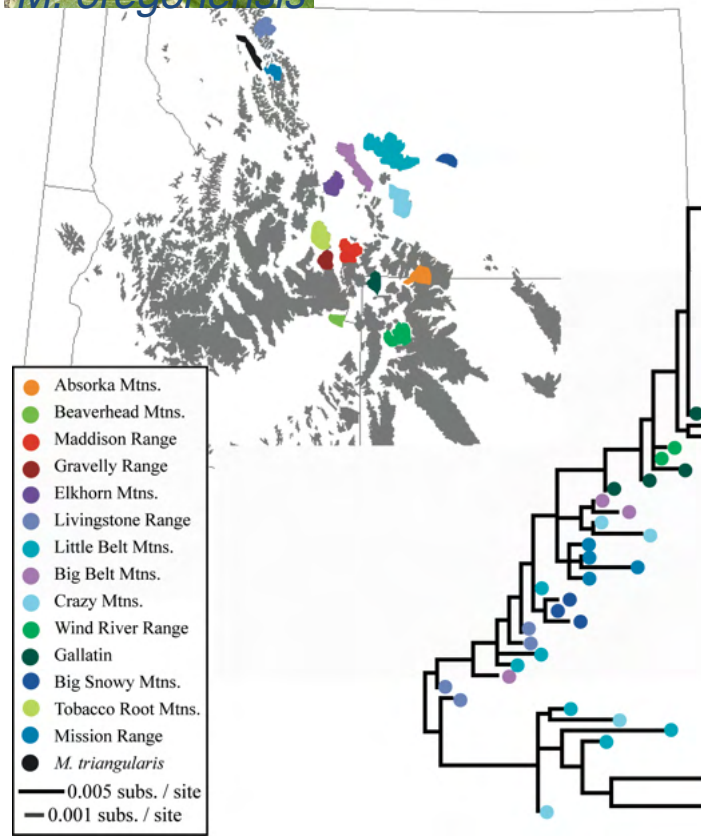
step 2: compare the relationships among mtDNA sequences/haplotypes to the geographic distribution of haplotypes

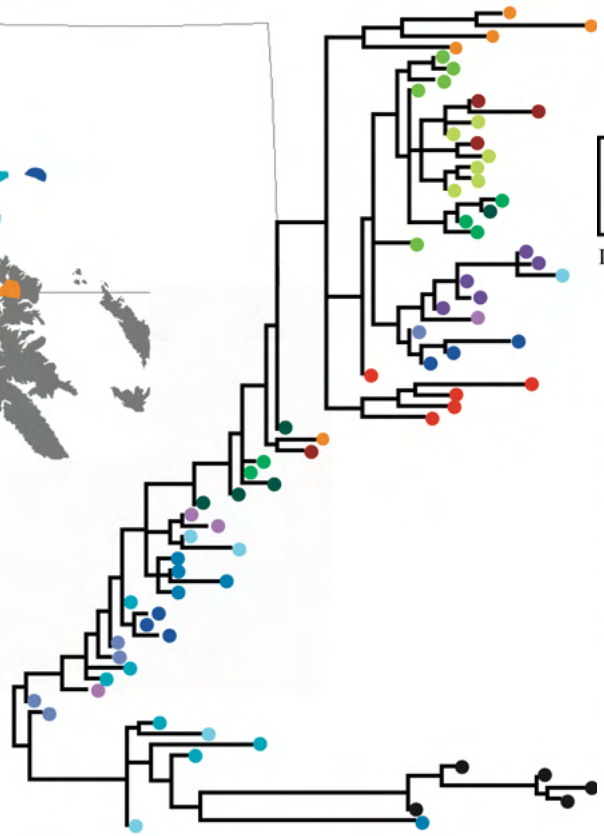


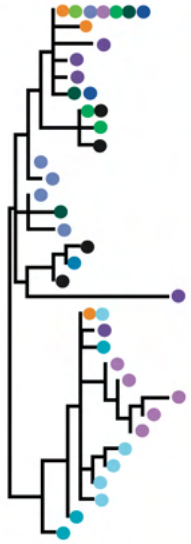
grasshopper haplotypes across populations (color coded by the mountain top where individual was collected)

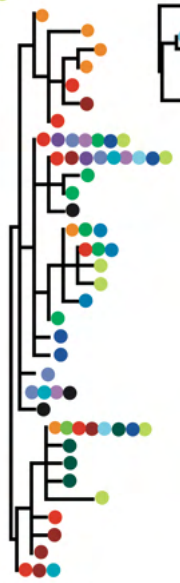
But different loci have different gene trees

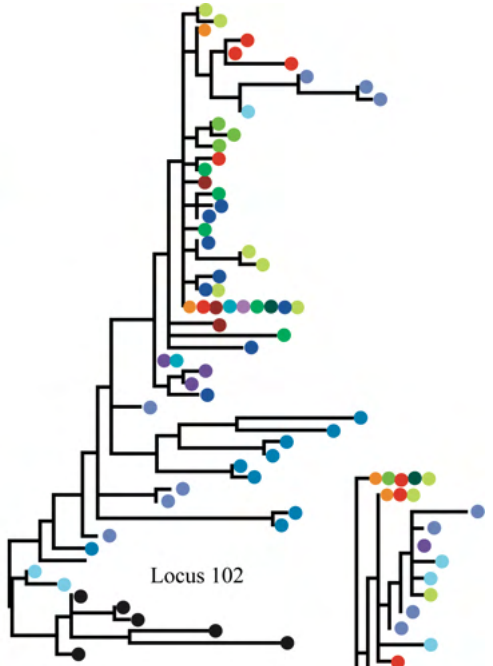
Phylogenetic relationships among populations?
(i.e., what's the underlying geographic history of divergence)?

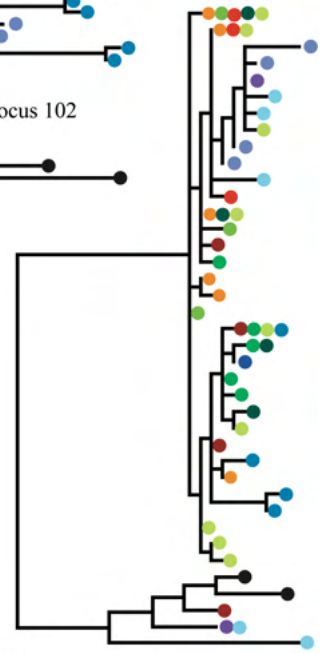








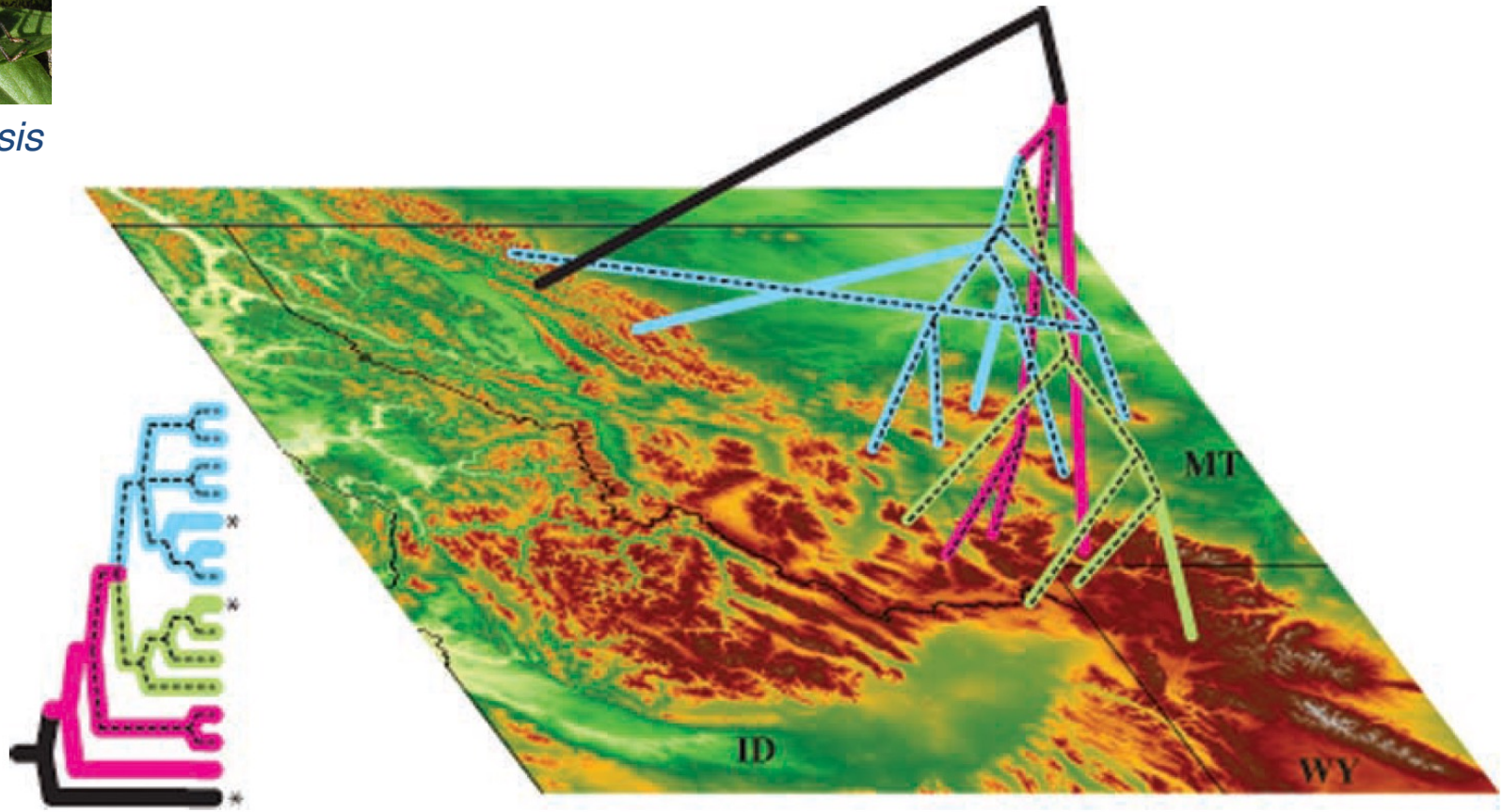




Knowles & Carstens (2007) *Evolution* 61:477

Infer a population tree using coalescent modeling containing information on the geography and timing of divergence

M. oregonensis



Knowles & Carstens (2007) *Evolution 61:477*

- Why models are important:

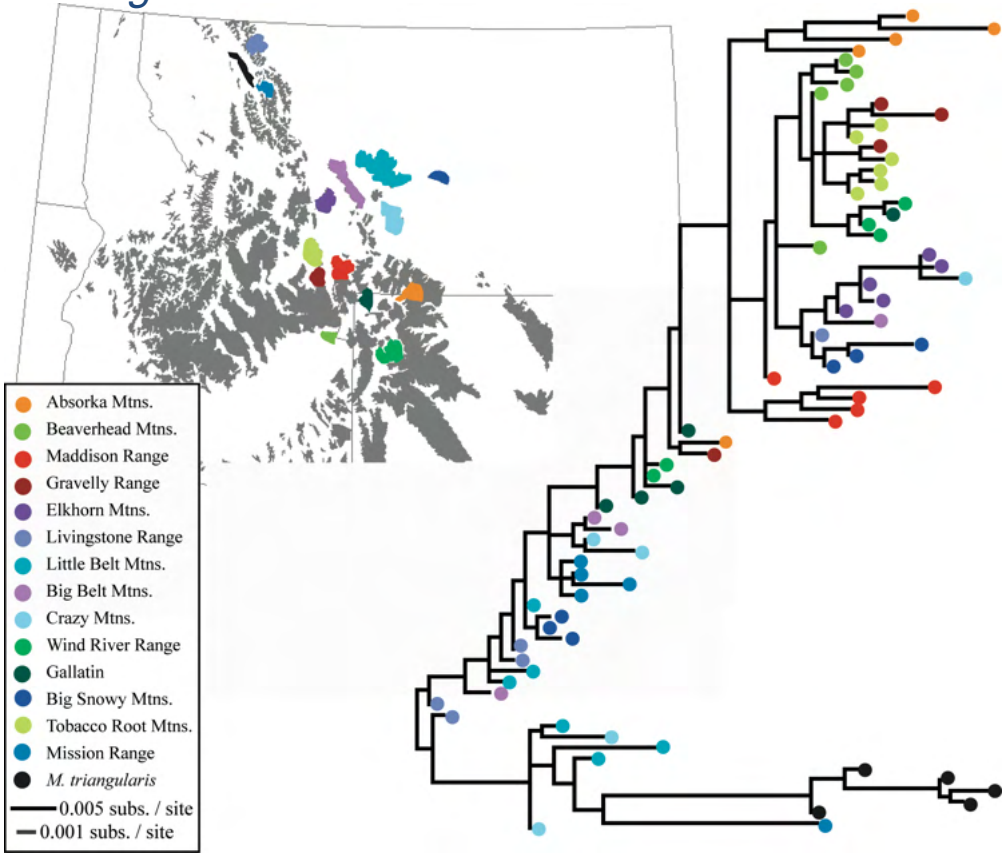
Different processes can produce similar genetic patterns

Cause for the lack of monophyly of populations?

M. oregonensis

Retention of ancestral polymorphism due to recent isolation?

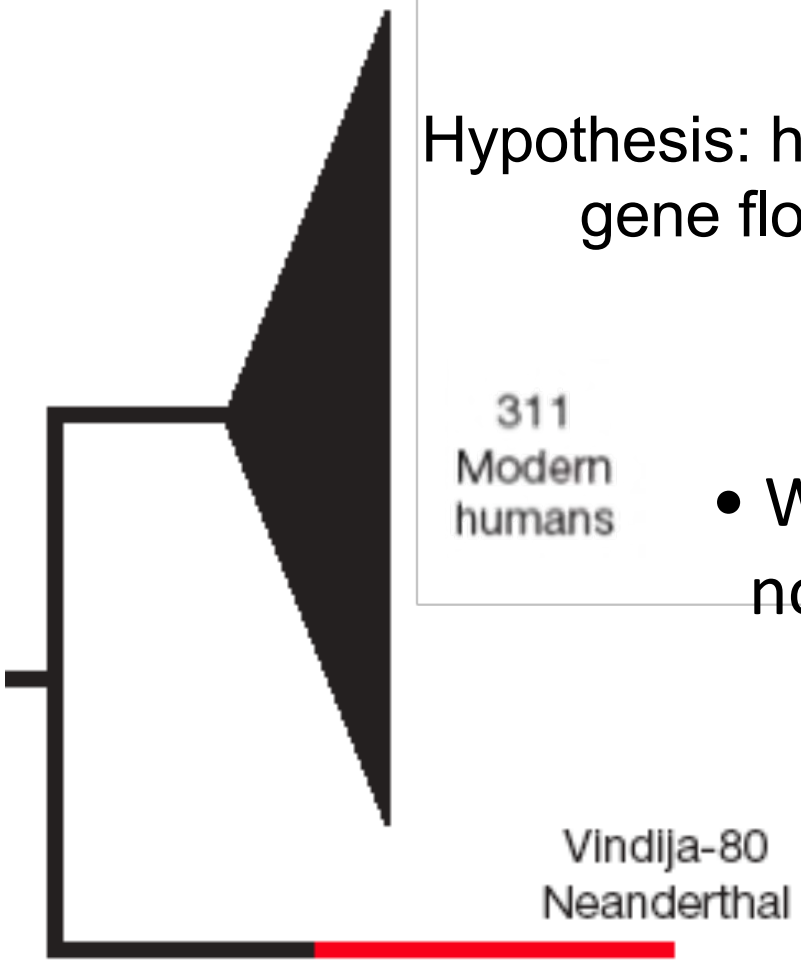
or migration?



Knowles & Carstens (2007) *Evolution 61:477*

$F_{ST} = 0.15$

Interbreeding between Neanderthals and humans?



Hypothesis: humans replaced Neanderthals without gene flow between the separate species

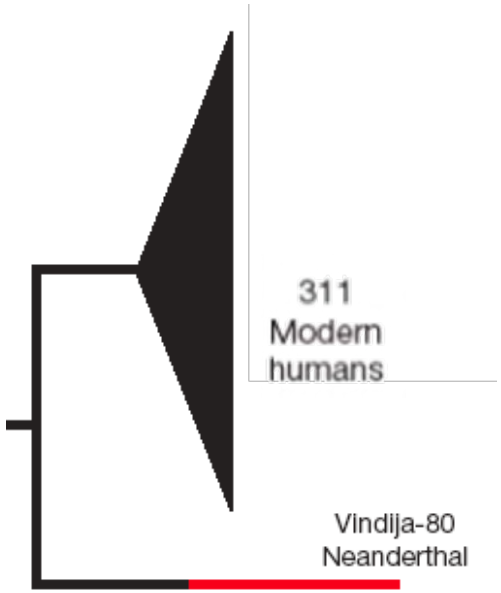
- What are two bits of evidence for no gene flow based on this tree?

Problem with interpreting gene tree as evidence of "divergence with no gene flow"

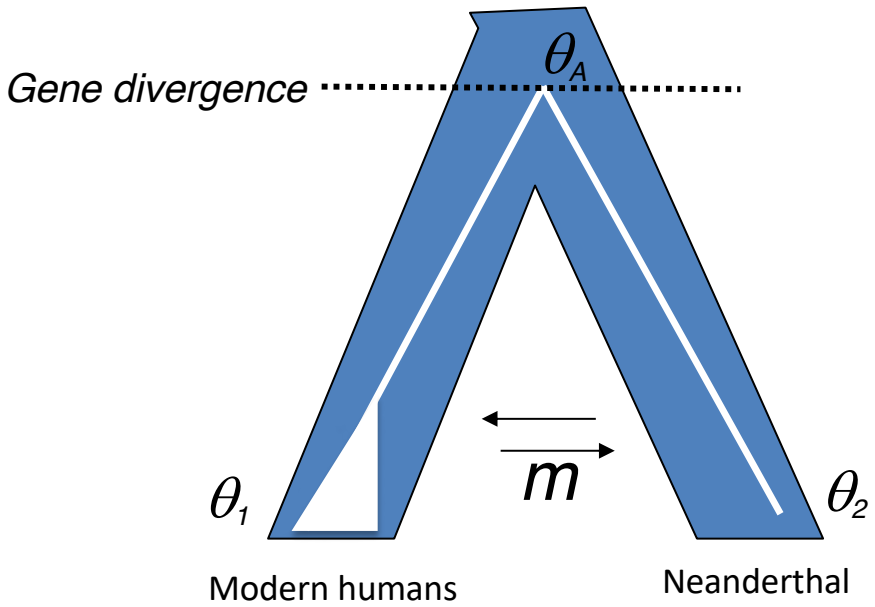
Nordberg

Interbreeding between Neanderthals and humans?

- Model-based test: is this gene tree compatible with ancient gene flow between humans and Neanderthal



(Bayesian program IM)



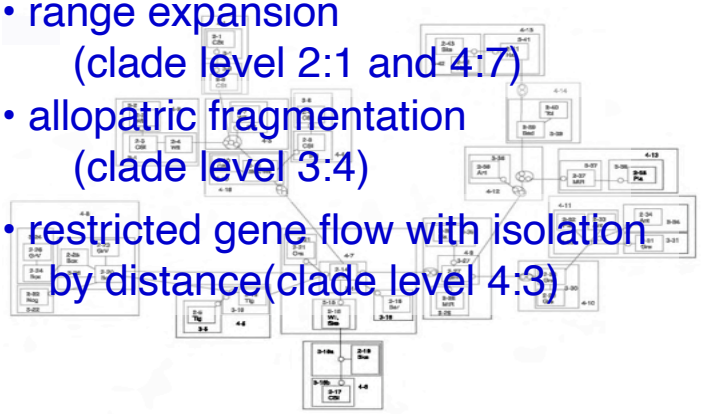
Result: yes, tree is compatible; does this mean there was gene flow?

- Not necessarily because with single gene there is not a lot of power to evaluate the hypothesis of divergence with gene flow

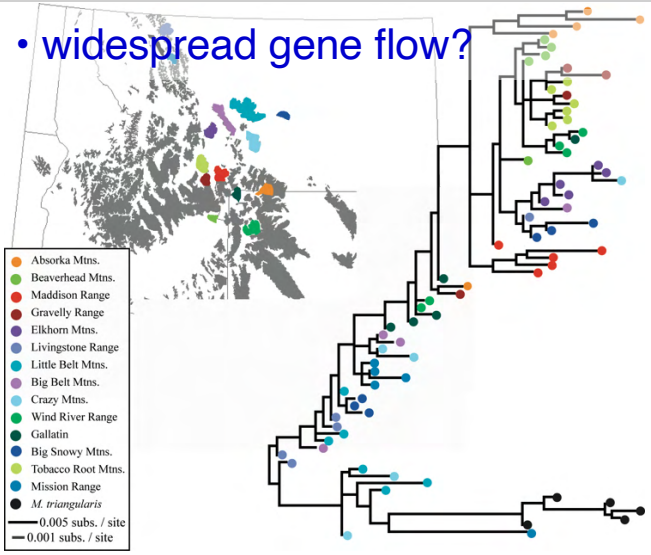
Nordberg

Equating a gene tree (or network) with a species' history is not appropriate for making inferences about evolutionary processes

- range expansion
  (clade level 2:1 and 4:7)
- allopatric fragmentation
  (clade level 3:4)
- restricted gene flow with isolation by distance(clade level 4:3)



- widespread gene flow?



Without a model:

- inferred processes may (*or may not*) be accurate because different processes can produce a similar pattern in genetic data and gene trees may differ across loci

- no measure of the uncertainty/support surrounding hypotheses or statistical framework for evaluating competing hypotheses

- no framework to incorporate additional data (e.g., geologic or ecological information)

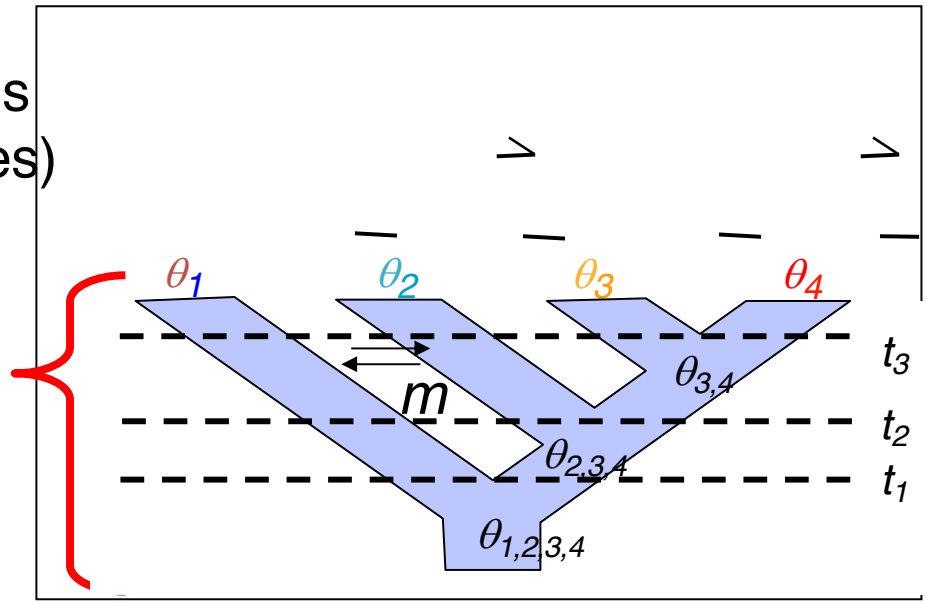
- inherent lack of power when individual loci analyzed separately, and discordance among loci is uninterpretable

Knowles 2009 Annu Rev Ecol Syst

With model-based approaches

1) accommodate and make full use of multilocus data (individual gene trees differ so trying to interpret their patterns would lead you to many different stories)

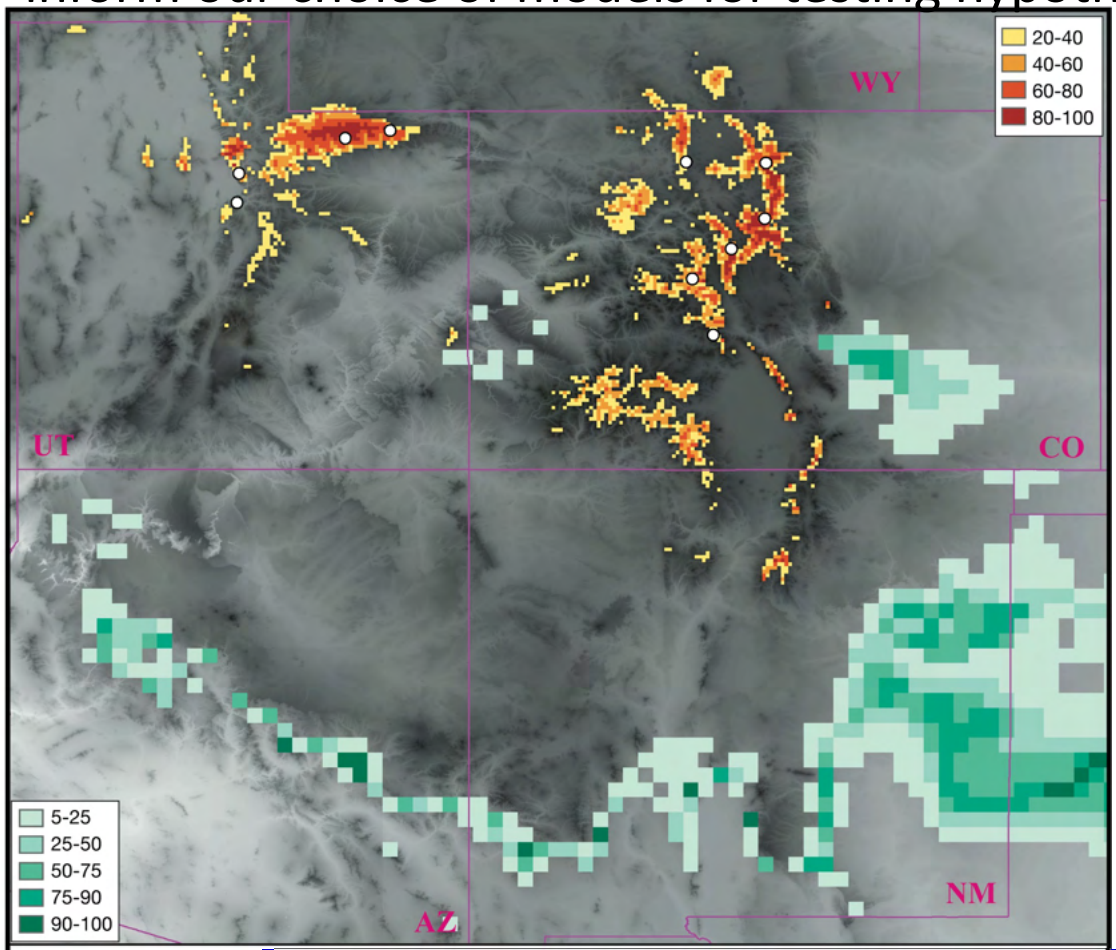
Explicit model of a species' history



2) estimate evolutionary parameters (e.g., population size, migration rates, divergence times, or demographic changes like expansions or bottlenecks, the geographic coordinates of the ancestral population)

3) test alternative hypotheses/models (e.g., distinguish between a hierarchical vicariant divergence model versus a stepping-stone colonization model, or isolation by distance)

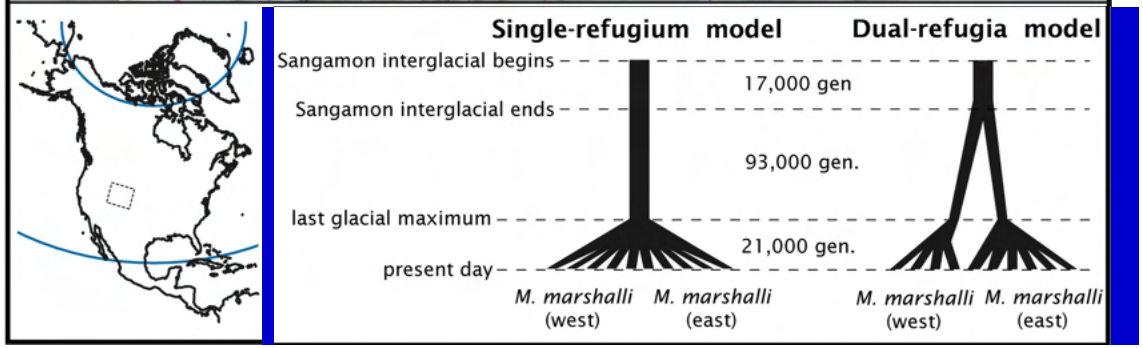
4) Incorporate additional non-genetic sources information to inform our choice of models for testing hypotheses



Coupled genetic and ecological-niche models to test hypotheses about ancestral refuges

- Projections of current distribution
- Projections of past distribution 21,000 years ago

(based on 19 bioclimatic variables; analyzed with MAXENT)

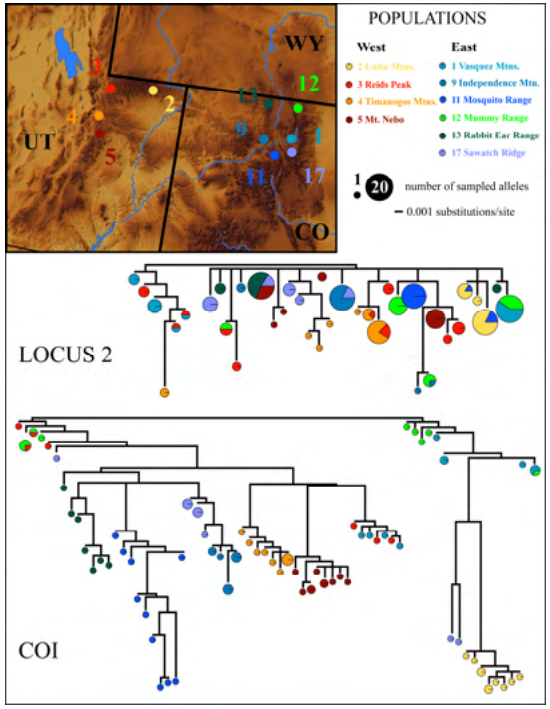


Allopatric ancestral glacial refugia populations promoted speciation

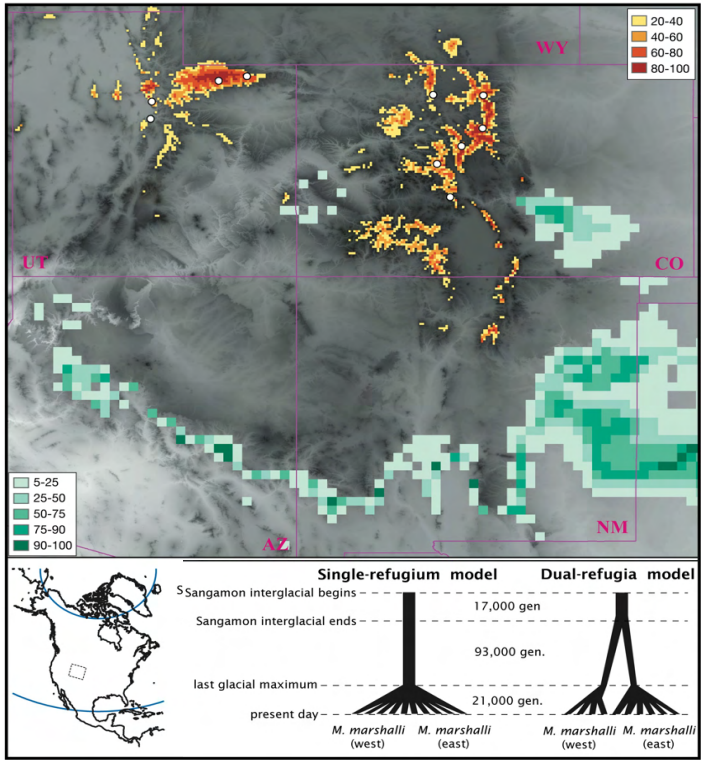
Knowles et al. 2007 Current Biology 17:1-7.

Coupled genetic and ecological-niche model:

With sequence data from multiple loci, we could reject the fragmentation of a single refugial population hypothesis, suggesting divergence among multiple refugia promoted species divergence.



- simulate gene trees and evaluate the degree of discord between gene tree and population trees to generate an expected distribution for the degree of discordance if the data had evolved under a model of founding from a single ancestral population



Knowles et al. 2007 Current Biology 17:1-7.

Fascinating and spectacular diversity of the genus *Melanoplus*!

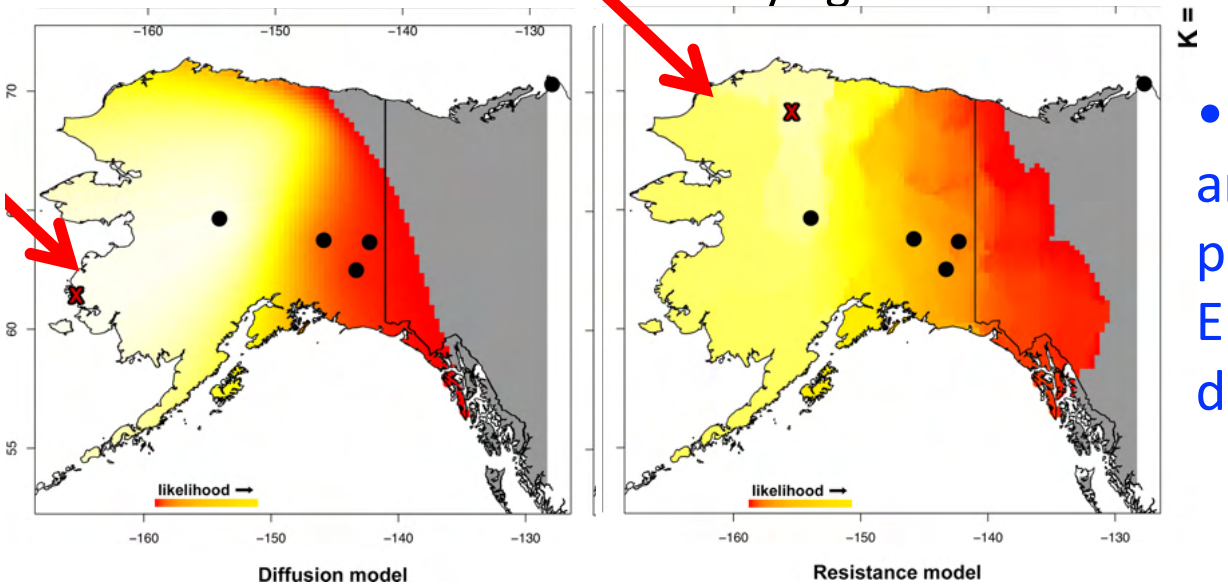
and other Melanoplinae genera

There are many different reasons why it is desirable to combine genetic data with other types of information (e.g., geographic or distributional data, ecological information, species functional traits, etc.)

Why is it desirable to combine genetic data with other types of information?

Capture biologically reality!

- Euclidean distance
- resistance distance (based on underlying environmental setting)



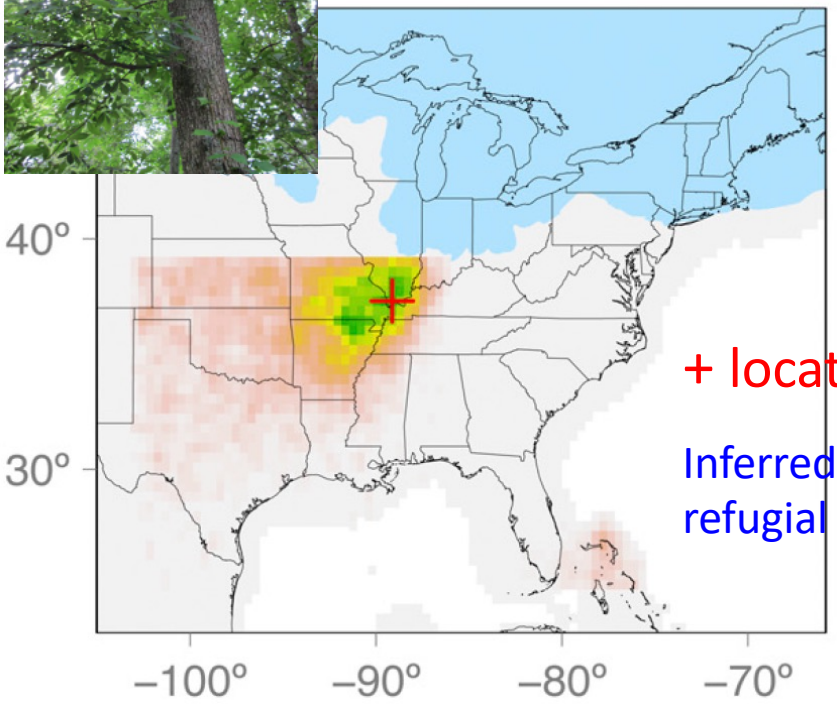
- Direction and location of ancestral source of expanding population differs between Euclidean and resistance distance (He et al. 2017)

Likelihood surface of location of source population during expansion (He et al. 2017) based on allele frequency gradients, represented by Ψ-statistics (Peter & Slatkin 2013)

He et al. 2017. Inferring the geographic origin of a range expansion: latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program X-ORIGIN. *Mol. Ecol.* 26:6908-6920. DOI: 10.1111/mec.14380

Use genetic data to corroborate inferences based on other data types

ENMs do not provide precise location of Pleistocene refuge for hickory trees



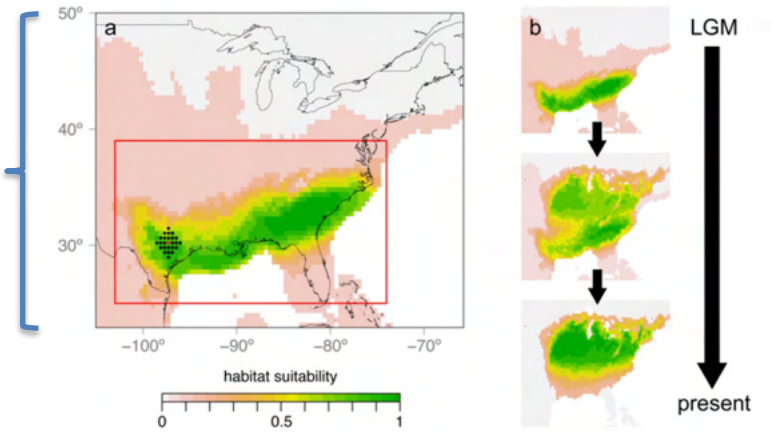


Fig. 1. Schematic overview of demographic simulations. (*A*) Simulations were initiated in the LGM landscape (shown here for *C. cordiformis*) from a central deme (see red dot as an example) plus an area extending three additional demes (black dots) in all directions. Different geographic sources of

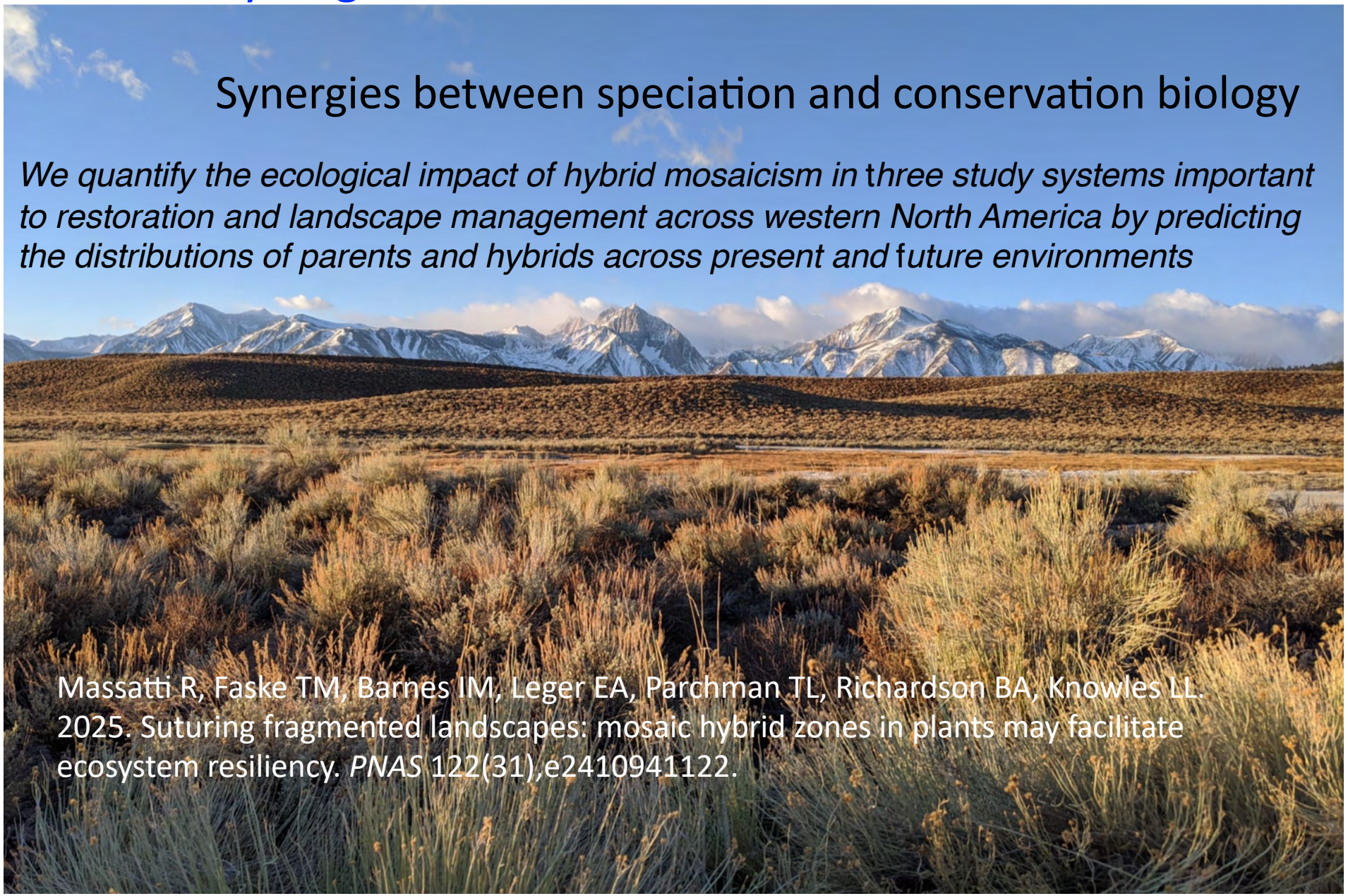
+ location of a macrofossil of the hickory species!

Inferred likelihood of geographic coordinates of ancestral refugial population overlap with the macrofossil

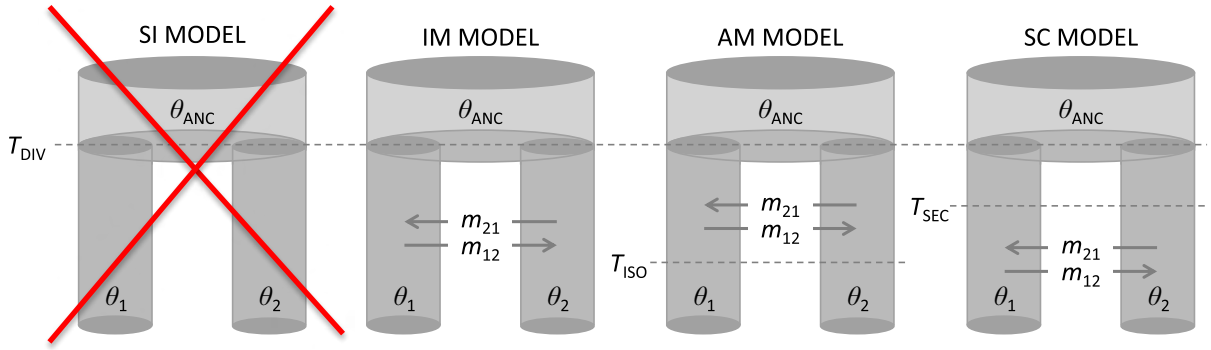
Fig. 2. Estimated expansion origins (Ω; red cross) in *C. cordiformis* (*A*) and *C. ovata* (*B*). The shading of pixels depicts a probability surface (kernel density) showing the likelihood that each pixel served as the expansion origin relative to the pixel with the highest likelihood (i.e., Ω). Glaciated regions are shown in blue. The results presented in *A* and *B* are based on retention of four and three PC axes of variation in genetic summary statistics, respectively. Results based on retaining additional PC axes are presented in *SI Appendix*, Figs. S2 and S3.

Bemmels JB, Knowles LL, Dick CW (2019) Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. *PNAS* 116:8431-8436.

By combining genetic data with other types of information we can build synergies between fields

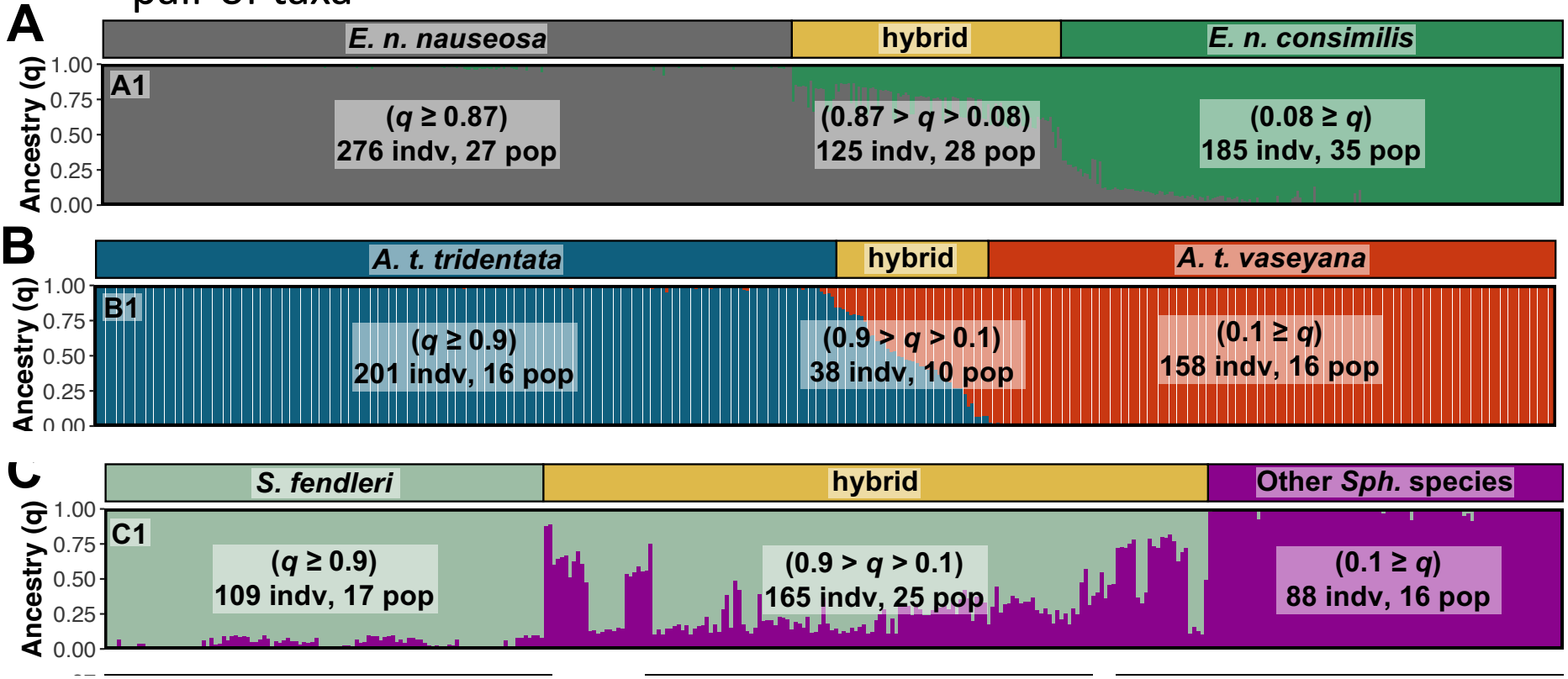


• Divergence with gene flow heavily favored over strict isolation among pairs of diverging taxa in three different foundational shrubs

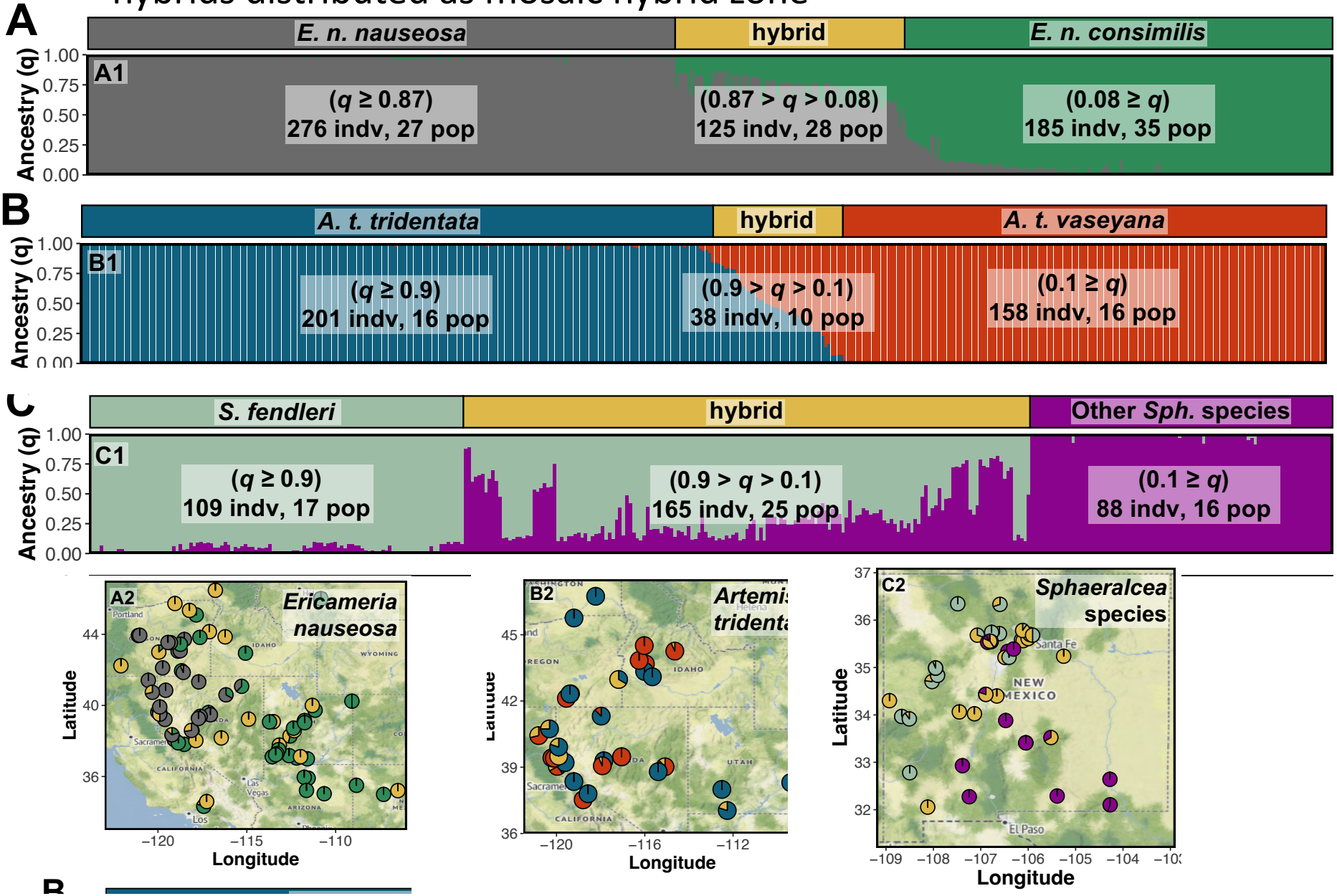


Analysis of SNPs from RADseq data using FastSimCoal

- despite gene flow, parental species remain genetically distinct in each pair of taxa

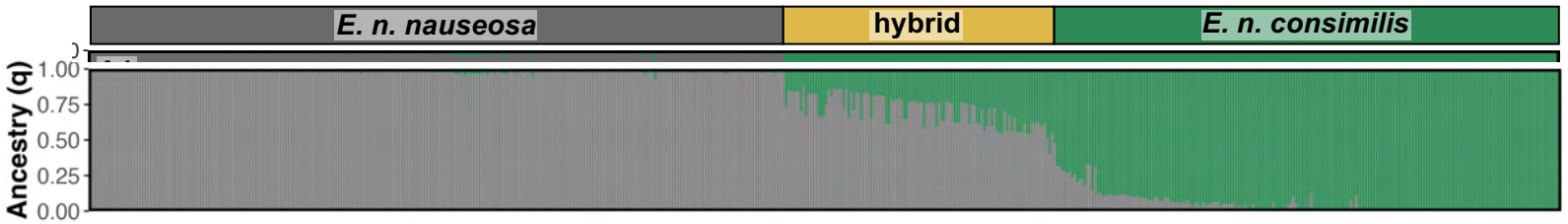


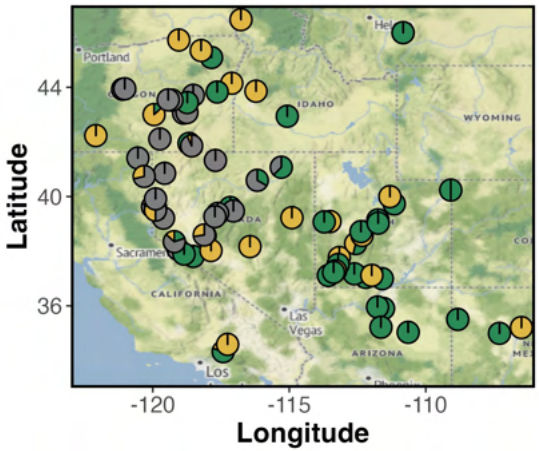
- despite gene flow, parental species remain genetically distinct
- hybrids distributed as mosaic hybrid zone

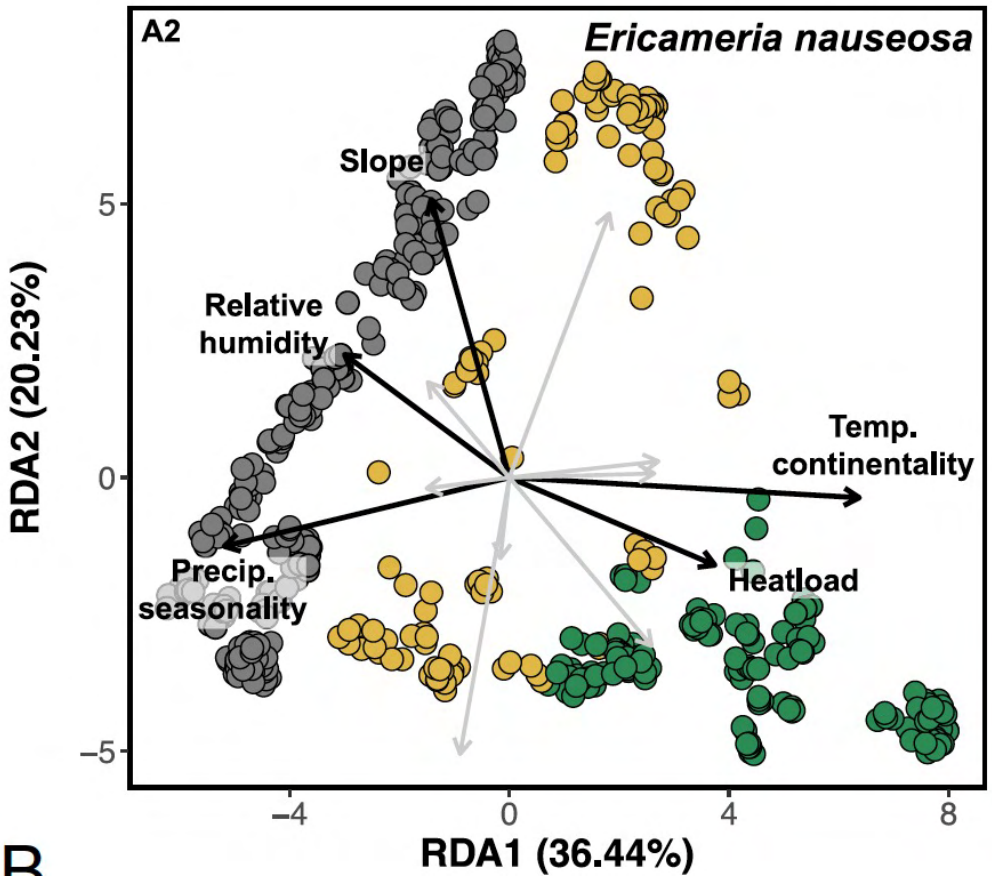


- Tests of association between genotypes and environmental variables show that hybrids occupy intermediate environmental space

Rubber rabbitbrush (*Ericameria nauseosa*)



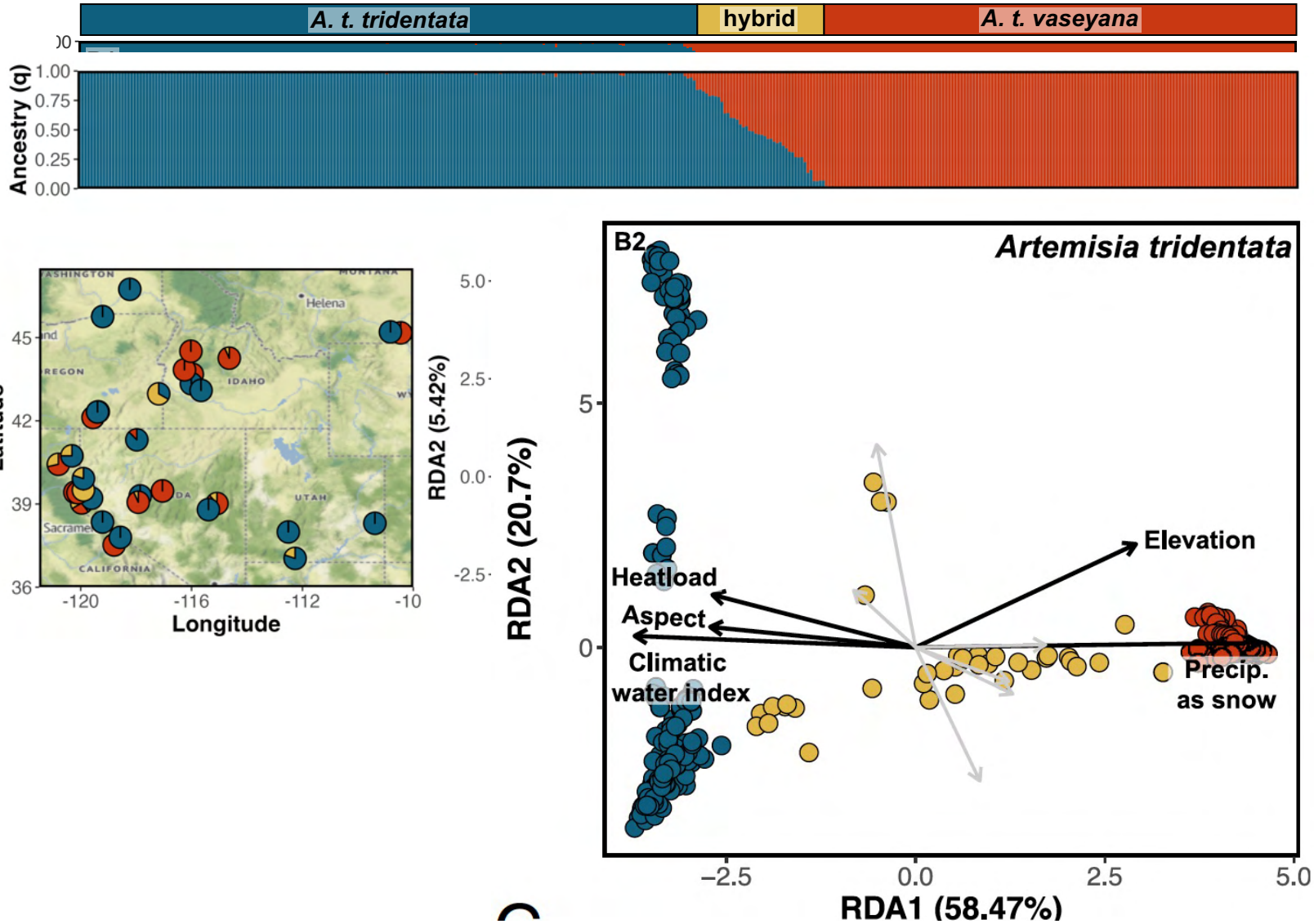




Redundancy analysis (RDA) illustrates how environmental variation predicts genetic variation between the lineages and hybrids. Black lines and labels denote the top five environmental variables predicting ancestry in a multivariate framework.

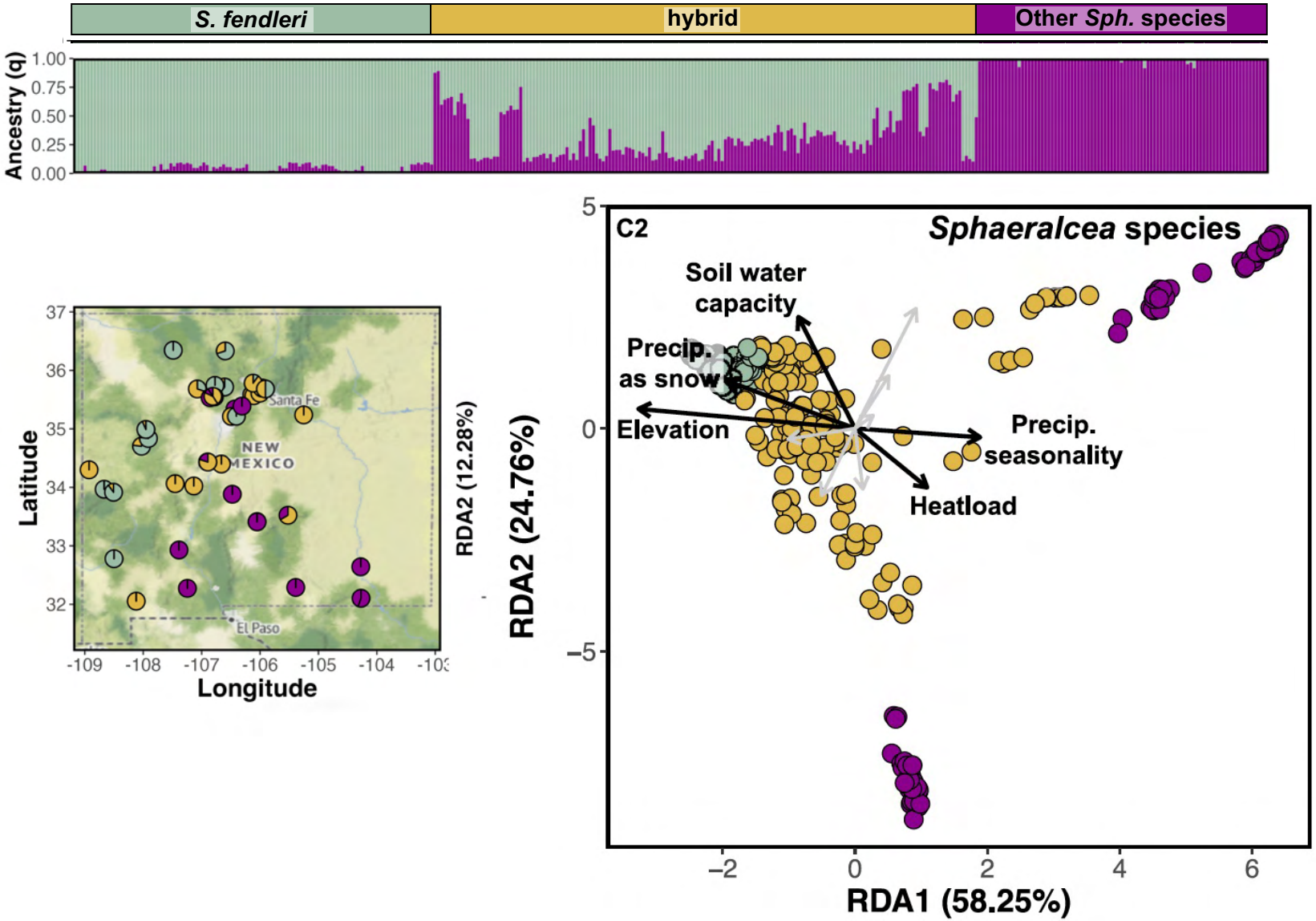
- Tests of association between genotypes and environmental variables show that hybrids occupy intermediate environmental space

Big sagebrush (*Artemisia tridentata*)



- Tests of association between genotypes and environmental variables show that hybrids occupy intermediate environmental space

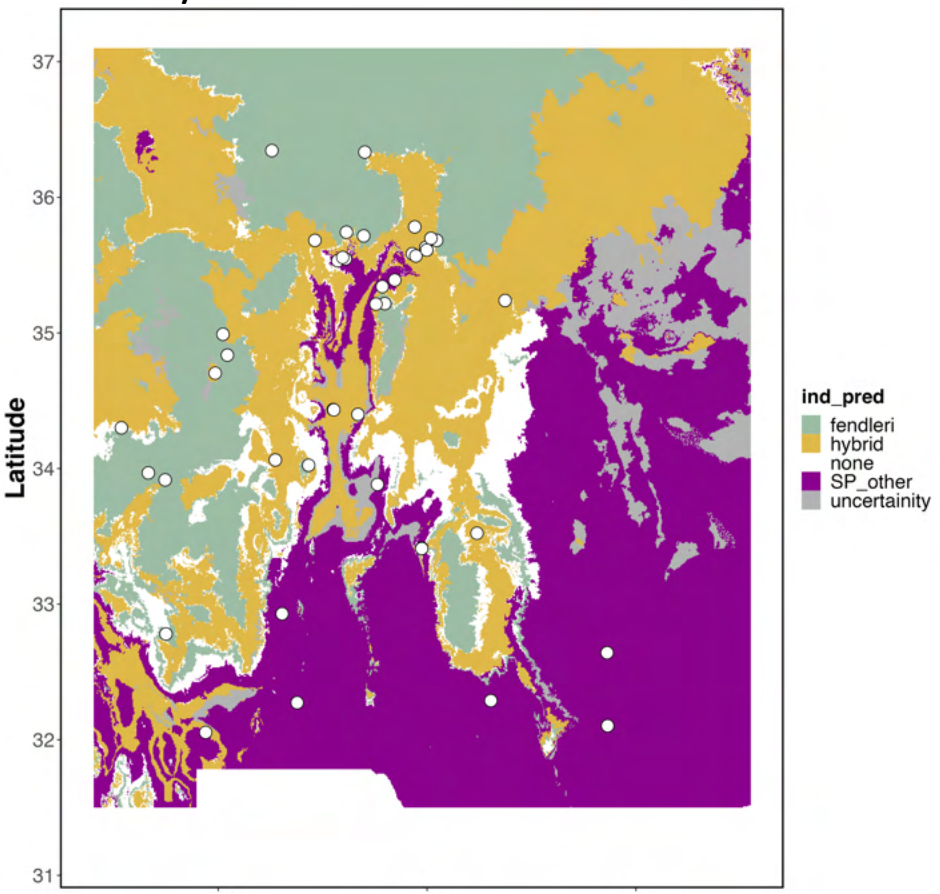
Globemallows (*Sphaeralcea*)

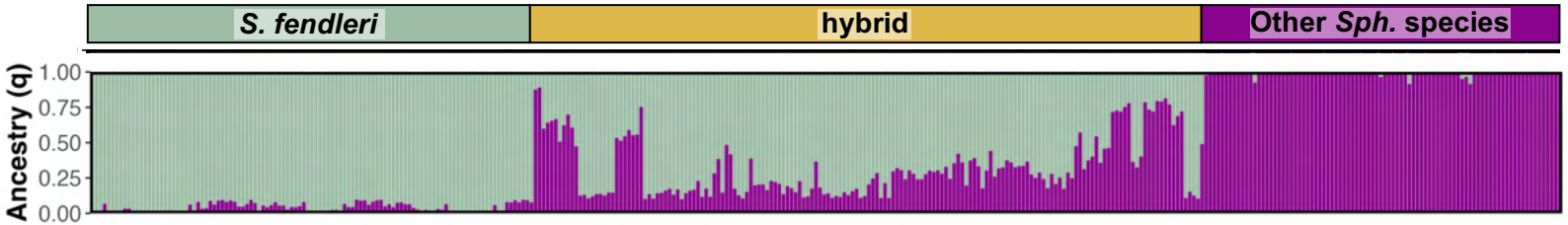


Hybrids act as sutures across fragmented landscapes:
speciation history informing restoration efforts

- species occupy more geographic area by virtue of the hybrids occupying unique environmental space
- underappreciated ecological role, especially for species that play a foundational role in ecosystems (sagebrush, rabbitbrush)
- many ramifications for restoration in current and future climates

Projected occurrence of parental taxa and hybrids from Random Forest model





Why is it desirable to combine genetic data with other types of information?

Cool questions! Biological insights!!

Does microhabitat affect responses to climate change

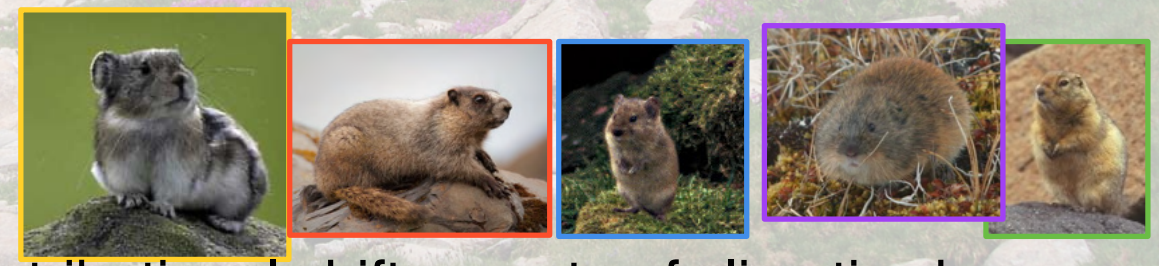
Massatti & Knowles (2014, 2016) *Evolution,* Mol. Ecol.

Role of habitat stability in structuring genetic variation

He et al (2013) *Evolution*

Present versus past distributions as drivers of species 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.*

Statistical inference in phylogeography:

Need to define a model

to see how variation in the parameters (e.g., mutation rates, migration rates, selection coefficients, number of population lineages) lead to specific patterns of genetic variation (e.g., variation among DNA sequences, among SNPs, partitioning of variation across populations, structuring of genetic variation across space and time, and among species)

How do we decide upon a model*:

- informed from information independent of the genetic data itself
  - that is, a specific biological narrative motivates the model

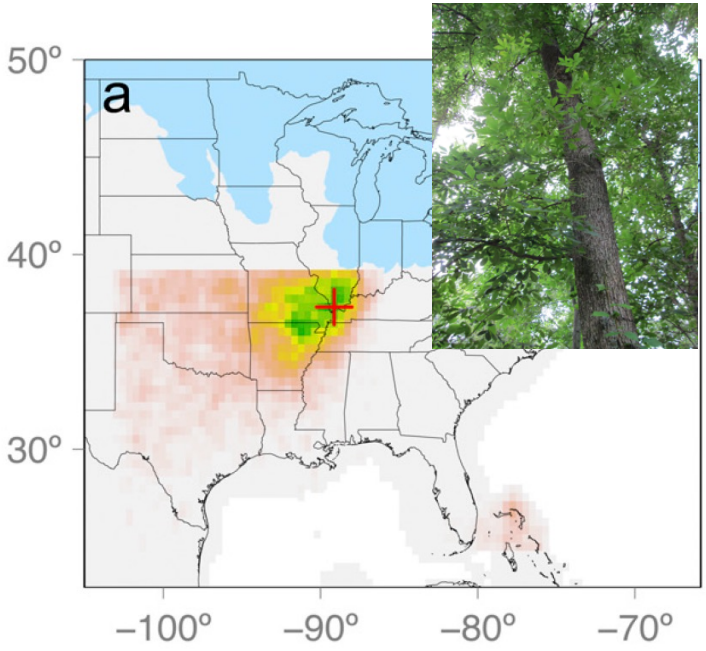
- models informed by genetic data

- generic models

* All models are simplifications, and vary in their relative degree of abstraction

How do we decide upon a model:

Expansion model used because of known displacement of hickory trees from current distribution by glacial ice sheet.



Inferred geographic coordinates of ancestral source population of expansion, where the geographic coordinate is a parameter in the model (see *He et al.* 2017. Inferring the geographic origin of a range expansion: latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program X-ORIGIN. *Mol. Ecol.* 26:6908-6920. DOI: 10.1111/mec.14380

Bemmels et al. 2019 PNAS 116:8431-8436

Geographic position of ancestral source populations differ among species.

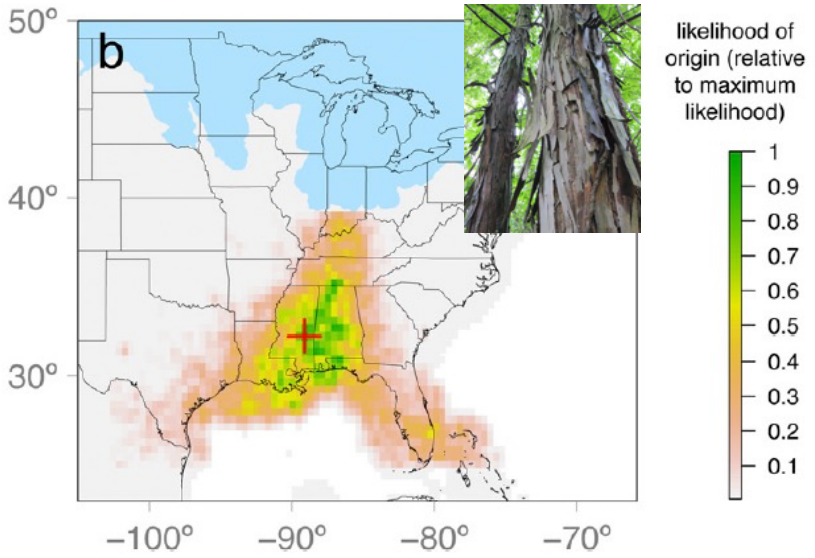
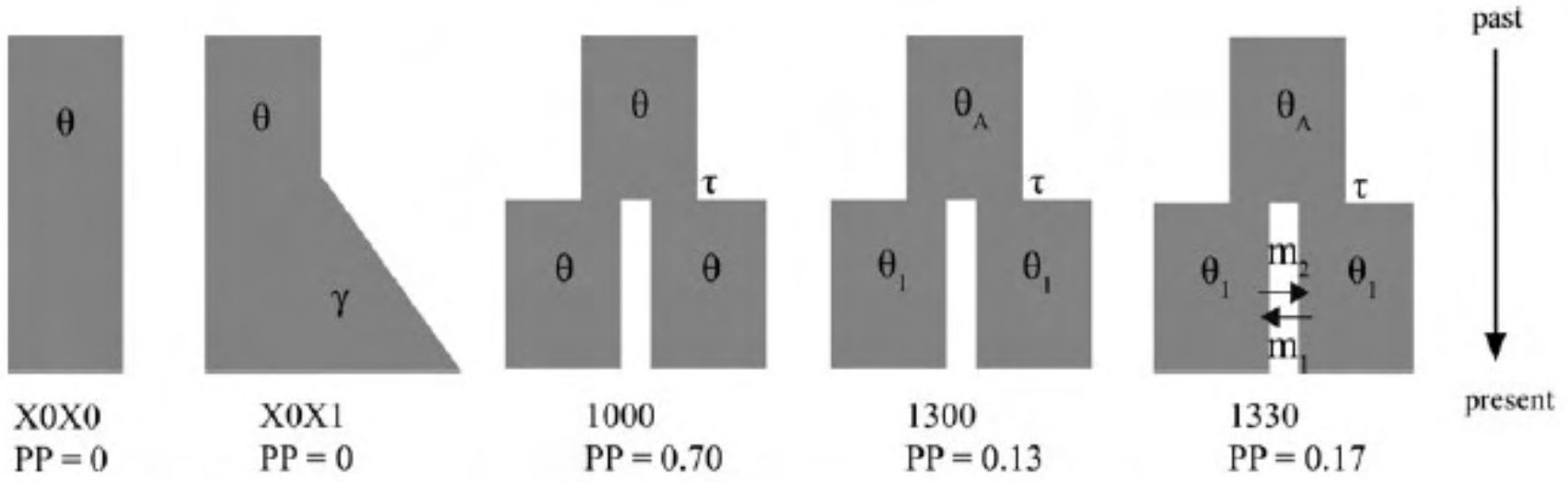


Fig. 2. Estimated expansion origins (Ω; red cross) in *C. cordiformis* (*A*) and *C. ovata* (*B*). The shading of pixels depicts a probability surface (kernel density) showing the likelihood that each pixel served as the expansion origin relative to the pixel with the highest likelihood (i.e., Ω). Glaciated regions are shown in blue. The results presented in *A* and *B* are based on retention of four and three PC axes of variation in genetic summary statistics, respectively. Results based on retaining additional PC axes are presented in *SI Appendix*, Figs. S2 and S3.

How do we decide upon a model:

Generic models in phylogeography

Tests of 142 objectively identified models (e.g., program like PHRAPL)



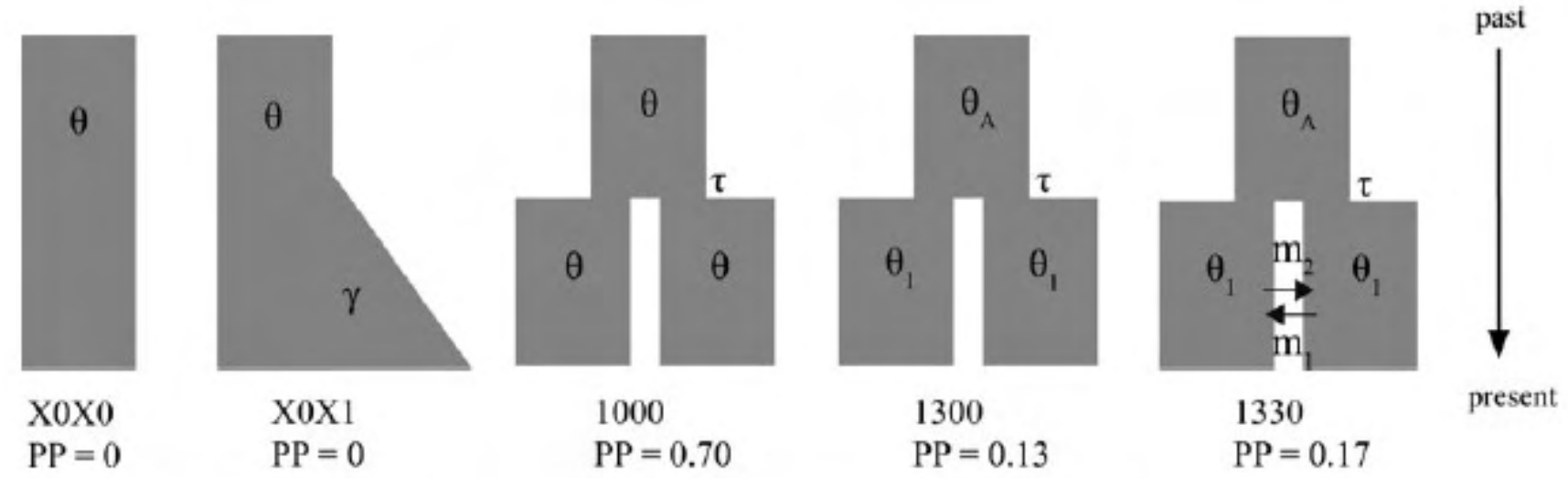
Pelletier & Carstens (2014 Mol. Ecol.)

- PHRAPL can create hundreds of possible histories that have a mixture of gene flow, population subdivision, and/or population size differences and compare these models using AIC (O'Meara)

Model choice in phylogeography: generic versus informed

- generic models

Tests of 142 objectively identified models

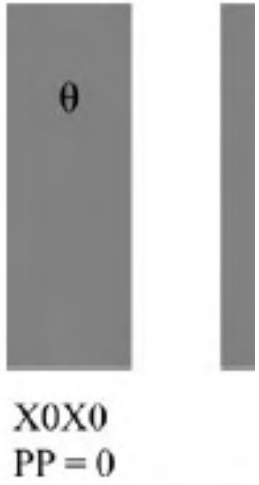


Pelletier & Carstens (2014 Mol. Ecol.)

Statistical procedures themselves may seem to provide a legitimacy to modeling decision – the advocacy of objective models in phylogeography

Model choice

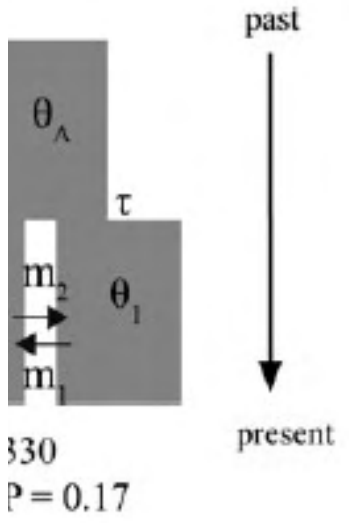
Tests of 142 c



- generic models

Table 3 List of all 143 models included in analyses. Model = τθmγ

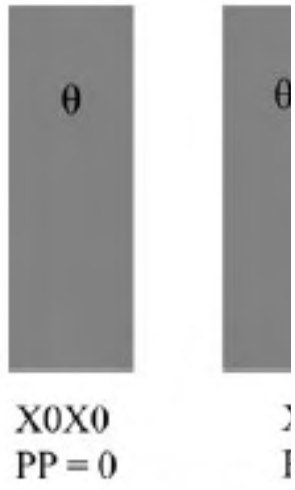
Model	Parameters	Mean	SD	Median	Posterior probability
1030	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, m_{21}$**	**0.792**	**1.124**	**0.000**	**0.024**
1232	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_2$	0.822	0.856	0.772	0.007
1200	$\tau, \theta_A = \theta_2, \theta_1$	0.836	0.985	0.499	0.004
1222	$\tau, \theta_A = \theta_2, \theta_1, m_{21}, \gamma_2$	0.846	0.982	0.542	0.006
1220	$\tau, \theta_A = \theta_2, \theta_1, m_{21}$	0.849	0.957	0.647	0.006
1231	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_1$	0.863	0.877	0.859	0.006
1221	$\tau, \theta_A = \theta_2, \theta_1, m_{21}, \gamma_1$	0.870	0.878	0.862	0.011
1031	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1$**	**0.886**	**1.133**	**0.000**	**0.020**
1230	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, m_{21}$	0.917	0.937	0.880	0.006
1033	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$**	**0.923**	**1.170**	**0.000**	**0.018**
0131	$\theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1$	0.930	1.024	0.779	0.007
0130	$\theta_A = \theta_1, \theta_2, m_{12}, m_{21}$	0.949	0.881	1.055	0.010
1023	**$\tau, \theta_A = \theta_1 = \theta_2, m_{21}, \gamma_1, \gamma_2$**	**0.956**	**1.154**	**0.000**	**0.024**
1201	$\tau, \theta_A = \theta_2, \theta_1, \gamma_1$	0.975	1.026	0.866	0.006
0030	**$\theta_A = \theta_1 = \theta_2, m_{12}, m_{21}$**	**0.977**	**1.210**	**0.000**	**0.024**
1211	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, \gamma_1$	0.990	1.042	0.927	0.007
0020	**$\theta_A = \theta_1 = \theta_2, m_{12}, m_{21}$**	**0.991**	**1.264**	**0.000**	**0.017**
1132	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_2$	0.995	0.981	0.986	0.007
0031	**$\theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1$**	**0.996**	**1.303**	**0.000**	**0.020**
0022	**$\theta_A = \theta_1 = \theta_2, m_{21}, \gamma_2$**	**1.003**	**1.241**	**0.000**	**0.025**
1131	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1$	1.011	0.967	1.013	0.004
1032	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_2$**	**1.013**	**1.212**	**0.000**	**0.031**
1212	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, \gamma_2$	1.015	0.986	1.083	0.003
1233	$\tau, \theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.021	0.946	1.121	0.010
1203	$\tau, \theta_A = \theta_2, \theta_1, \gamma_1, \gamma_2$	1.024	1.058	1.002	0.010
0233	$\theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.026	0.985	1.118	0.004
1110	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, \gamma_1$	1.030	1.003	1.118	0.007
0222	$\theta_A = \theta_2, \theta_1, m_{21}, \gamma_2$	1.031	1.112	0.921	0.008
1130	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, m_{21}$	1.031	0.976	1.084	0.006
0112	$\theta_A = \theta_1, \theta_2, m_{12}, \gamma_2$	1.032	0.991	1.121	0.007
0032	**$\theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_2$**	**1.033**	**1.212**	**0.000**	**0.020**
0110	$\theta_A = \theta_1, \theta_2, m_{12}, \gamma_1$	1.034	1.031	1.070	0.004
1020	$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.035	1.196	0.000	0.015
0012	**$\theta_A = \theta_1 = \theta_2, m_{12}, \gamma_2$**	**1.038**	**1.272**	**0.000**	**0.018**
1213	$\tau, \theta_A = \theta_2 = \theta_1, m_{12}, \gamma_1, \gamma_2$	1.041	1.053	1.121	0.003
0220	$\theta_A = \theta_2, \theta_1, m_{21}$	1.041	0.965	1.121	0.010
1013	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, \gamma_1, \gamma_2$**	**1.042**	**1.227**	**0.543**	**0.024**
0231	$\theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_1$	1.048	1.104	0.997	0.007
1111	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, \gamma_1$	1.050	1.027	1.098	0.013
0013	**$\theta_A = \theta_1 = \theta_2, m_{12}, \gamma_1, \gamma_2$**	**1.056**	**1.254**	**0.000**	**0.021**
0133	**$\theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$**	**1.057**	**1.107**	**1.028**	**0.001**
0033	**$\theta_A = \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$**	**1.059**	**1.289**	**0.000**	**0.031**
1002	$\tau, \theta_A = \theta_1 = \theta_2, \gamma_2$	1.084	1.261	0.000	0.008
1331	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1$	1.098	1.093	1.081	0.000
0132	$\theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_2$	1.101	0.991	1.129	0.007
0210	$\theta_A = \theta_2, \theta_1, m_{12}$	1.102	1.111	1.040	0.001
1321	$\tau, \theta_A, \theta_1 = \theta_2, m_{21}, \gamma_1$	1.108	1.012	1.124	0.000
1123	$\tau, \theta_A = \theta_1, \theta_2, m_{21}, \gamma_1, \gamma_2$	1.118	1.094	1.121	0.003
1021	**$\tau, \theta_A = \theta_1 = \theta_2, m_{21}, \gamma_1$**	**1.119**	**1.323**	**0.000**	**0.036**
1113	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, \gamma_1, \gamma_2$	1.132	1.042	1.129	0.003
1010	$\tau, \theta_A = \theta_1 = \theta_2, m_{12}$	1.135	1.284	0.558	0.013
1112	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, \gamma_1$	1.135	0.943	1.137	0.006
1101	$\tau, \theta_A = \theta_1, \theta_2, \gamma_1$	1.136	1.048	1.129	0.006
1011	**$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, \gamma_1$**	**1.148**	**1.274**	**0.739**	**0.021**
0023	**$\theta_A = \theta_1 = \theta_2, m_{21}, \gamma_1, \gamma_2$**	**1.154**	**1.311**	**0.500**	**0.020**



& Carstens (2014 Mol. Ecol.)

Model choice i

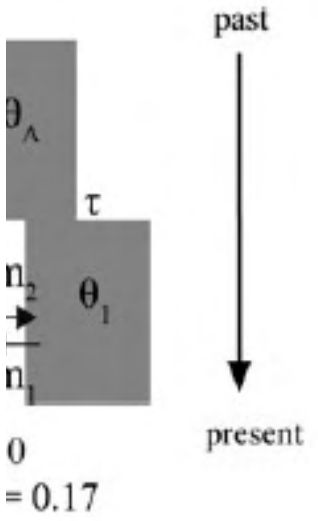
Tests of 142 ob



- generic models

Table 3 *Continued*

Model	Parameters	Mean	SD	Median	Posterior probability
0230	$\theta_A = \theta_2, \theta_1, m_{12}, m_{21}$	1.172	1.022	1.135	0.003
0321	$\theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1$	1.173	1.106	1.129	0.003
1000*	$\tau, \theta_A = \theta_1 = \theta_2$	1.178	1.261	0.971	0.015
1202	$\tau, \theta_A = \theta_1 = \theta_2, \gamma_2$	1.180	1.163	1.124	0.004
0223	$\theta_A = \theta_2, \theta_1, m_{21}, \gamma_1, \gamma_2$	1.181	1.173	1.124	0.007
1001	$\tau, \theta_A = \theta_1 = \theta_2, \gamma_1$	**1.187**	**1.328**	**0.752**	**0.021**
0011	$\theta_A = \theta_1 = \theta_2, m_{12}, \gamma_1$	**1.198**	**1.298**	**0.931**	**0.022**
0213	$\theta_A = \theta_2, \theta_1, m_{12}, \gamma_1, \gamma_2$	1.199	1.117	1.135	0.004
1102	$\tau, \theta_A = \theta_1, \theta_2, \gamma_2$	1.205	1.217	1.129	0.004
1121	$\tau, \theta_A = \theta_1, \theta_2, m_{21}, \gamma_1$	1.211	1.141	1.137	0.010
1022	$\tau, \theta_A = \theta_1 = \theta_2, m_{21}, \gamma_2$	**1.214**	**1.308**	**1.011**	**0.021**
1012	$\tau, \theta_A = \theta_1 = \theta_2, m_{12}, \gamma_2$	**1.270**	**1.324**	**1.129**	**0.021**
1332	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_2$	1.271	1.159	1.179	0.003
1322	$\tau, \theta_A, \theta_1 = \theta_2, m_{21}, \gamma_2$	1.280	1.087	1.233	0.000
0212	$\theta_A = \theta_2, \theta_1, m_{12}, \gamma_2$	1.281	1.181	1.140	0.001
1312	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, \gamma_2$	1.286	1.105	1.221	0.001
1323	$\tau, \theta_A, \theta_1 = \theta_2, m_{21}, \gamma_1, \gamma_2$	1.312	1.075	1.239	0.001
0123	$\theta_A = \theta_1, \theta_2, m_{21}, \gamma_1, \gamma_2$	1.312	1.189	1.192	0.007
1003	$\tau, \theta_A = \theta_1 = \theta_2, \gamma_1, \gamma_2$	1.321	1.443	1.122	0.007
0313	$\theta_A, \theta_1 = \theta_2, m_{12}, \gamma_1, \gamma_2$	1.327	1.207	1.182	0.001
1433	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.327	0.998	1.269	0.000
0312	$\theta_A, \theta_1 = \theta_2, m_{12}, \gamma_2$	1.328	1.201	1.209	0.004
0211	$\theta_A = \theta_2, \theta_1, m_{12}, \gamma_1$	1.333	1.195	1.256	0.006
1320	$\tau, \theta_A, \theta_1 = \theta_2, m_{21}$	1.336	1.235	1.180	0.001
1403	$\tau, \theta_A, \theta_1, \theta_2, \gamma_1, \gamma_2$	1.350	1.011	1.298	0.000
1330*	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, m_{21}$	1.351	1.274	1.225	0.006
0323	$\theta_A, \theta_1 = \theta_2, m_{21}, \gamma_1, \gamma_2$	1.353	1.170	1.259	0.003
1333	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.357	1.127	1.277	0.003
1103	$\tau, \theta_A = \theta_1, \theta_2, \gamma_1, \gamma_2$	1.400	1.186	1.408	0.003
1423	$\tau, \theta_A, \theta_1, \theta_2, m_{21}, \gamma_1, \gamma_2$	1.408	1.502	1.182	0.001
0331	$\theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1$	1.424	1.314	1.368	0.000
0311	$\theta_A, \theta_1 = \theta_2, m_{12}, \gamma_1$	1.475	1.353	1.353	0.003
1432	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_2$	1.500	1.297	1.360	0.000
1402	$\tau, \theta_A, \theta_1, \theta_2, \gamma_2$	1.543	1.101	1.545	0.003
0413	$\theta_A, \theta_1, \theta_2, m_{12}, \gamma_1, \gamma_2$	1.570	1.139	1.545	0.006
0412	$\theta_A, \theta_1, \theta_2, m_{12}, \gamma_2$	1.575	1.172	1.516	0.001
0322	$\theta_A, \theta_1 = \theta_2, m_{21}, \gamma_2$	1.591	1.493	1.481	0.001
1303	$\tau, \theta_A, \theta_1 = \theta_2, \gamma_1, \gamma_2$	1.591	1.303	1.610	0.003
1301	$\tau, \theta_A, \theta_1 = \theta_2, \gamma_1$	1.621	1.428	1.554	0.001
1300*	$\tau, \theta_A, \theta_1 = \theta_2$	1.630	1.342	1.562	0.004
1313	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, \gamma_1, \gamma_2$	1.676	3.419	1.164	0.007
0423	$\theta_A, \theta_1, \theta_2, m_{21}, \gamma_1, \gamma_2$	1.710	1.358	1.593	0.000
0430	$\theta_A, \theta_1, \theta_2, m_{12}, m_{21}$	1.715	1.294	1.620	0.000
0113	$\theta_A, \theta_1 = \theta_2, m_{12}, \gamma_1, \gamma_2$	1.715	5.727	1.068	0.004
0411	$\theta_A, \theta_1, \theta_2, m_{12}, \gamma_1$	1.717	1.259	1.665	0.003
0422	$\theta_A, \theta_1, \theta_2, m_{21}, \gamma_2$	1.759	1.417	1.614	0.000
1401	$\tau, \theta_A, \theta_1, \theta_2, \gamma_1$	1.781	1.835	1.505	0.001
0433	$\theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	1.843	1.773	1.597	0.000
0021	$\theta_A = \theta_1 = \theta_2, m_{21}, \gamma_1$	1.867	4.813	0.673	0.014
0221	$\theta_A = \theta_2, \theta_1, m_{21}, \gamma_1$	1.934	6.915	0.937	0.006
1400	$\tau, \theta_A, \theta_1, \theta_2$	2.098	1.697	1.899	0.000
0232	$\theta_A = \theta_2, \theta_1, m_{12}, m_{21}, \gamma_2$	2.186	7.859	1.121	0.007
0122	$\theta_A = \theta_1, \theta_2, m_{21}, \gamma_2$	2.356	7.532	1.254	0.006
1122	$\tau, \theta_A = \theta_1, \theta_2, m_{21}, \gamma_2$	2.551	8.798	1.283	0.003
1133	$\tau, \theta_A = \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	2.748	12.927	0.814	0.008
1410	$\tau, \theta_A, \theta_1, \theta_2, m_{12}$	2.790	7.890	1.673	0.003



Carstens (2014 Mol. Ecol.)

Model choice in phylogeography: generic versus informed

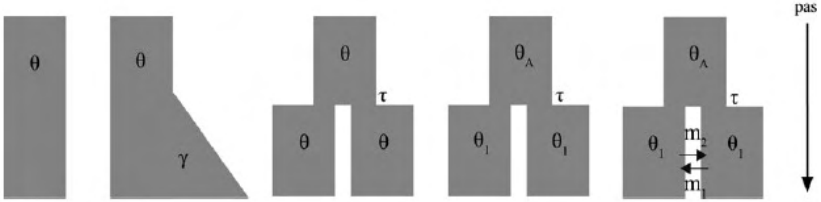
Table 3 *Continued*

Model	Parameters	Mean	SD	Median	Posterior probability
1420	$\tau, \theta_A, \theta_1, \theta_2, m_{21}$	2.819	9.142	1.557	0.001
0330	$\theta_A, \theta_1 = \theta_2, m_{12}, m_{21}$	3.156	11.980	1.608	0.000
0431	$\theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1$	3.388	12.338	1.687	0.001
0432	$\theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_2$	3.769	15.818	1.606	0.003
1210	$\tau, \theta_A = \theta_2, \theta_1, m_{12}$	4.007	21.699	0.880	0.010
0310	$\theta_A, \theta_1 = \theta_2, m_{12}$	4.405	20.648	1.670	0.001
0421	$\theta_A, \theta_1, \theta_2, m_{21}, \gamma_1$	4.761	18.586	1.563	0.000
1223	$\tau, \theta_A = \theta_2, \theta_1, m_{21}, \gamma_1, \gamma_2$	4.813	27.942	0.880	0.007
0410	$\theta_A, \theta_1, \theta_2, m_{12}$	4.840	19.483	1.684	0.000
0333	$\theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_1, \gamma_2$	4.841	24.764	1.304	0.004
1411	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, \gamma_1$	4.949	22.725	1.182	0.000
0320	$\theta_A, \theta_1 = \theta_2, m_{21}$	5.184	25.275	1.771	0.000
1431	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, m_{21}, \gamma_1$	5.539	28.987	1.440	0.000
1421	$\tau, \theta_A, \theta_1, \theta_2, m_{21}, \gamma_1$	5.618	22.805	1.418	0.001
1311	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}, \gamma_1$	5.721	32.177	1.137	0.001
0111	$\theta_A = \theta_1, \theta_2, m_{12}, \gamma_1$	5.804	32.950	1.143	0.008
0420	$\theta_A, \theta_1, \theta_2, m_{21}$	6.037	28.946	1.629	0.001
1412	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, \gamma_2$	6.186	23.177	1.611	0.003
0010	$\theta_A = \theta_1 = \theta_2, m_{12}$	6.223	36.293	0.000	0.017
1413	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, \gamma_1, \gamma_2$	8.209	48.083	1.344	0.000
1430	$\tau, \theta_A, \theta_1, \theta_2, m_{12}, m_{21}$	8.661	50.499	1.516	0.001
1422	$\tau, \theta_A, \theta_1, \theta_2, m_{21}, \gamma_2$	9.269	45.089	1.344	0.006
0121	$\theta_A = \theta_1, \theta_2, m_{21}, \gamma_1$	9.369	56.607	1.327	0.004
1302	$\tau, \theta_A, \theta_1 = \theta_2, \gamma_2$	9.386	44.243	1.233	0.004
0120	$\theta_A = \theta_1, \theta_2, m_{21}$	9.466	57.924	1.189	0.004
1310	$\tau, \theta_A, \theta_1 = \theta_2, m_{12}$	9.812	60.333	1.206	0.000
1100	$\tau, \theta_A = \theta_1, \theta_2$	10.795	68.438	1.121	0.007
0332	$\theta_A, \theta_1 = \theta_2, m_{12}, m_{21}, \gamma_2$	13.053	82.999	1.415	0.004
1120	$\tau, \theta_A = \theta_1, \theta_2, m_{21}$	14.667	54.818	1.365	0.007
X0X1*	θ_A, γ_1	16.013	5.576	15.576	0.000
X0X0*	θ_A	17.048	7.013	16.115	0.000
0000	$\theta_A = \theta_1 = \theta_2$				

For each model: τθmγ

The answer is model 1023!

Divergence time (τ)	Theta (θ)	Migration (m)	Population expansion (γ)
0: island model 1: divergence at time (τ) X: pamixia Prior: 0.001–5 (4N generations)	0: $\theta_A = \theta_1 = \theta_2$ 1: $\theta_A = \theta_1, \theta_2$ 2: $\theta_A = \theta_2, \theta_1$ 3: $\theta_A, \theta_1 = \theta_2$ 4: $\theta_A, \theta_1, \theta_2$ Prior: 0.01–10 per locus	0: no migration 1: m_{12} 2: m_{21} 3: m_{12}, m_{21} X: na/pamixia Prior: 0–5 migrants per generation	0: no expansion 1: γ_1 2: γ_2 3: γ_1, γ_2 Prior: 0.1–9 (exponential)

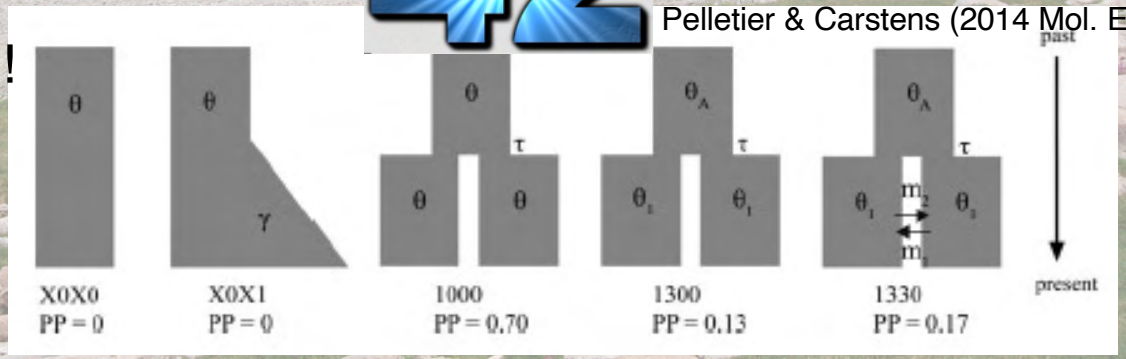


Biological insights depend on the questions we (the scientist) ask!

- Should we expect (or want) or computer programs to define the questions we ask!?!

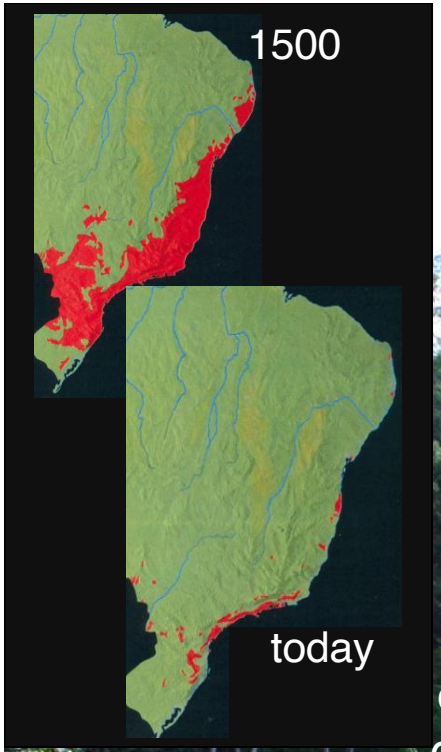
The answer is: 42

The answer is model 1023!



Pelletier & Carstens (2014 Mol. Ecol.)

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



Model-based approach:

Forecasting spatial patterns of diversity in poorly explored, highly threatened ecosystems

Model-based approach:

Directly model historical processes through a combination of ecological-niche models under paleoclimates and genetic analyses, discovered a central region in the Brazilian Atlantic forest that served as a biodiversity refuge during climatic extremes.

H. semilineatus

Hypsiboas albomarginatus

H. faber

Carnaval et al. 2009. *Science*

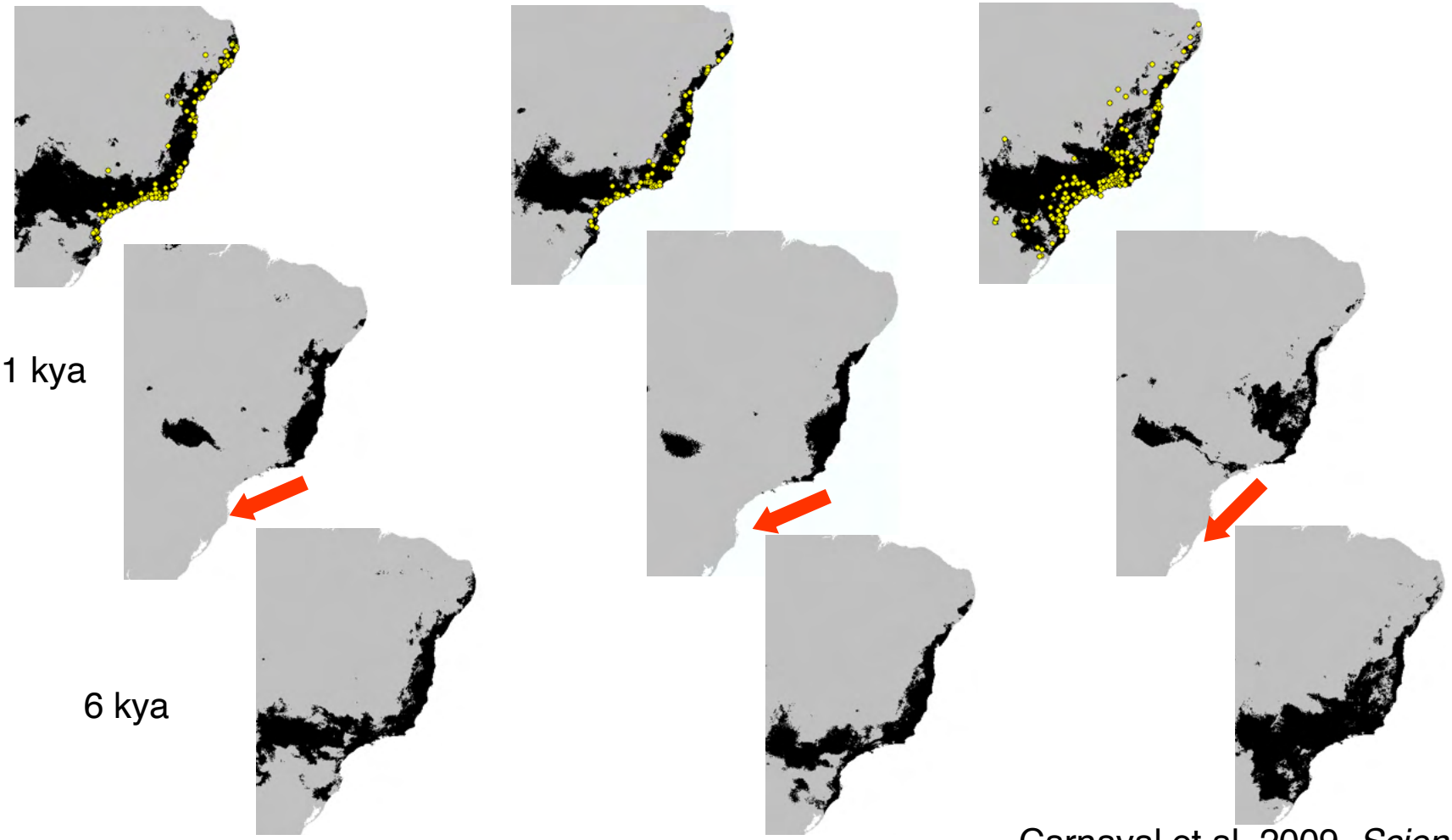
Model species distributions under current conditions and climatic extremes (based on climatic niches with MAXENT)



Now

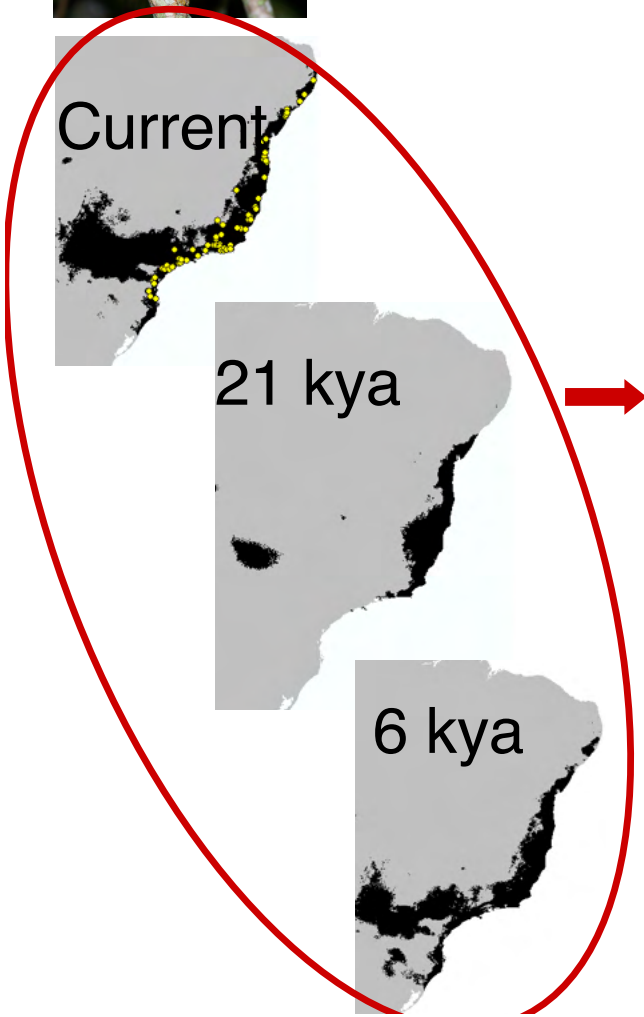
21 kya

6 kya



Carnaval et al. 2009. *Science*

Model species distributions under current conditions and climatic extremes (based on climatic niches with MAXENT)



Predicted region of stability

Maps of stable and unstable areas raise specific hypotheses about regional differences in persistence and hence diversity, which lead to phylogeographic predictions that can be tested with molecular data

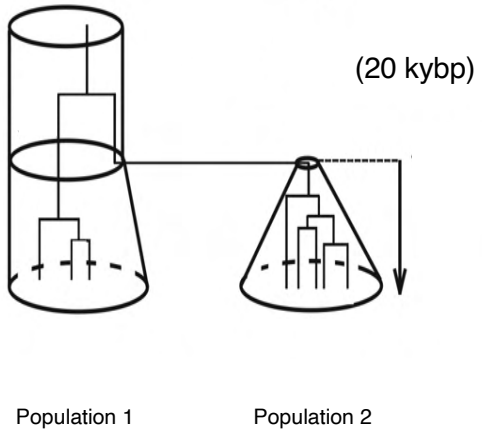
Carnaval et al. 2009. *Science*

Different demographic scenarios motivated by stable/unstable areas:

Results support community responses for both models using hierarchical Approximate Bayesian Computation:

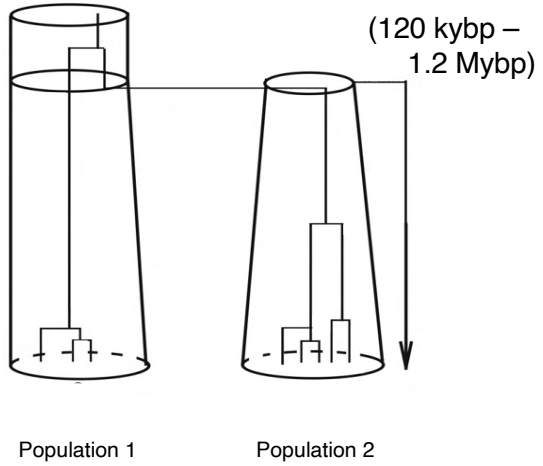
(i) simultaneous, multi-species colonization of unstable areas from adjacent refugial populations since the LGM

recent colonization



(ii) assemblage-scale, long-term persistence of populations in isolated refugial areas (i.e., temporally stable regions)

long-term persistence



Carnaval et al. 2009. *Science*

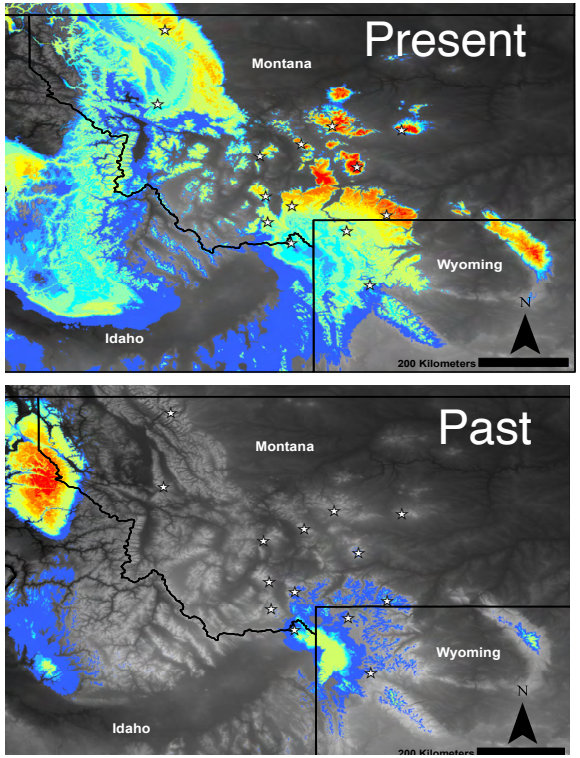
* All models are simplifications, but they vary in their relative degree of abstraction

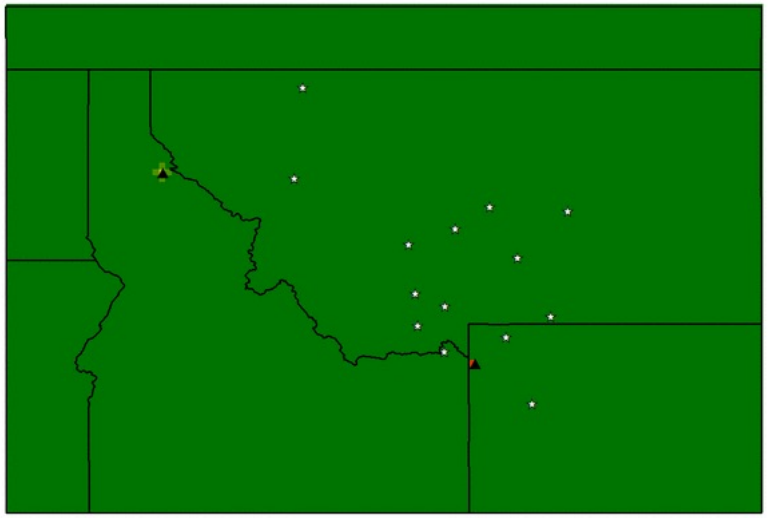
Different ways to model population expansion:

(i) Model as population size change with no spatial aspect of expansion (e.g., Brazilian Atlantic forest areas of instability associated with recent expansion)

(ii) Model expansion process across landscape explicitly

ENM based on paleoclimatic data 6kya





iDDC: Generate species-specific expectations for patterns of genetic variation

He, Edwards & Knowles, Evolution 2013

integrative
Distributional
Demographic
Coalescent
modeling

Distributional model
(i.e., ecological niche model) with
predictions on probability of occurrence
across the landscape

⇩

Demographic model
informed by habitat
suitabilities

⇩

Spatially-explicit coalescent
simulations based on
demographic model

⇩

Tests of hypotheses/models
using ABC

Habitat suitability scores

40	20	10	5
100	60	20	10
100	100	40	40
80	80	60	60

K(m)

400 (40)	200 (20)	100 (10)	50 (5)
1000 (100)	600 (60)	200 (20)	100 (10)
1000 (100)	1000 (100)	400 (40)	400 (40)
800 (80)	800 (80)	600 (60)	600 (60)

Carrying capacity: k_i

Gene coalescence across the landscape

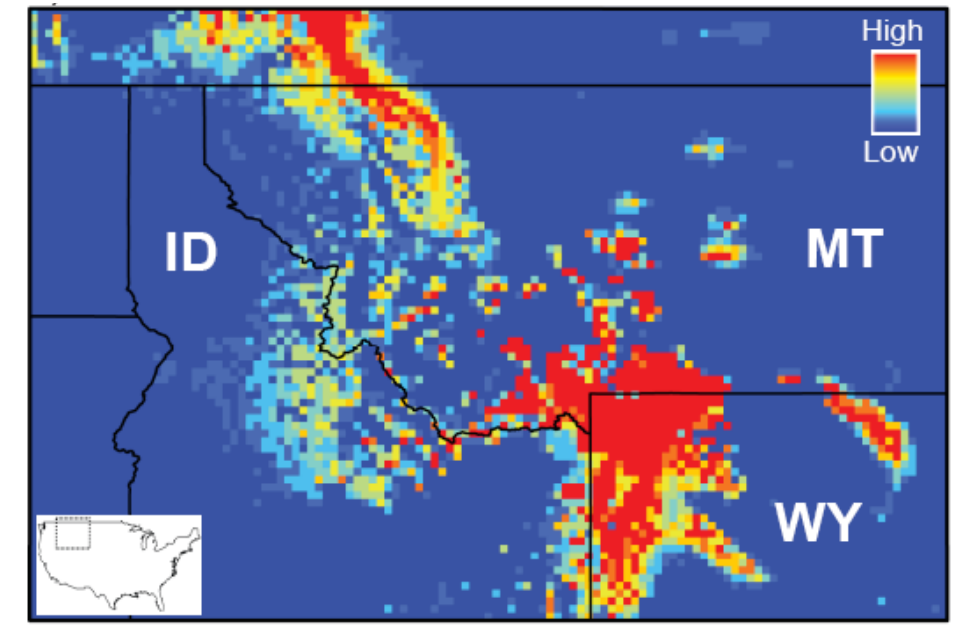
SPLATCHE2

Also could use program SLim

also Knowles & Serrano Mol. Ecol., 2010

iDDC: integrative Distributional, Demographic, Coalescent modeling

SPECIES-SPECIFIC
Spatially explicit quantitative information about probabilities of occurrence based on habitat suitability



low habitat suitability

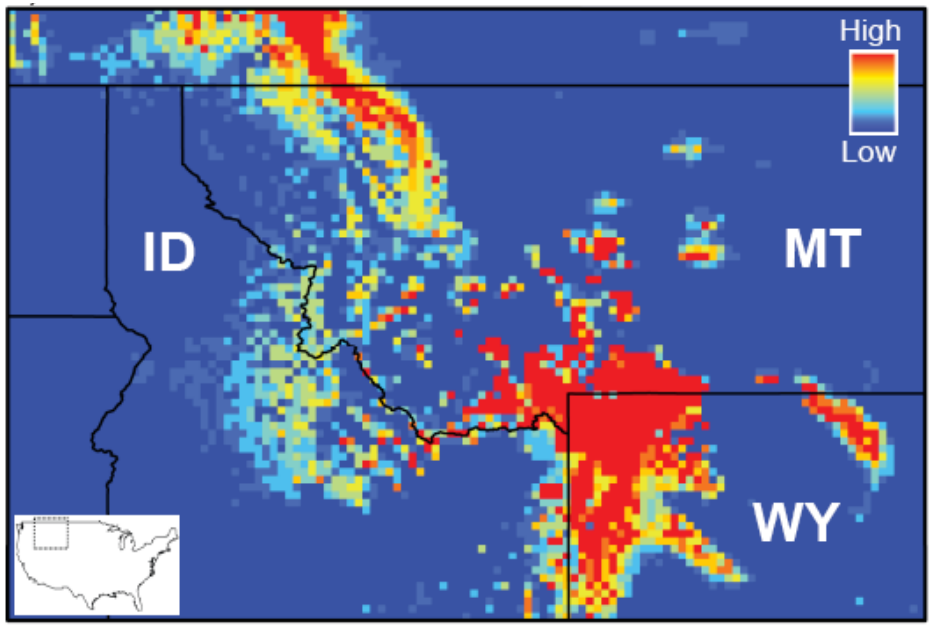
high habitat suitability

Habitat suitability scores

40	20	10	5
100	60	20	10
100	100	40	40
80	80	60	60

iDDC: integrative Distributional, Demographic, Coalescent modeling

Spatially explicit
probabilities of occurrence based on
habitat suitability



low habitat suitability — high habitat suitability

SPECIES-SPECIFIC Spatially explicit demographic model

- carrying capacity: *k*
- migration rate: m
- logistic growth rate: r

K(m)

400 (40)	200 (20)	100 (10)	50 (5)
1000 (100)	600 (60)	200 (20)	100 (10)
1000 (100)	1000 (100)	400 (40)	400 (40)
800 (80)	800 (80)	600 (60)	600 (60)

low *K* — high *K*

low *m* — High *m*

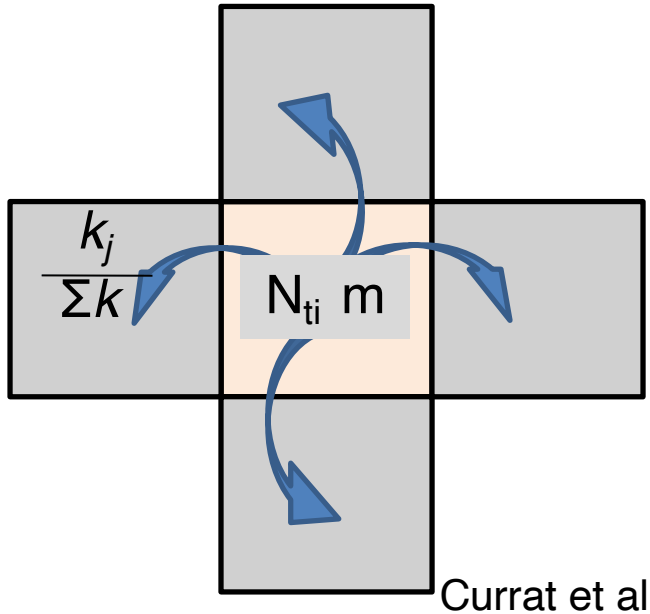
e.g.: SPECIES-SPECIFIC Demographic model:

At each generation:
-the population density, N_{ti}, of each deme is logistically regulated
-followed by a migration step
-the population densities and number of immigrants (N_{ti} and m) are stored and used during the genetic simulations

- carrying capacity: k_i

- # of emigrants leaving deme i: $N_{ti}\, m$

- # of immigrants entering deme j: $\frac{k_j}{\Sigma k}$



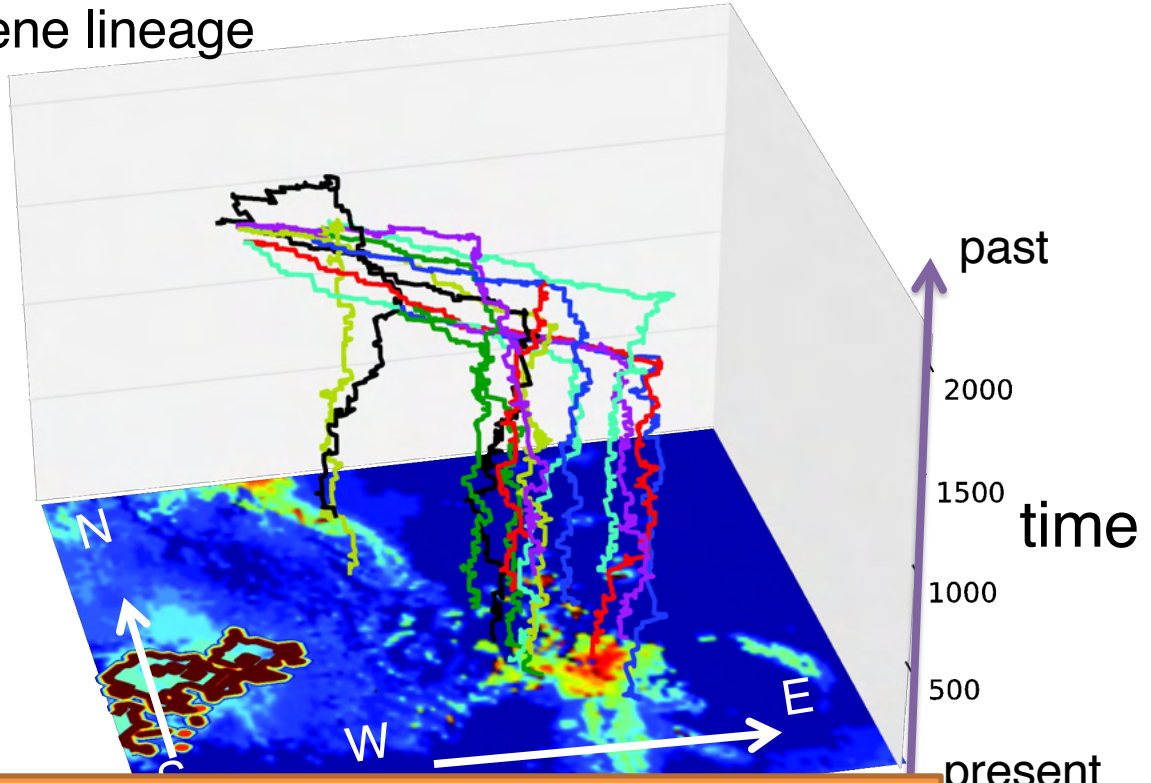
Currat et al. 2004

iDDC: integrative Distributional, Demographic, Coalescent modeling

- spatially-explicit genealogies to generate genetic patterns

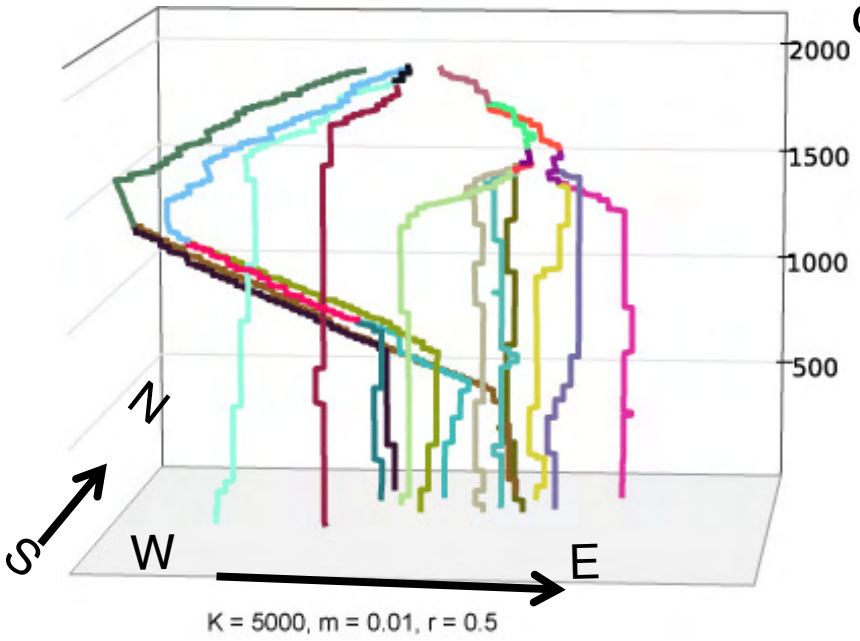
- at each generation (looking backwards in time), and depending on the local population sizes and migration rates from the demographic model, which are specified by ENMs, genes have probability of:

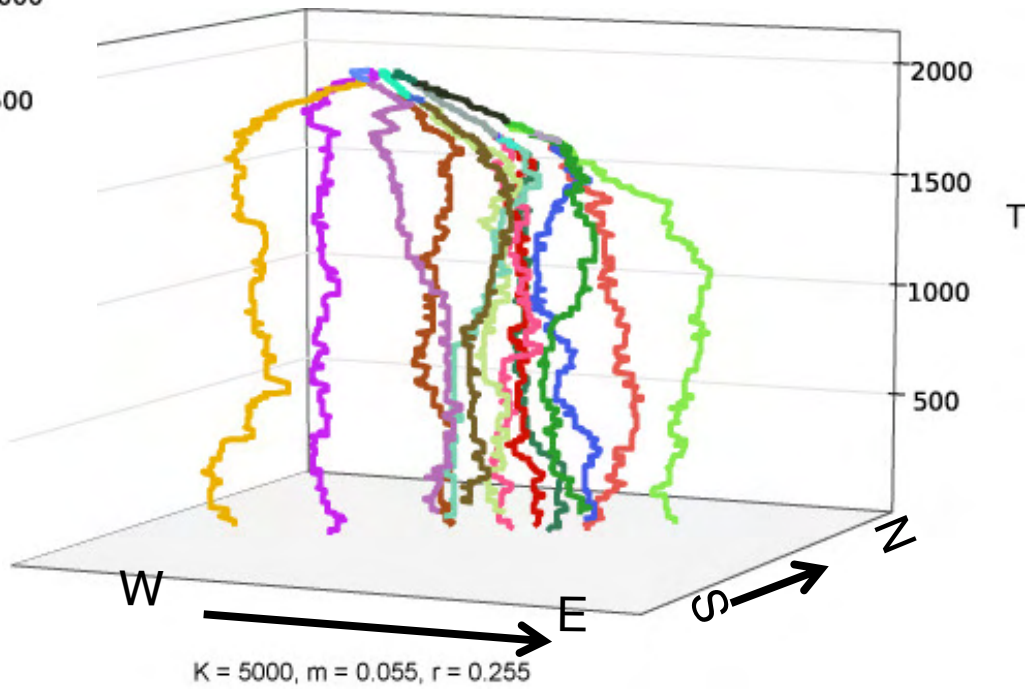
(i) staying in the same deme,
(ii) move to a different deme, or
(iii) coalesce with another gene lineage



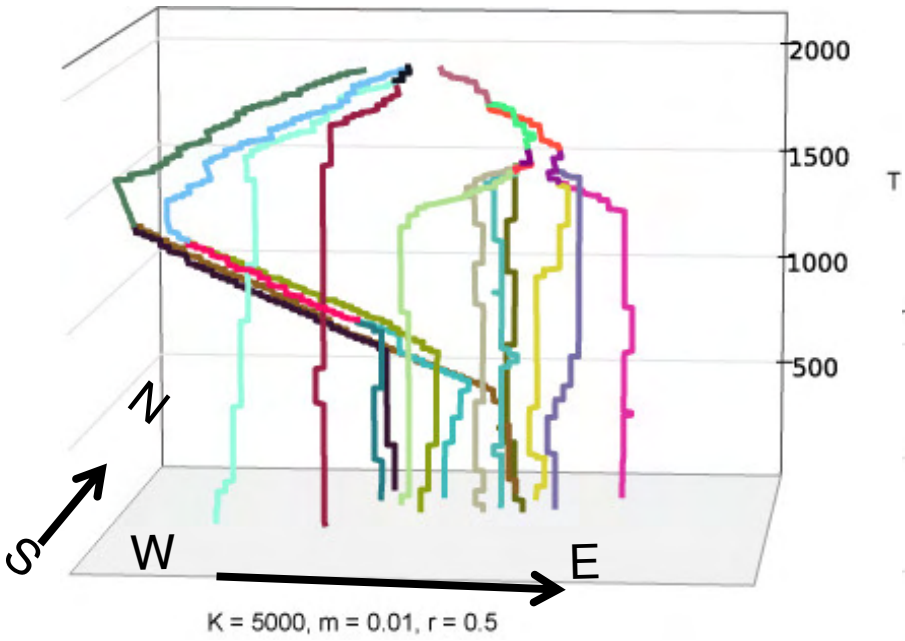
Suggested software: Quesztal package (Becheler and Knowles 2022)

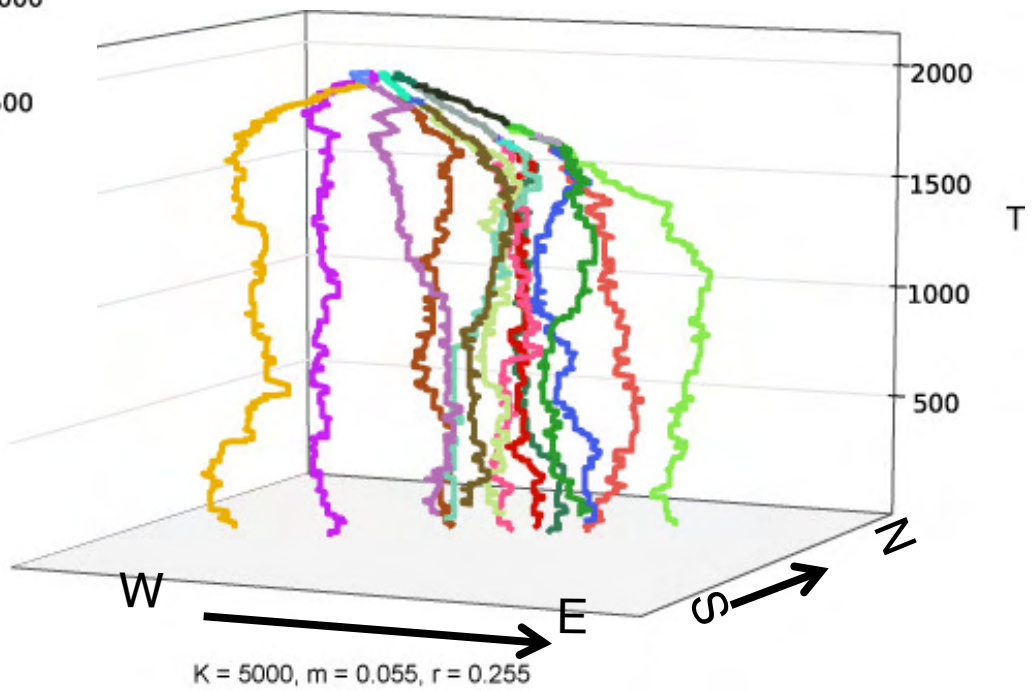
Under different demographic parameters (e.g., different *k* and *m*), same set of sampled populations would have different histories because the geographic location and timing of coalescence will differ.





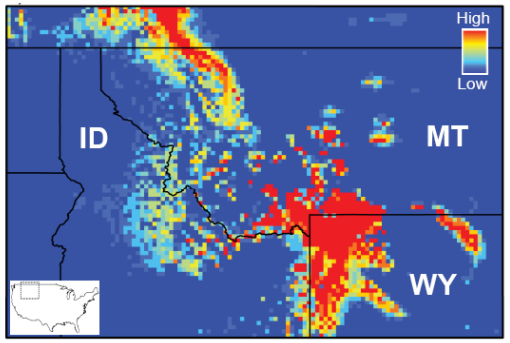
Simulate predicted patterns of genetic variation for set of parameters under the model





Mutations accumulate along the branches of the genealogy according to a Poisson process with rate μt, so different gene genealogies will produce difference in genetic diversity and the geographic distribution of genetic variation

Species-distribution model (SDM) generates predictions on probability of occurrence across the landscape

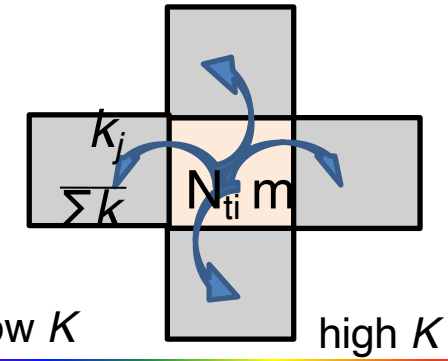


low habitat suitability — high habitat suitability

Habitat suitability scores

40	20	10	5
100	60	20	10
100	100	40	40
80	80	60	60

Spatially explicit demographic model (localized population densities, migration and growth rates)



low *K* — high *K*

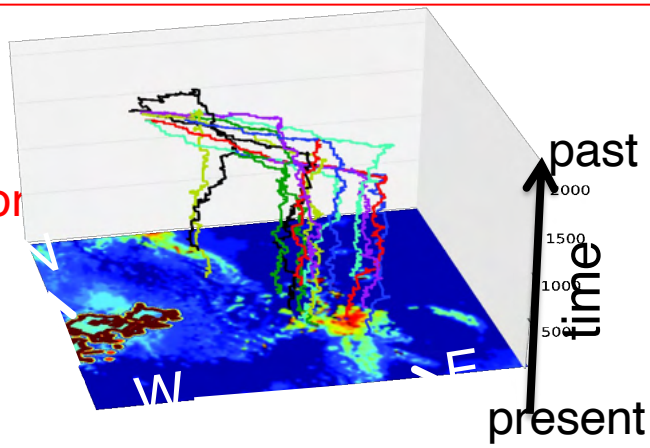
low *m* and *r* — High *m* and *r*

K(m)

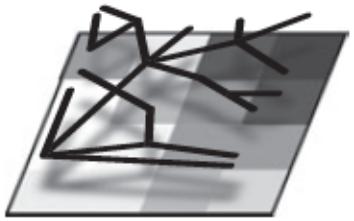
400 (40)	200 (20)	100 (10)	50 (5)
1000 (100)	600 (60)	200 (20)	100 (10)
1000 (100)	1000 (100)	400 (40)	400 (40)
800 (80)	800 (80)	600 (60)	600 (60)

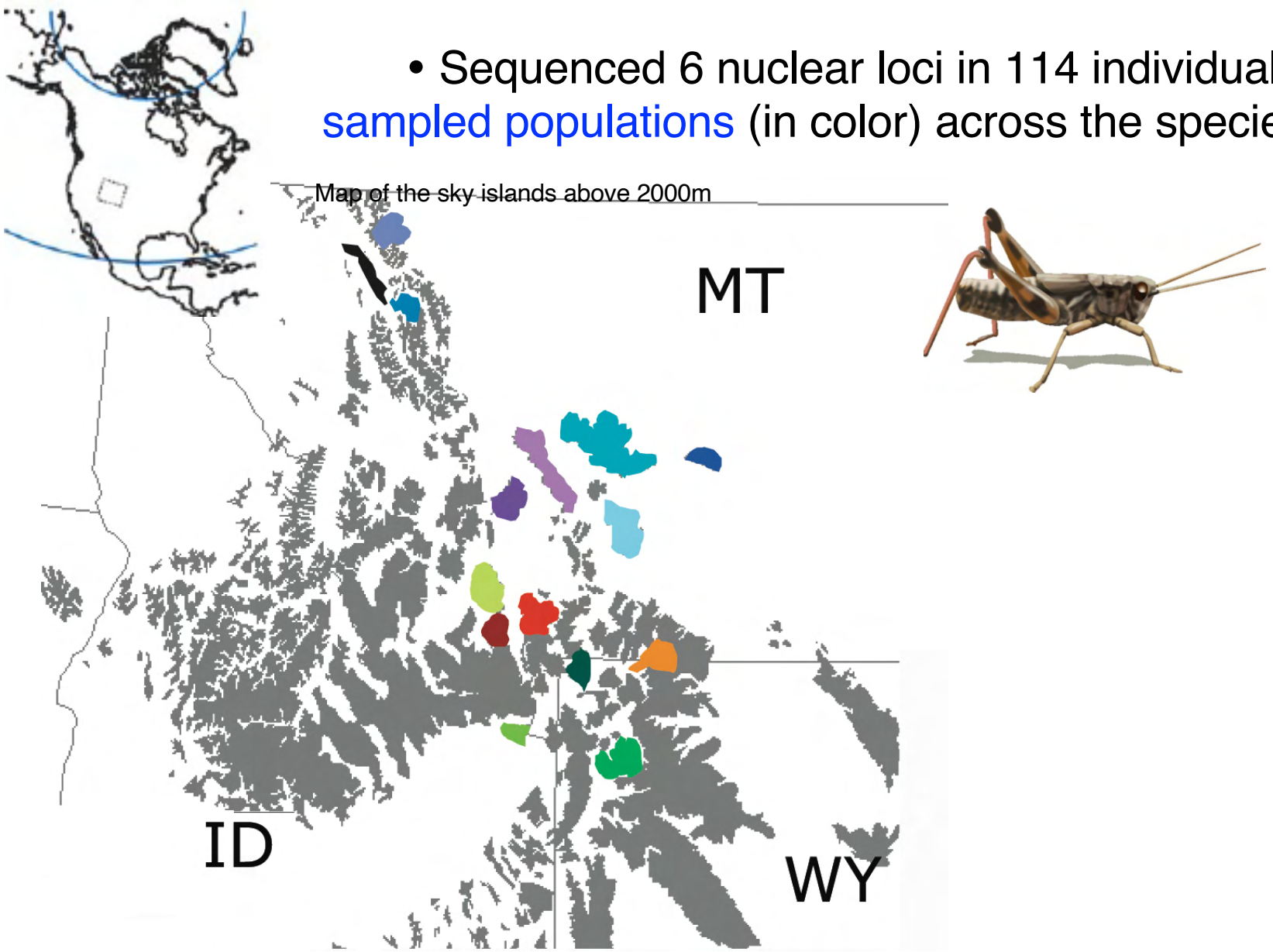
Carrying capacity: k_i

Spatially explicit coalescent model to generate predicted patterns of genetic variation for the empirically sampled population localities



Gene coalescence across the landscape





• Sequenced 6 nuclear loci in 114 individuals of sampled populations (in color) across the species range

Massatti R, Knowles LL (2020) The historical context of contemporary climatic adaptation: a case study in the climatically dynamic and environmentally complex southwestern United States. *Ecography* 43:735-746. doi.org/10.1111/ecog.04840

iDDC : Model Selection & Parameter Estimation using Approximate Bayesian Computation (ABC)

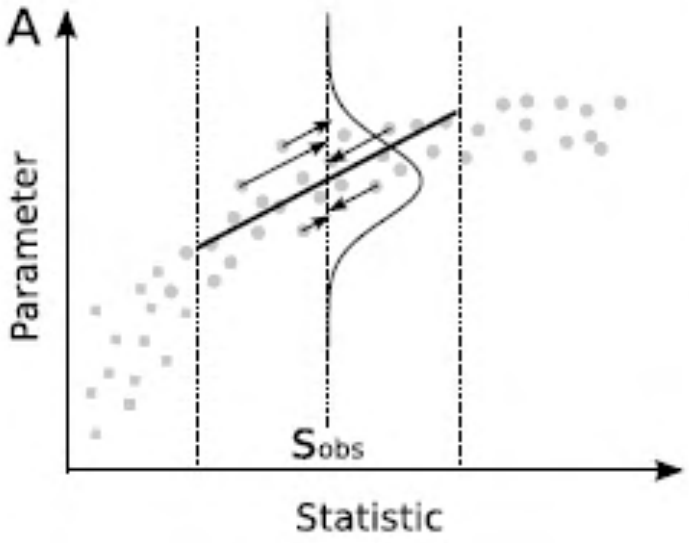
Model

Simulated datasets → Summary statistics (simulated data)

See Beaumont et al. 2002

Advantages:
- computational efficiency compared to ML methods
- allow for complex models

Summary statistics empirical data → Estimated parameters



We can identify sets of parameters for specific models that produce simulated data that matches the empirical data.

Suggested software: abctoolbox (Wegmann et al. 2010)

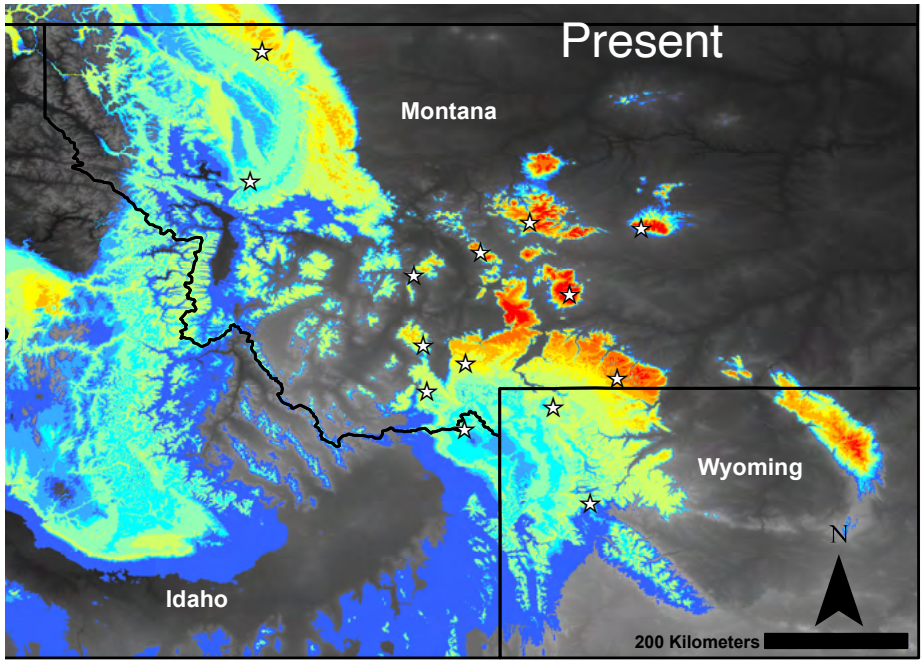
What geographic configuration of sky island populations promotes species divergence?

Population connectivity determined by contemporary sky island distribution

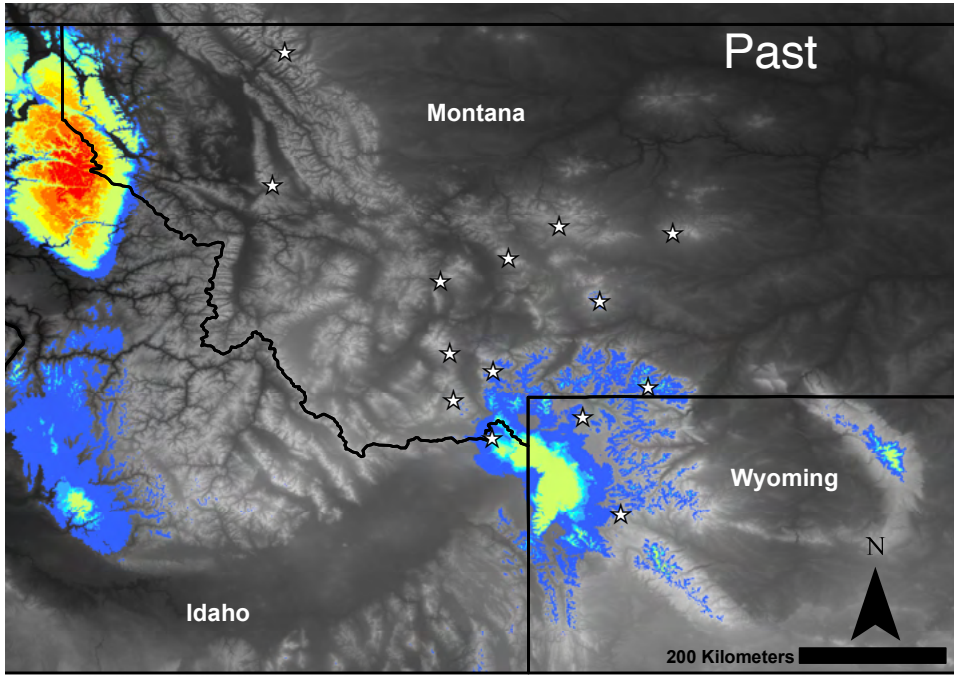
versus

Colonization of present sky island distribution from glacial refugia

19 bioclimatic variables used in modeling distributions



ENM based on current environmental data



ENM based on paleoclimatic data 6kya

- grasshoppers are flightless habitat specialists restricted to montane meadows

iDDC tests of drivers of divergence

population connectivity determined by contemporary sky island distribution

Choose model

Colonization of present sky island distribution from glacial refugia

Setting priors for parameters (K and m)

Simulations: demographic -> Coalescent -> DNA

Calculate summary statistics

Retain simulations whose SS are close to empirical ones

Model Selection and parameter estimation

Knowledge of geologic history and natural history were key in formulating hypotheses!

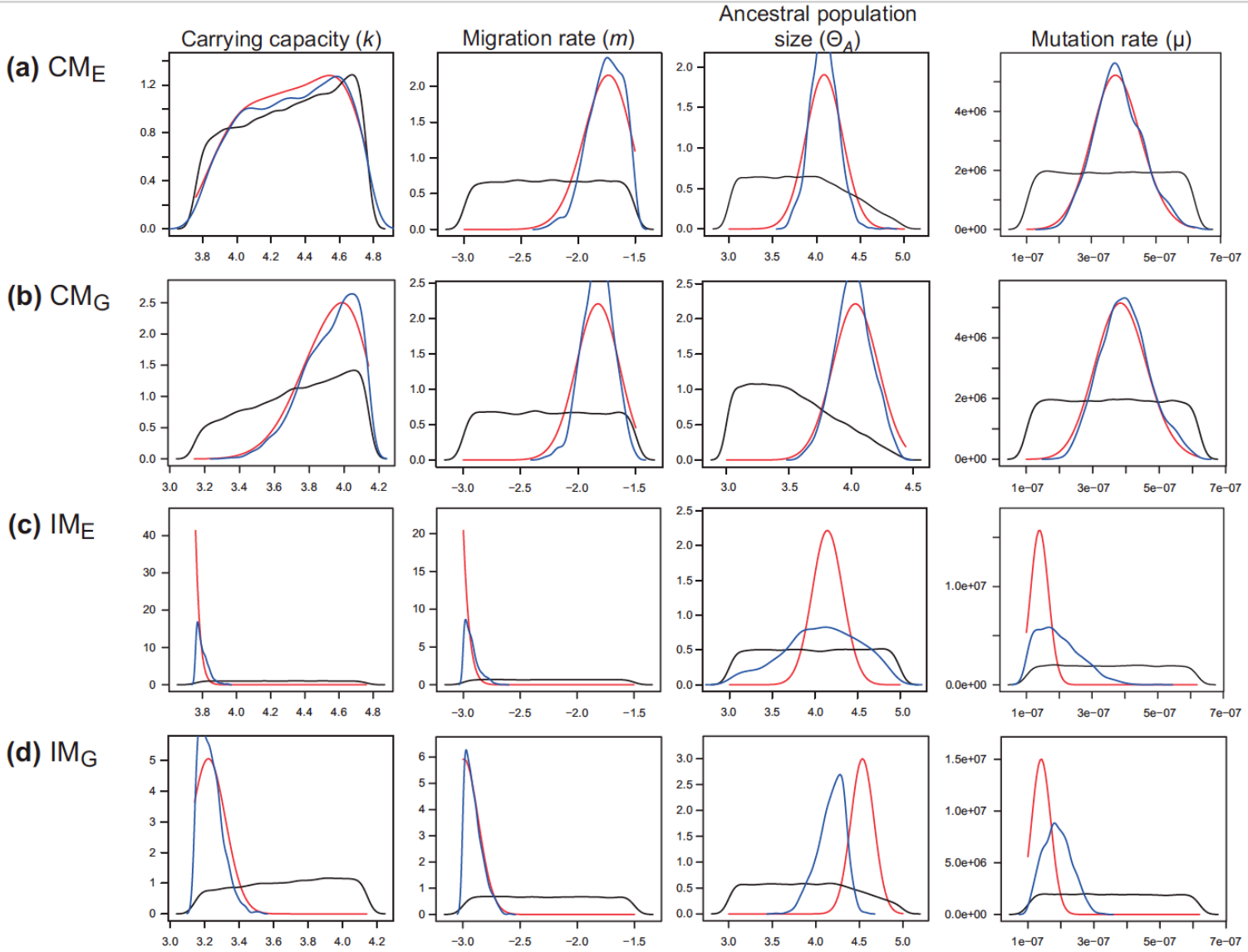
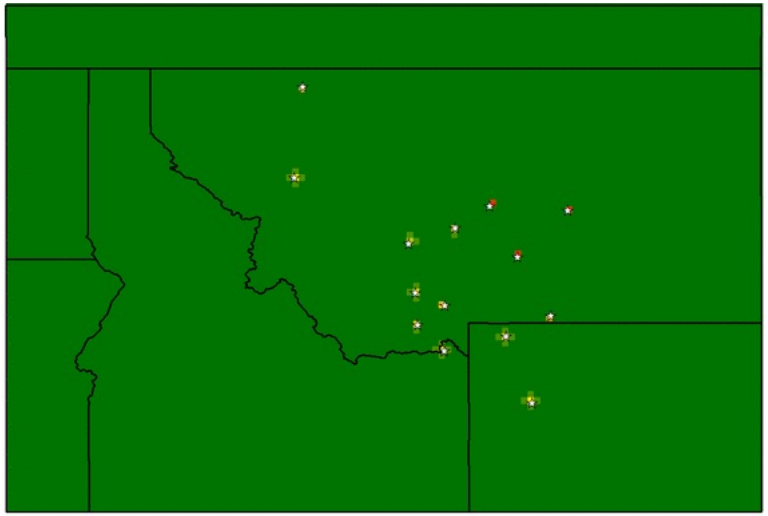
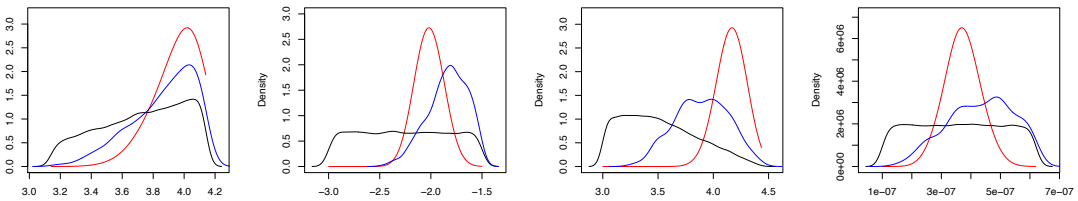
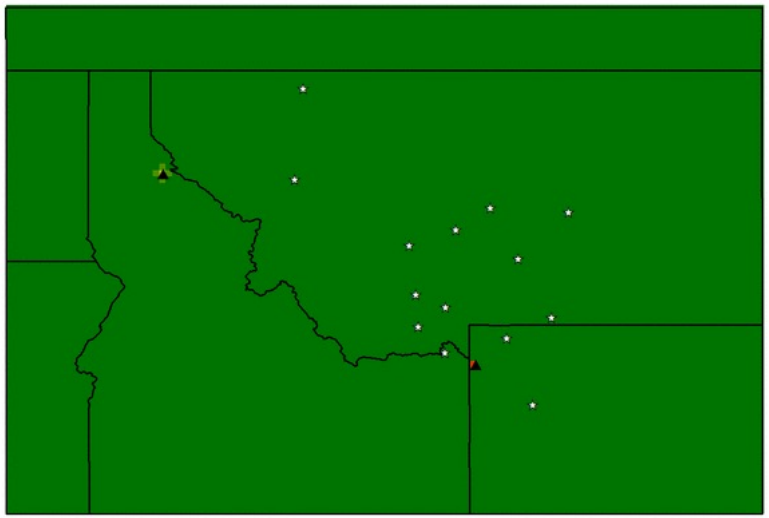


Figure 4. Posterior distribution (red line) of parameter estimates (i.e. carrying capacity, k, migration rate, m, ancestral population size θ_A, and mutation rate, μ) for each of the two colonization models, (a) CM_E and (b) CM_G, and the two sky island isolation models, (c) IM_E and (d) IM_G, where the subscripts E and G refer to connectivity patterns determined by either environmental heterogeneity or geographic distance, respectively. Results are based on a GLM regression adjustment of the 5000 closet simulations to each model. The distribution of the retained simulations (blue line) and the prior (black line) demonstrate the improvement that the GLM procedure had on parameter estimates and that the data contained information relevant to estimating the parameters.







Model tests based on comparing marginal likelihoods:

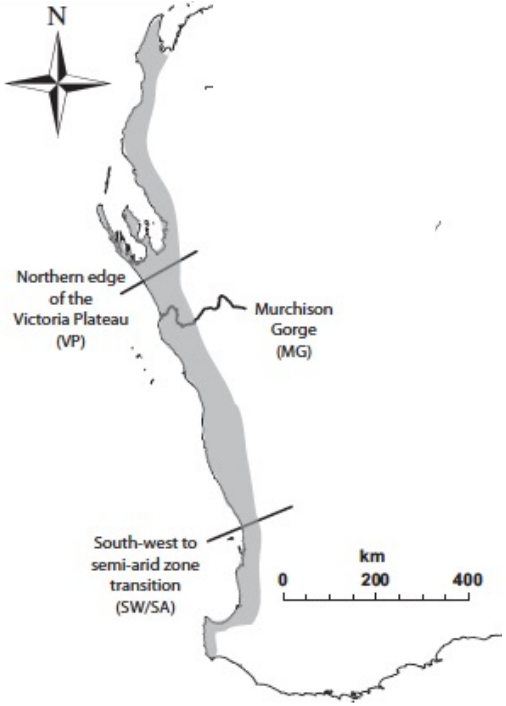
(i) population connectivity determined by contemporary sky island distribution

Patterns of genetic variation reflect: (ii) a colonization history from glacial refugia to present sky island distribution

Knowles LL, *Massatti R* (2017) Distributional shifts – not geographic isolation – as a probable driver of montane species divergence. *Ecography* 40:1475-1485.

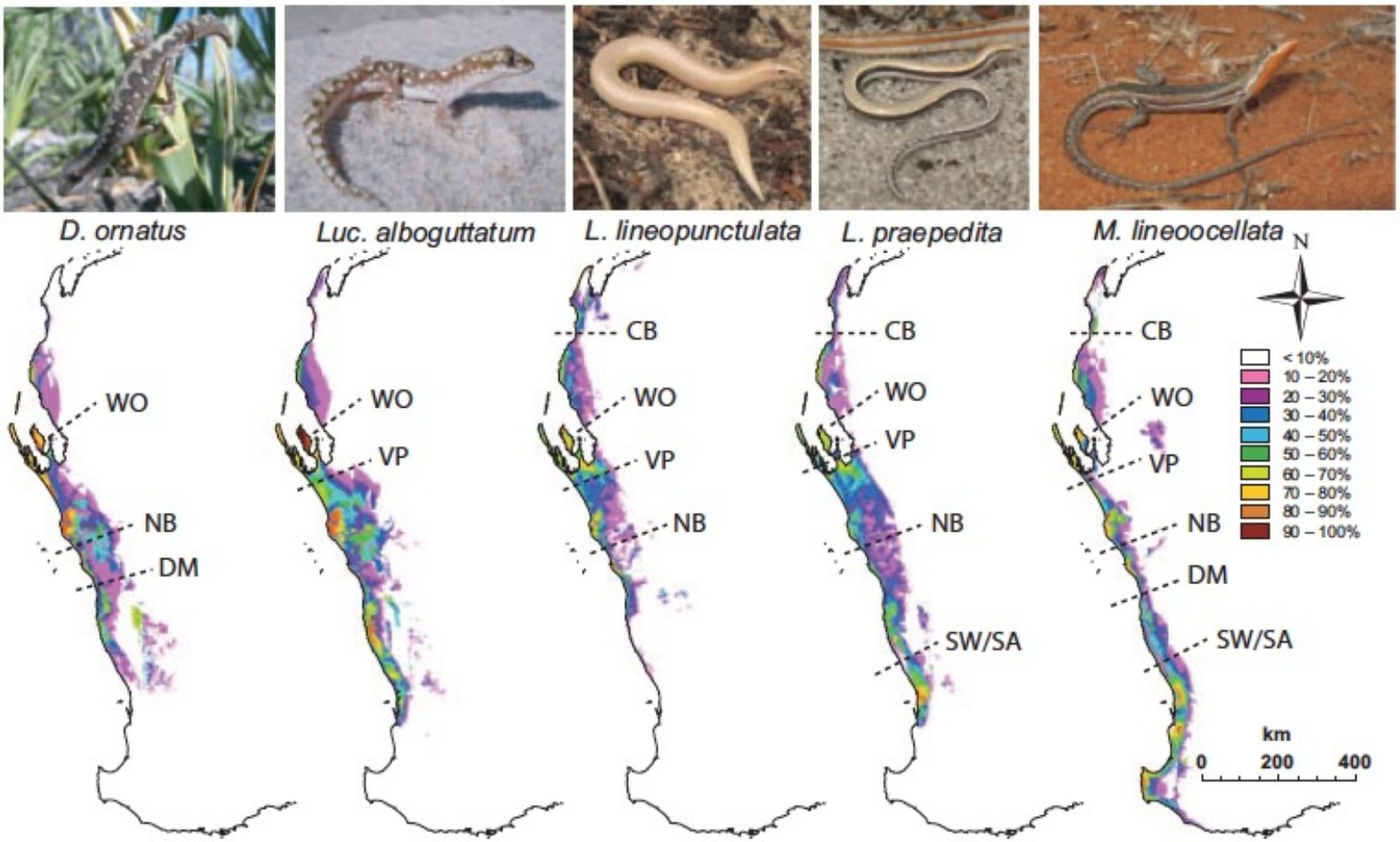
How do we decide upon a model:

Knowledge of geologic history, ecology and and natural history are key in formulating hypotheses!



Linear distribution of populations along SW coast suggests isolation-by-distance may be important in structuring patterns of genetic variation

Edwards, Keogh, Knowles (2012) *Mol. Ecol.*



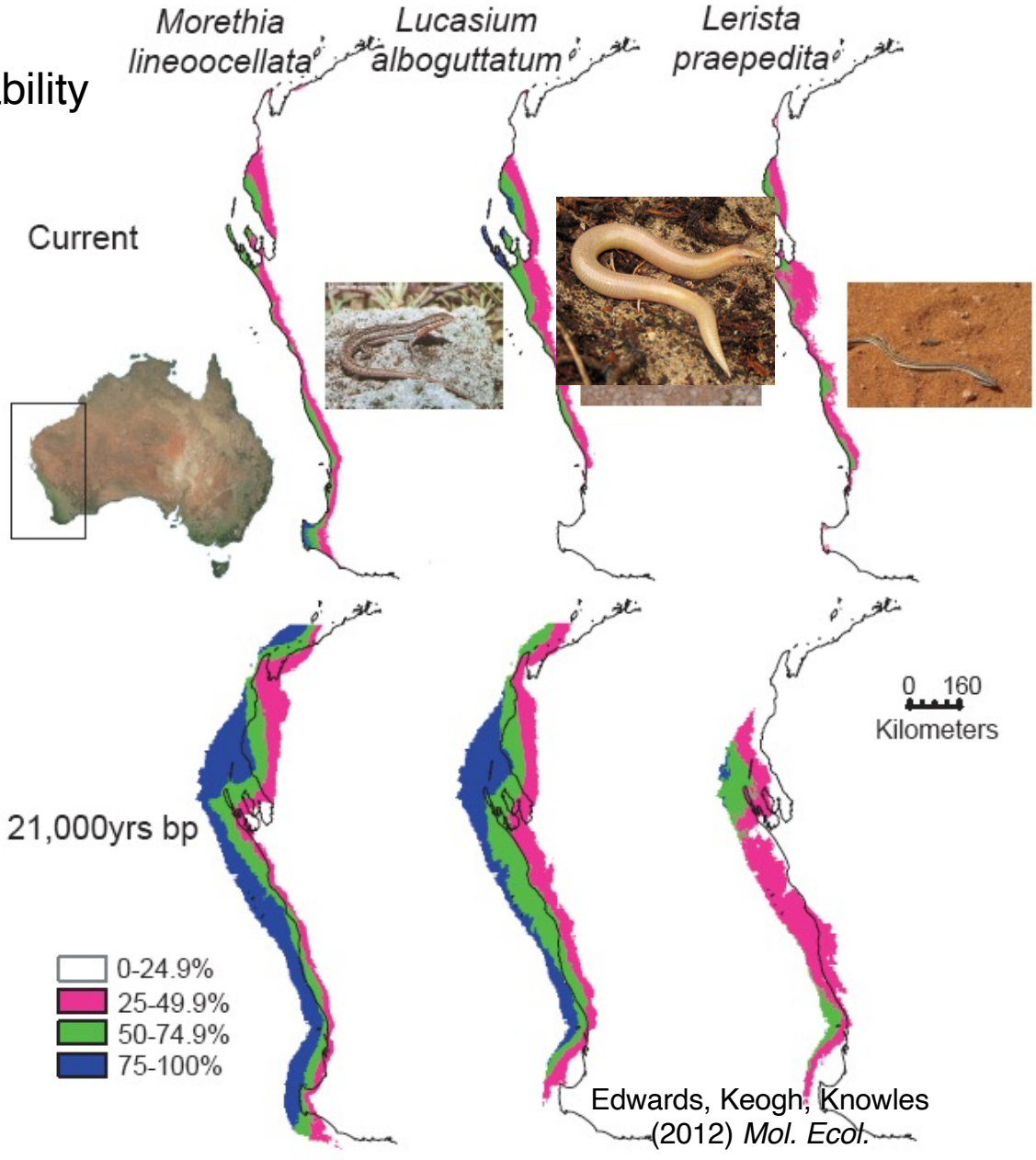
Species vary in their specialization to sand-dunes, suggesting habitat differences across space may be important in structuring patterns of genetic variation

Edwards, Keogh, Knowles (2012) *Mol. Ecol.*

Climatic conditions have changed over time

Current distribution (contemporary climatic data)

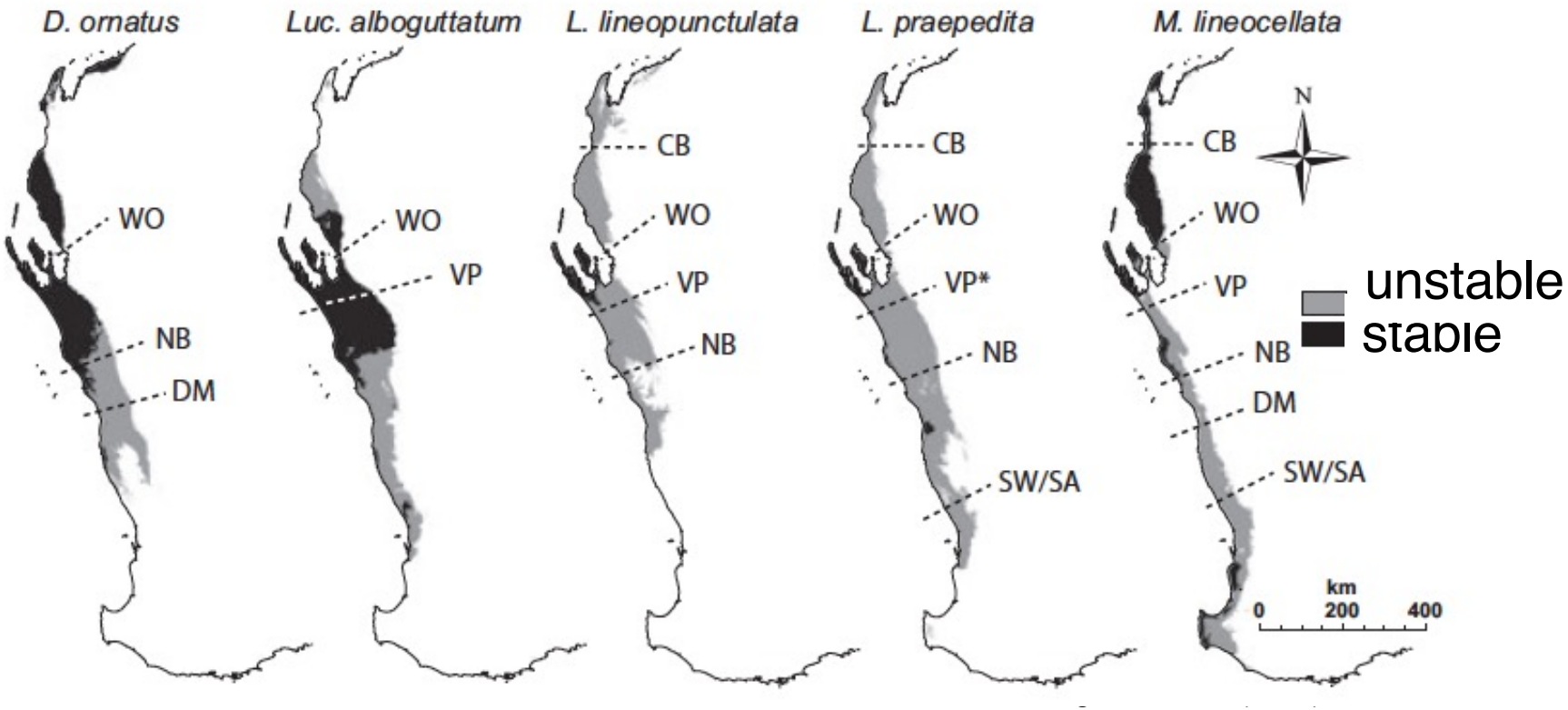
Habitat suitability



Past distribution (based on paleoclimatic data 21,000 yrs bp)

(based on ecological-niche models, ENMs, with MAXENT)

Edwards, Keogh, Knowles (2012) *Mol. Ecol.*



Differences in habitat stability due to climate-induced distributional shifts

Edwards et al. (2012) *Mol. Ecol.*

Spatially explicit coalescent model to capture movement across space and the different models are inspired by biology/geologic independent information

integrative
Distributional
Demographic
Coalescent
modeling

iDDC:

Distributional model
(i.e., ecological niche model)

⇩

Demographic model

⇩

Coalescent model

Hypotheses

- geographic isolation alone (IBD)

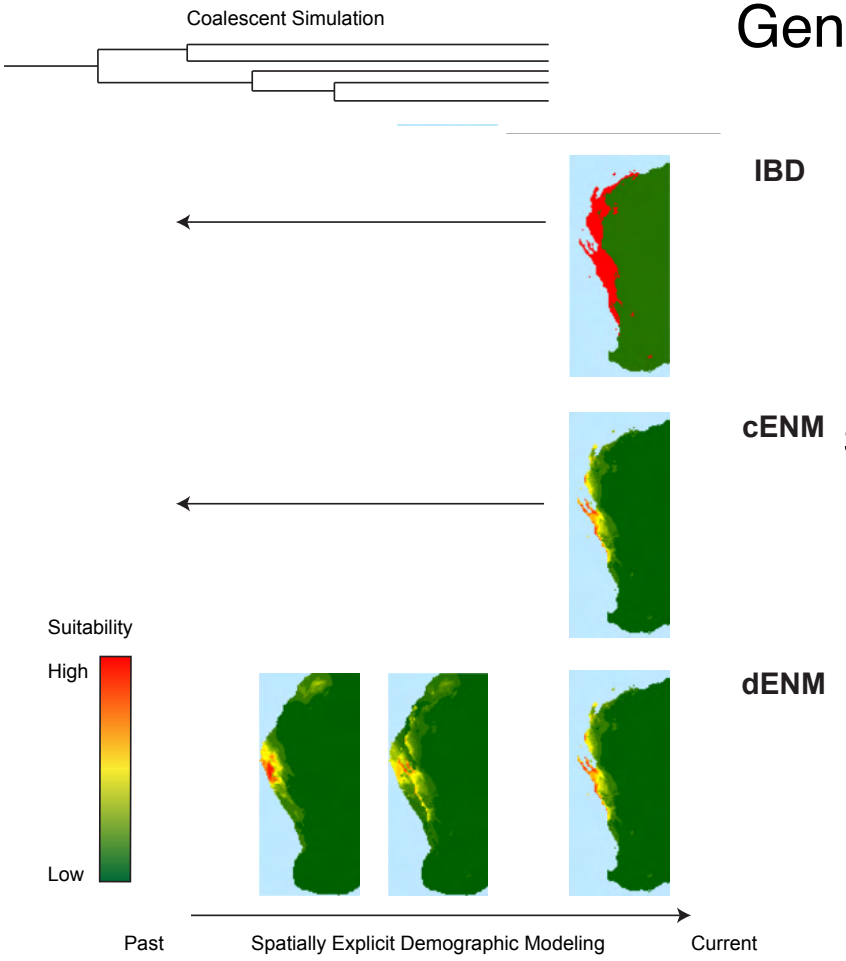
- population connectivity determined by current landscape, as measured from ENM

- population connectivity determined by distributional shifts associated with climate change, as modeled by current and paleoclimatic data

24 anonymous nuclear loci from 89 individuals sampled across the range of *Lerista* (shown by dots)

He, Edwards & Knowles (2013) *Evolution*

iDDC modeling:



Generate lots of simulated data sets under each model (IBD, cENM, dENM).

We can identify sets of parameters for specific models that produce simulated data that matches the empirical data.

Model Selection using Approximate Bayesian Computation (ABC)

He, Edwards & Knowles (2013) *Evolution*

Tests of hypotheses/models using ABC

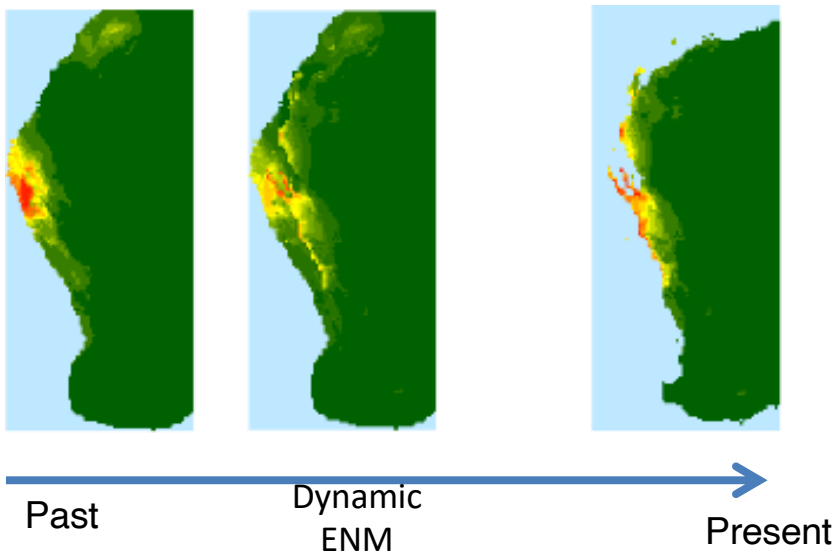
Choose model → Setting priors for parameters → Simulations: demographic -> Coalescent -> DNA → Calculate summary statistics → Retain simulations whose SS are close to empirical ones → Model Selection

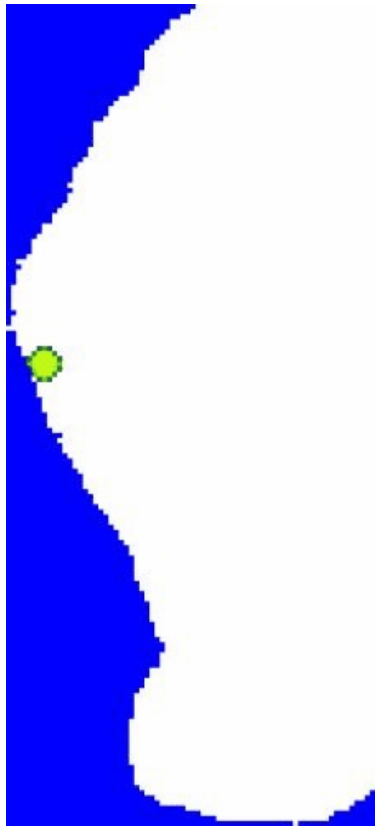
Comparison of Bayes factor showed that

Colonization by dynamic ENM

>> Isolation by contemporary ENM

> Isolation by distance



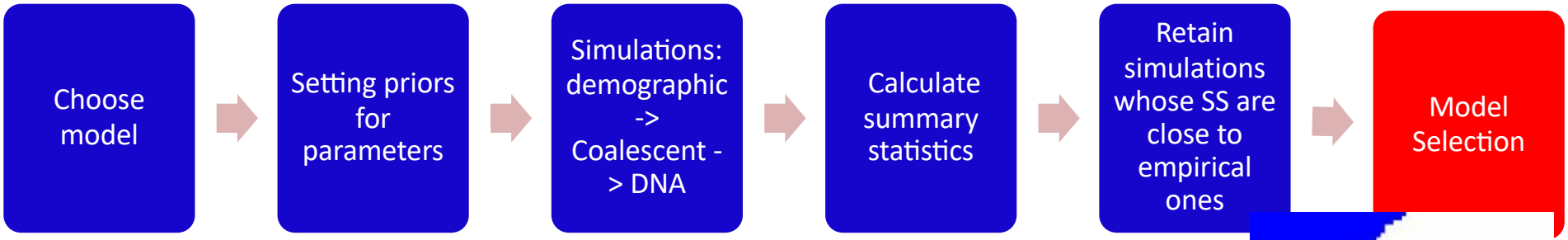


- Start from LGM refugia
- Colonize with changing layers of ENM

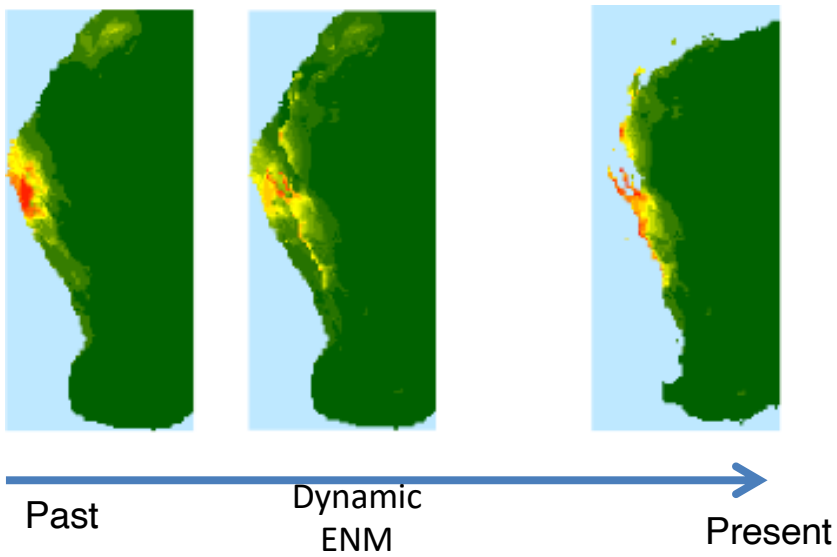
He, Edwards & Knowles (2013) *Evolution*

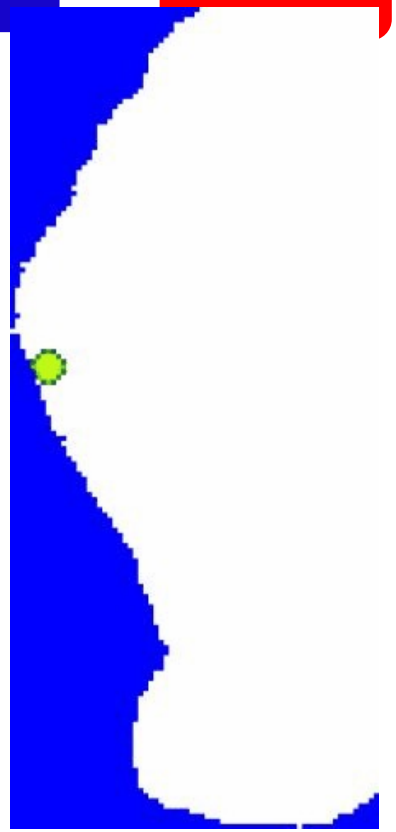
Is the most likely model a crappy model?

Compare the likelihood of the empirical data (i.e., the observation) under the model with the likelihoods of retained simulations under the model



Colonization by dynamic ENM





- Start from LGM refugia
- Colonize with changing layers of ENM

He, Edwards & Knowles (2013) *Evolution*

Advantages of iDDC:

- Flexible (expand to multiple species)
- Complex history
- Test of different historical processes
- Model verifications for ABC, e.g.:

- 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 higher likelihood than the observed data)

- Compute the coefficient of variation of each parameter explained by each PLSs of the summary statistics as an indicator for the power of the estimation

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

Challenges:

- iDDC is computationally intensive

Evolutionary applications of genomic data

what I'll emphasize:

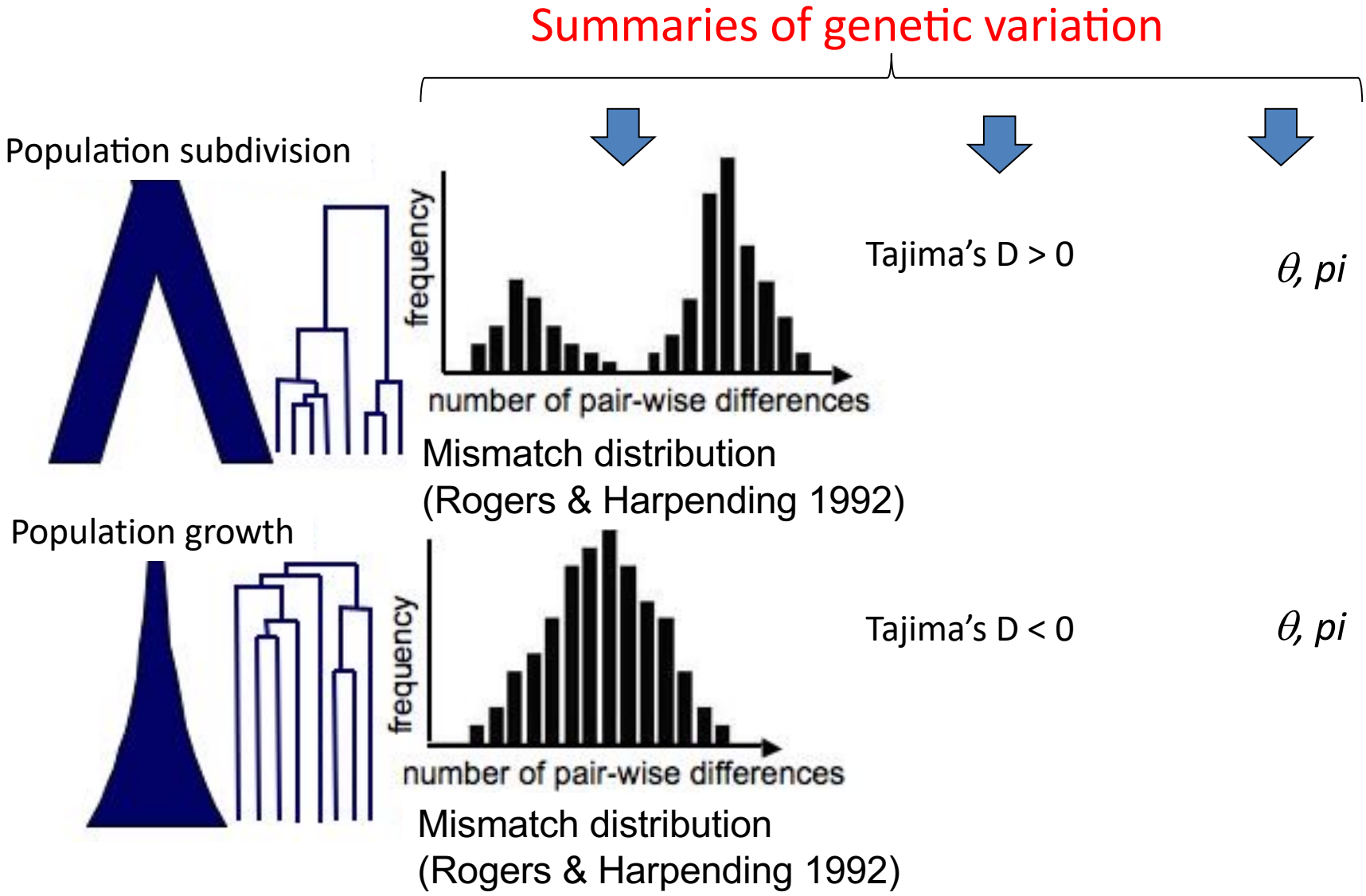
- Decisions/choices we make about model formulation

- Recognizing the subjectivity of model formulation itself when making inferences

- Decisions when applying to empirical data
(e.g., all the data, subset of data, what subset of data)

- Decide how to extract information from genetic data

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



Decisions about how to extract information from genetic data

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

- observed quantities are compared to expectations

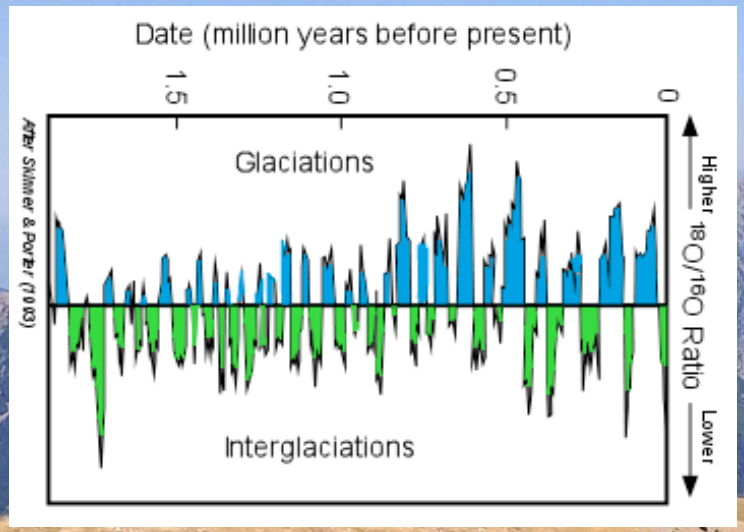
$\Rightarrow$ calculate full likelihood of the sequence data (computationally demanding, and may not work for complex models, but makes full use of the data)

Understanding the effects of rapid climate change on species diversity:

Did the frequent and repeated shifts in species distribution promote or inhibit divergence?

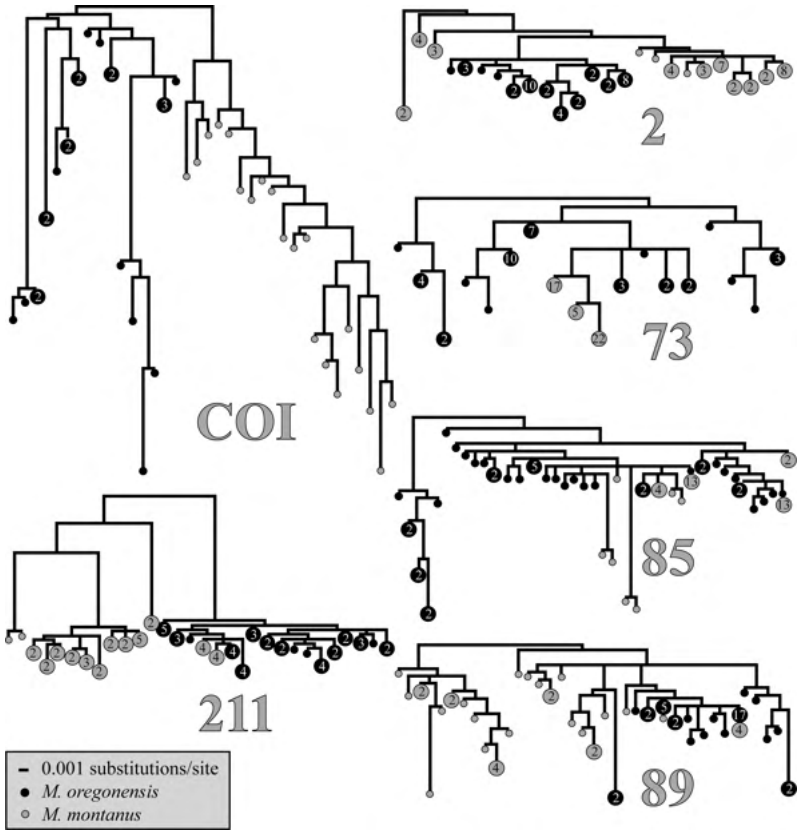
What is the timing of divergence?

- Pleistocene versus pre-Pleistocene?
- Glacial versus inter-glacial?



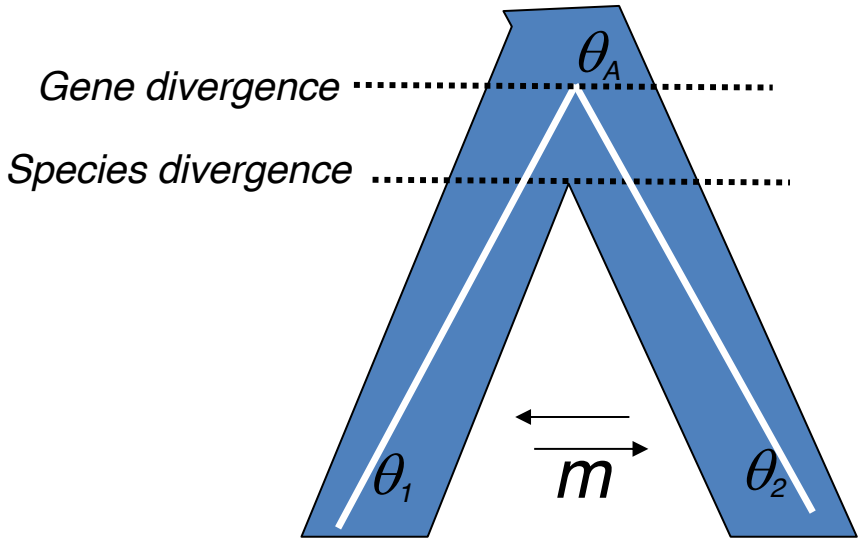
rstens & Knowles 2007, Mol. Ecol. 16:619-27.

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



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

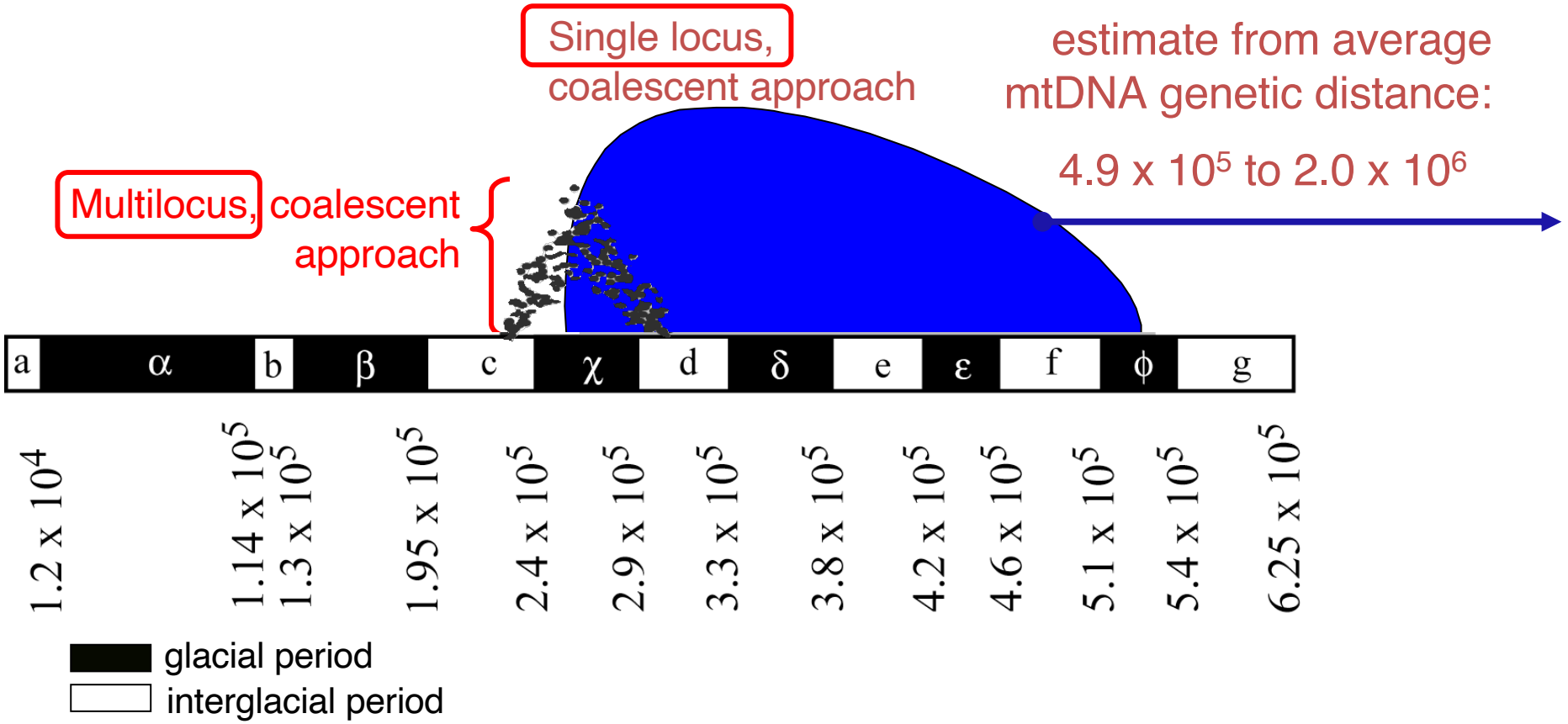
(Bayesian program IM)



Timing of divergence?

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

Precise estimate of T suggests species diverged during a glacial period

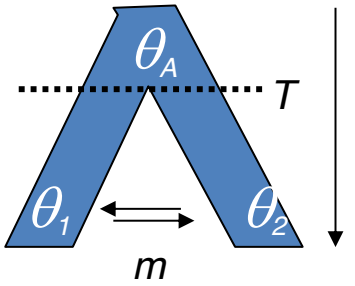


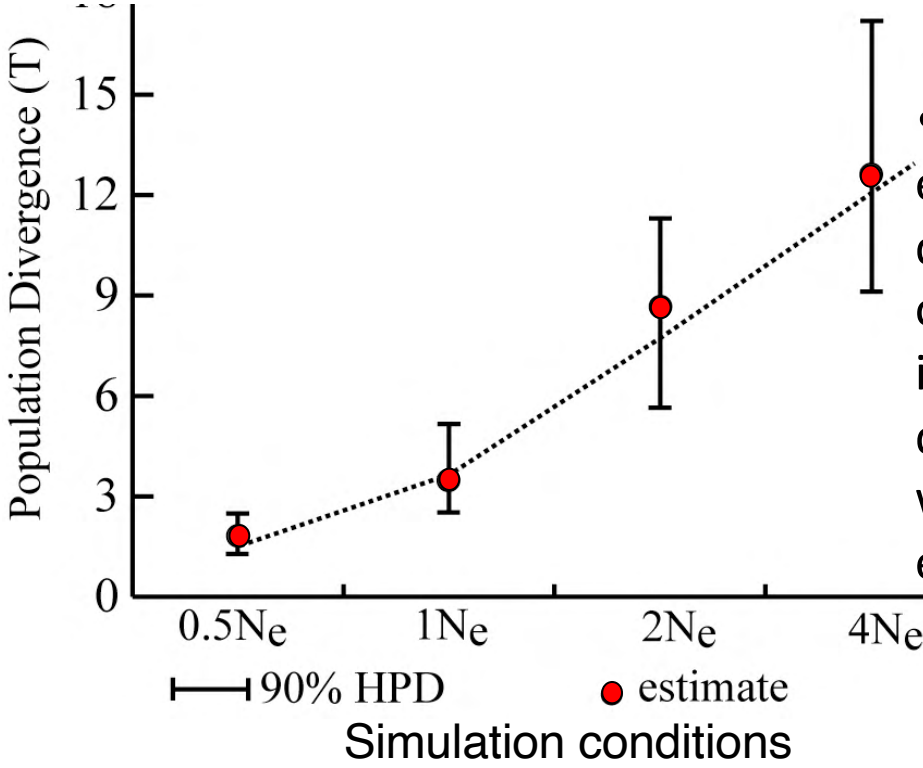
*same mutation rate used in the different approaches

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

Verified the accuracy of the speciation model given the data (only 6 loci)

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





- Simulate genetic data under models of evolution matching the empirical grasshopper data (i.e., same number of loci and same level of genetic diversity) and ask whether the inferred divergence time matches the divergence time used to simulate the data, where the parameter used in the model were estimated from the empirical data

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

How do we decide upon a model*:

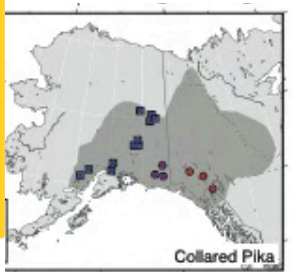
- informed from information independent of the genetic data itself
  – that is, a specific biological narrative motivates the model

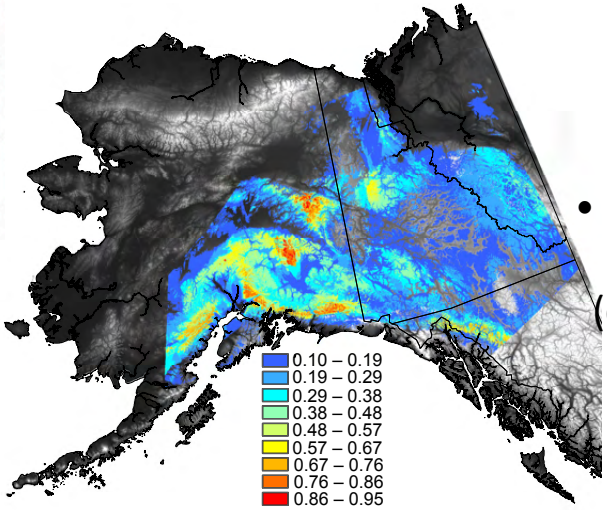
- models informed by the genetic data (…but be careful not to use same data twice)

- arbitrary/generic models

* All models are simplifications, and vary in the degree of their relative degree of abstraction

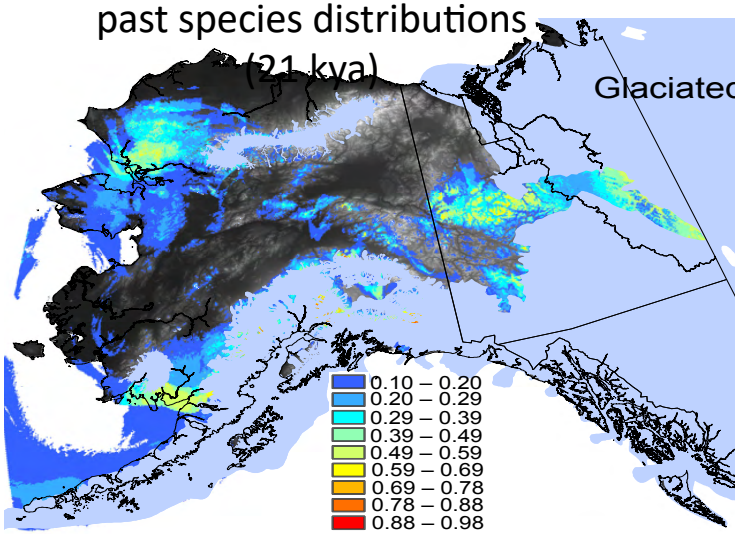
Informing model based on preliminary tests based on genetic data





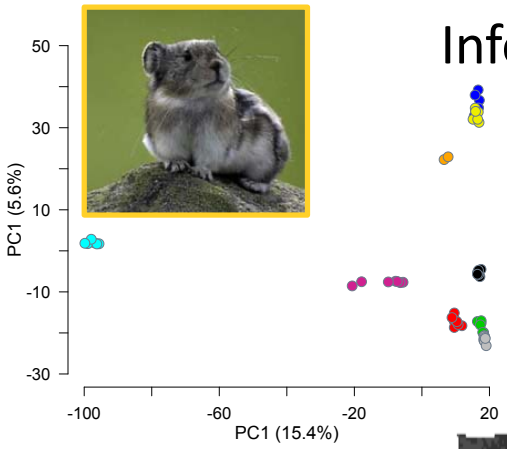
- Projected distribution from MAXENT based on contemporary bioclimatic variables (e.g., max and minimum temperatures and precipitation, etc)

- Paleoclimatic data used to model past species distributions (21 kya)



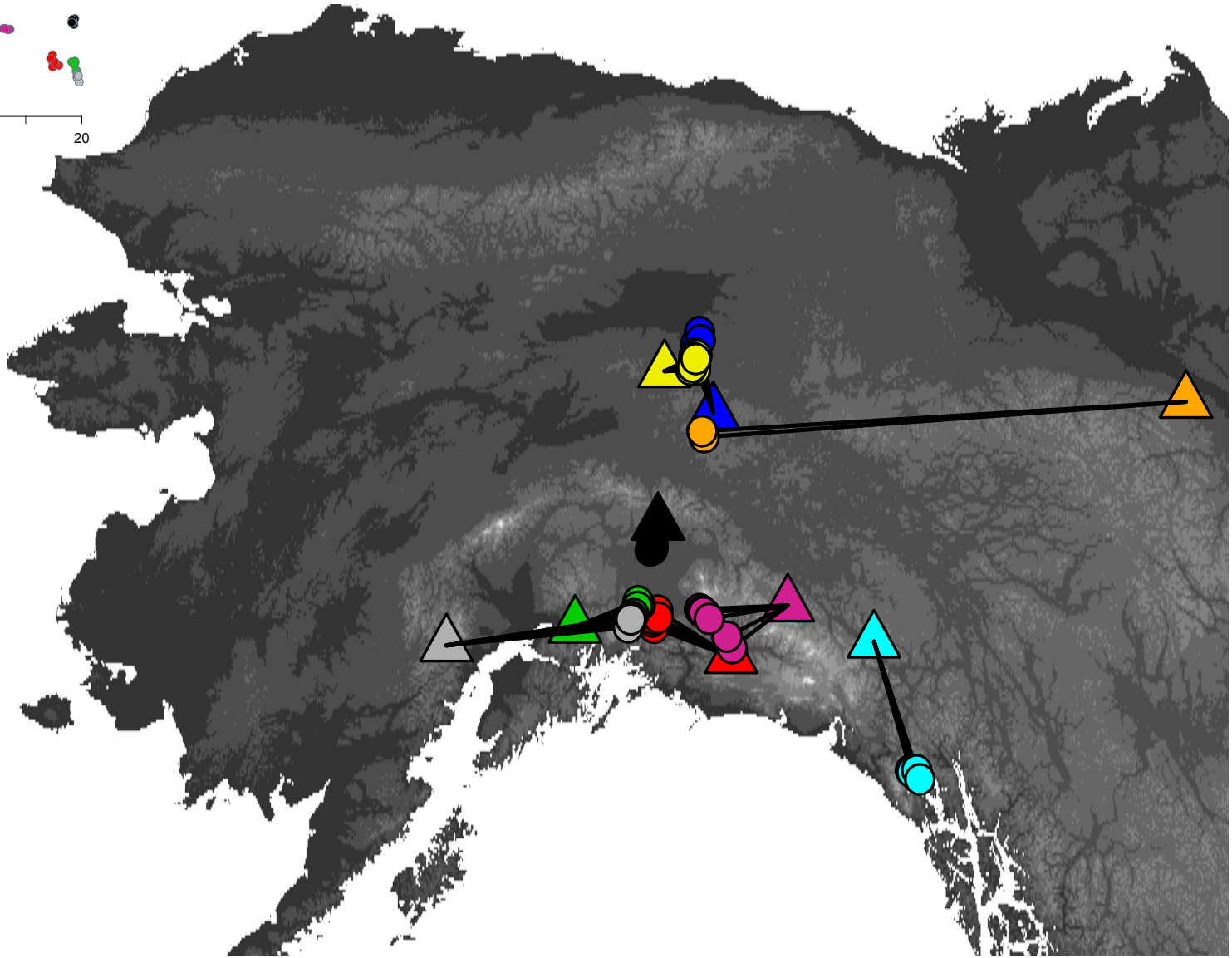
Sometimes ENMs not sufficient to define a model

Lanier et al. (2015) *Mol. Ecol.* 24:3688-3705

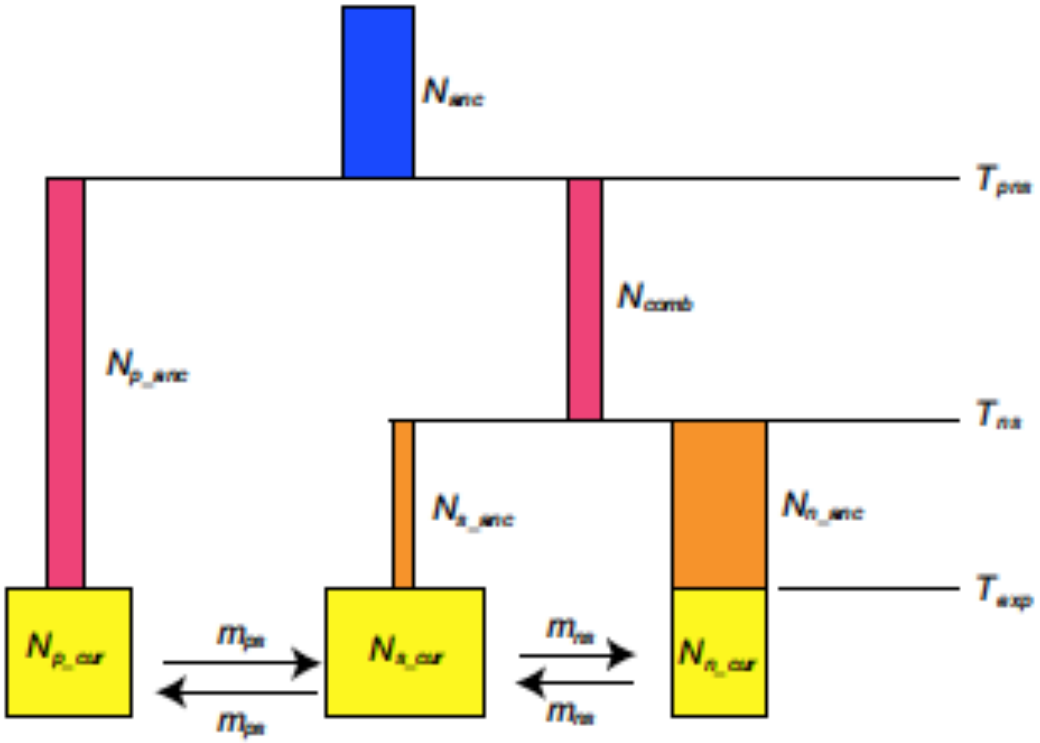


Informing model based on preliminary tests of genetic data:

Procrustes analysis (association between genotypes and geography in two dimensions)



To better understand the historical demographic trends for pika populations, we estimated divergence time, gene flow and population size changes among different populations using the site-frequency spectrum (SFS) using FastSimCoal.



10,892 variable SNPs

Fig. 2 Hypothesized demographic history of pika populations used in FASTSIMCOAL2 analyses. Pika ancestors diverged (Tpns generations ago) into ancestral populations of Pika Camp (Np_anc) and the other populations (Nns). Later, the divergence into southern (Ns_anc) and northern refugia (Nn_anc) occurred, and populations experienced recent expansions and exchanged migrants. The estimates of these parameters are listed in Table 4.

Lanier et al. (2015) *Mol. Ecol.* 24:3688-3705

• Our results indicate that contemporary factors alone (i.e., current habitat continuity and glacial corridors) are not sufficient to explain connectivity among populations of Collared Pikas across their range

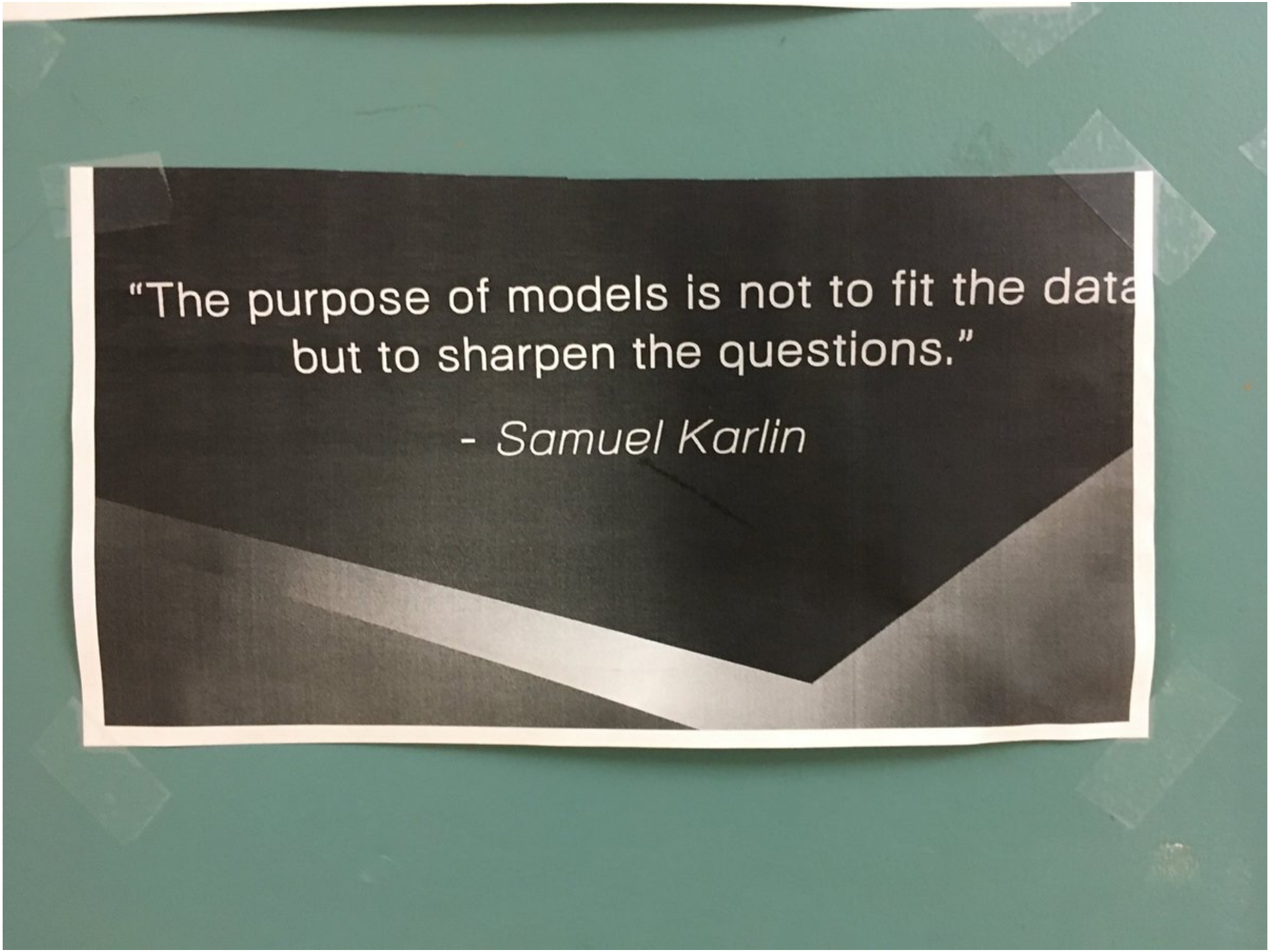
• Instead, the results provide strong support for the predominance of three divergent lineages, likely separated in different Pleistocene refugia

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL (2015) Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genetic variation in Collared Pikas (Ochotona collaris). Mol. Ecol. 24:3688-3705.

How do we know if we have the “right” model?

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

(i.e., YOUR knowledge about a biological system is key!)


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

Evolutionary applications of genomic data

- Accounting for species-specific traits
- Spatially explicit coalescent models
- Comparative analyses of genetic variation across species

Evolutionary applications of genomic data

- Accounting for species-specific traits
- Spatially explicit coalescent models
- Comparative analyses of genetic variation across species

Evolutionary applications of genomic data

What I'll emphasize:

- Decisions/choices we make about model formulation
- Recognizing the subjectivity of model formulation itself when making inferences
- Decisions when applying to empirical data

(e.g., all the data, subset of data, what subset of data)

Does microhabitat differences lead to differences among species in their responses to climate change?

- start with descriptive analysis to explore hypotheses, and follow-up with spatially explicit models to test hypotheses about why patterns of genetic variation differ among species (i.e., generate species-specific patterns of genetic variation)

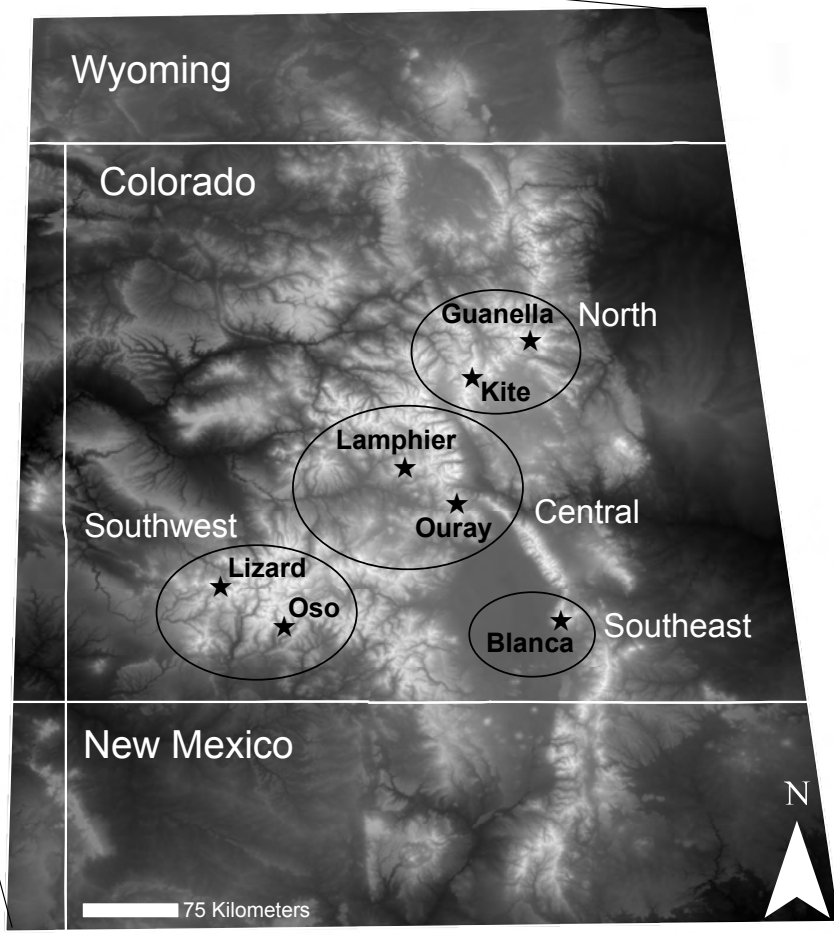
Knowledge of geologic history and natural history of the plants were key in formulating hypotheses!

Sky island community respond similarly to climate change? (use genetic tests and sampling design to evaluate this question)

Carex chalciolepis

Carex nova

Massatti & Knowles (2014 Evolution)



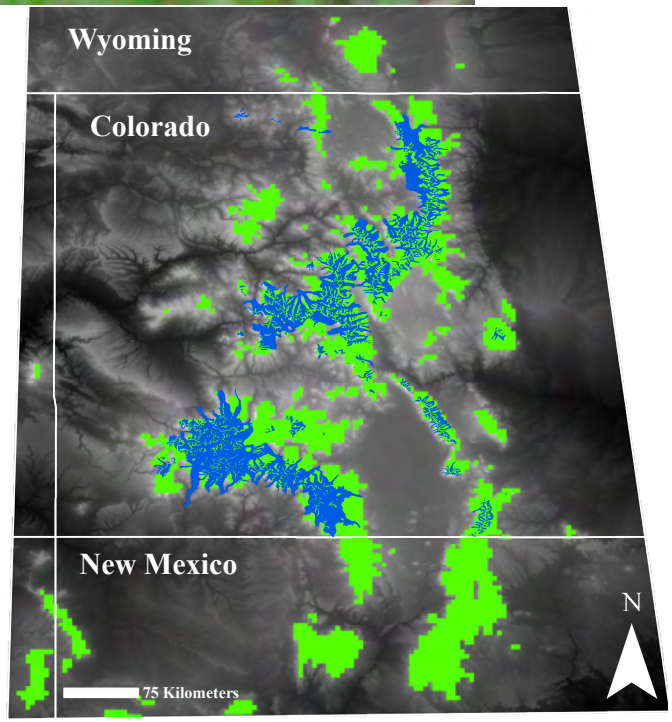
Sky island communities

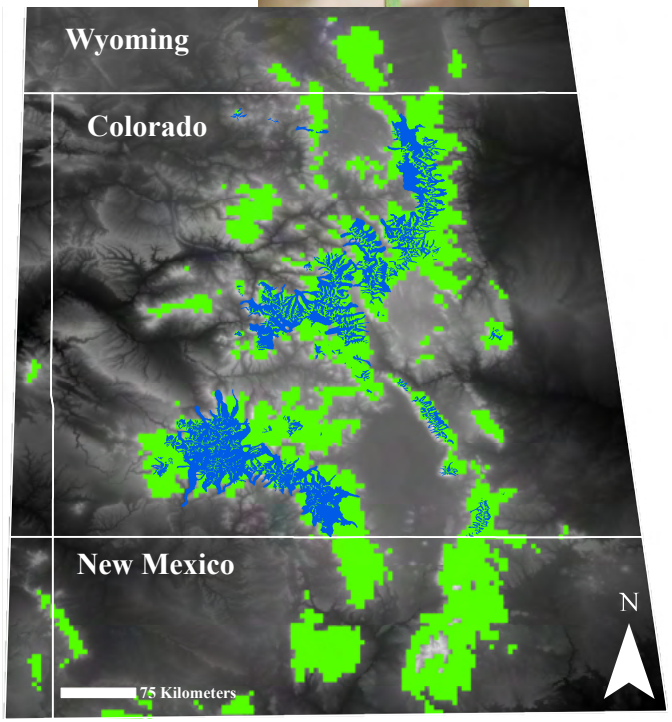
- co-distributed, abundant taxa with similar natural histories and dispersal abilities

Carex chalciolepis

C. nova

- so similar that ENMs project very similar past distributions

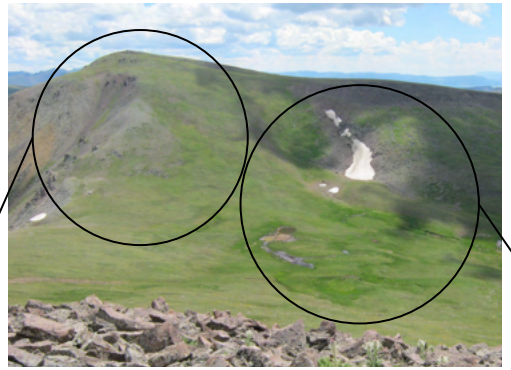




Taxa differ in microhabitats

inhabits slopes and ridges

Carex chalciolepis

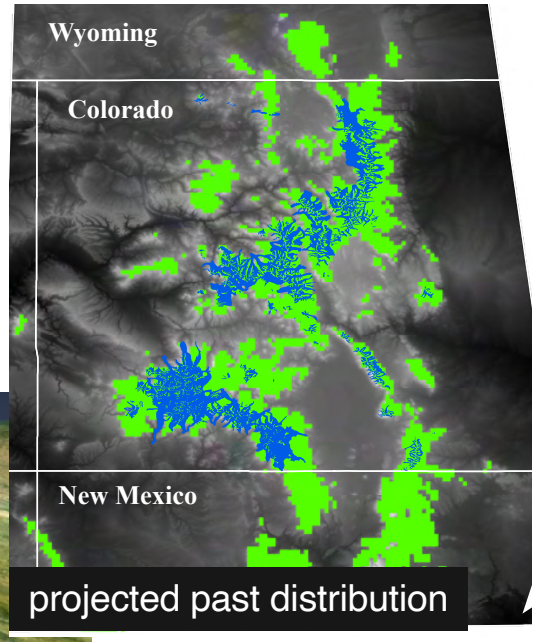


restricted to wetlands

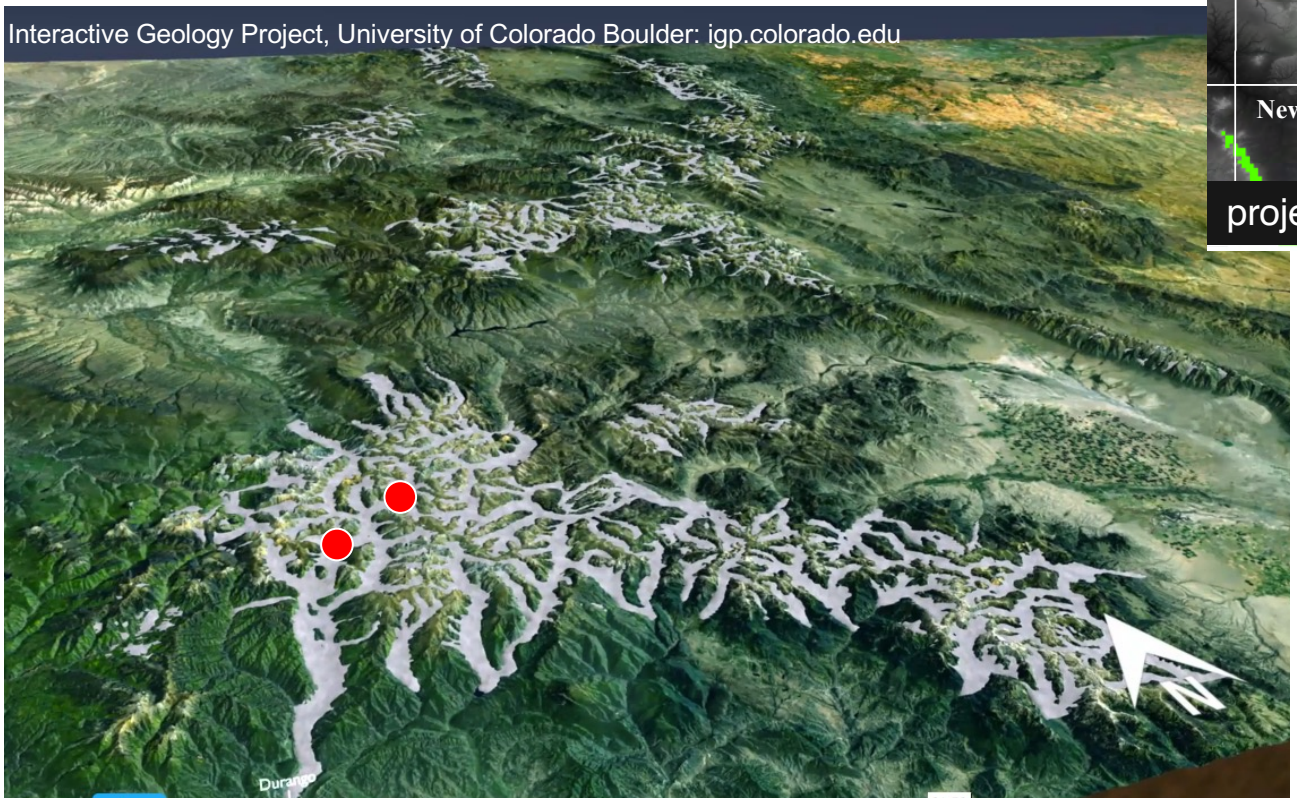
Carex nova

Given that ecological niche models (ENMs) are similar between species (both present and during LGM)…

why would we predict discord in patterns of genetic variation between the plant species?



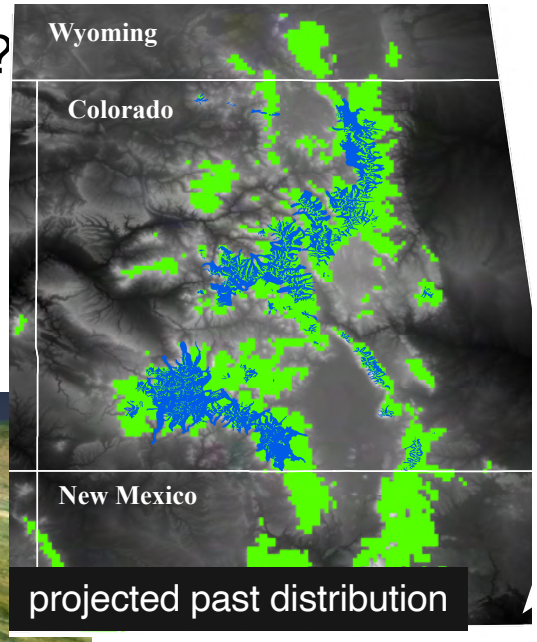
projected past distribution

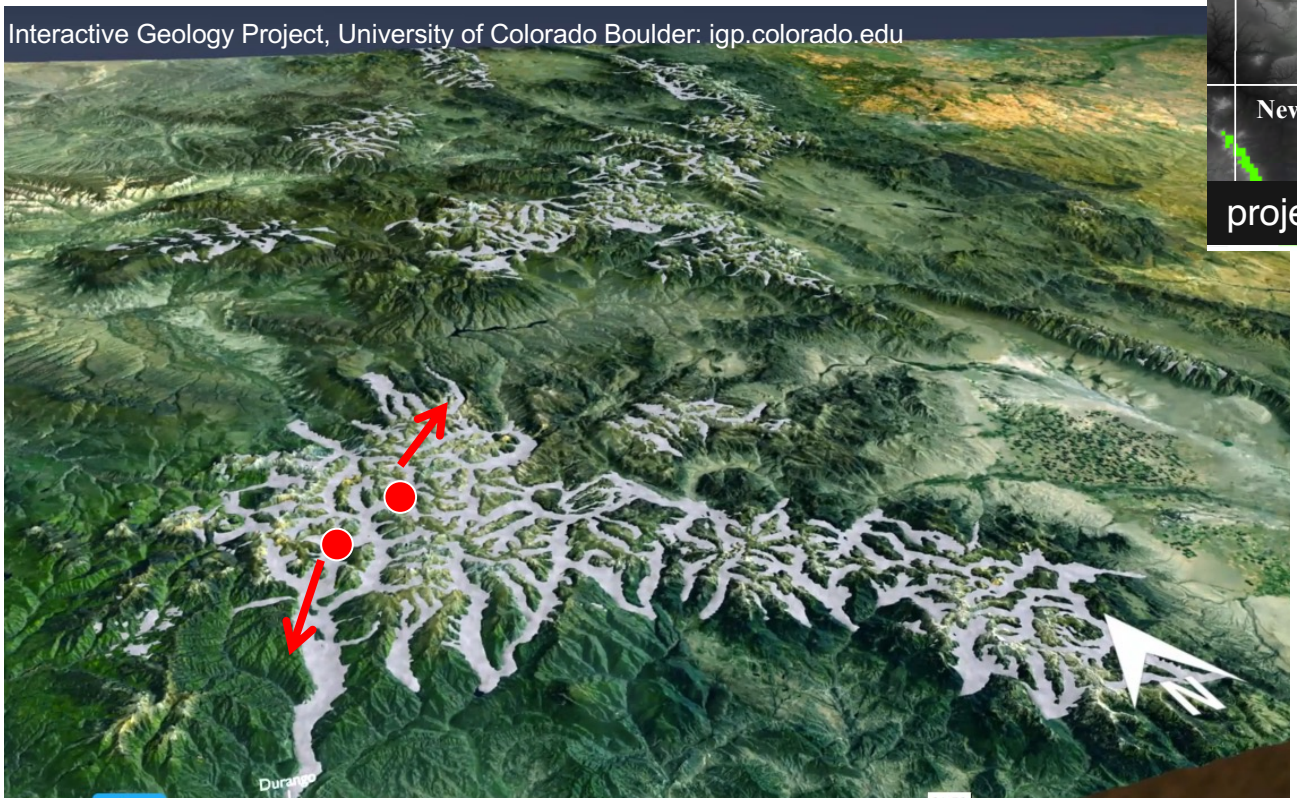


If microhabitat matters…

predict that glaciers in drainages would have displaced populations, but only in the wetland specialist

Why should microhabitat matter for sky island inhabitants?





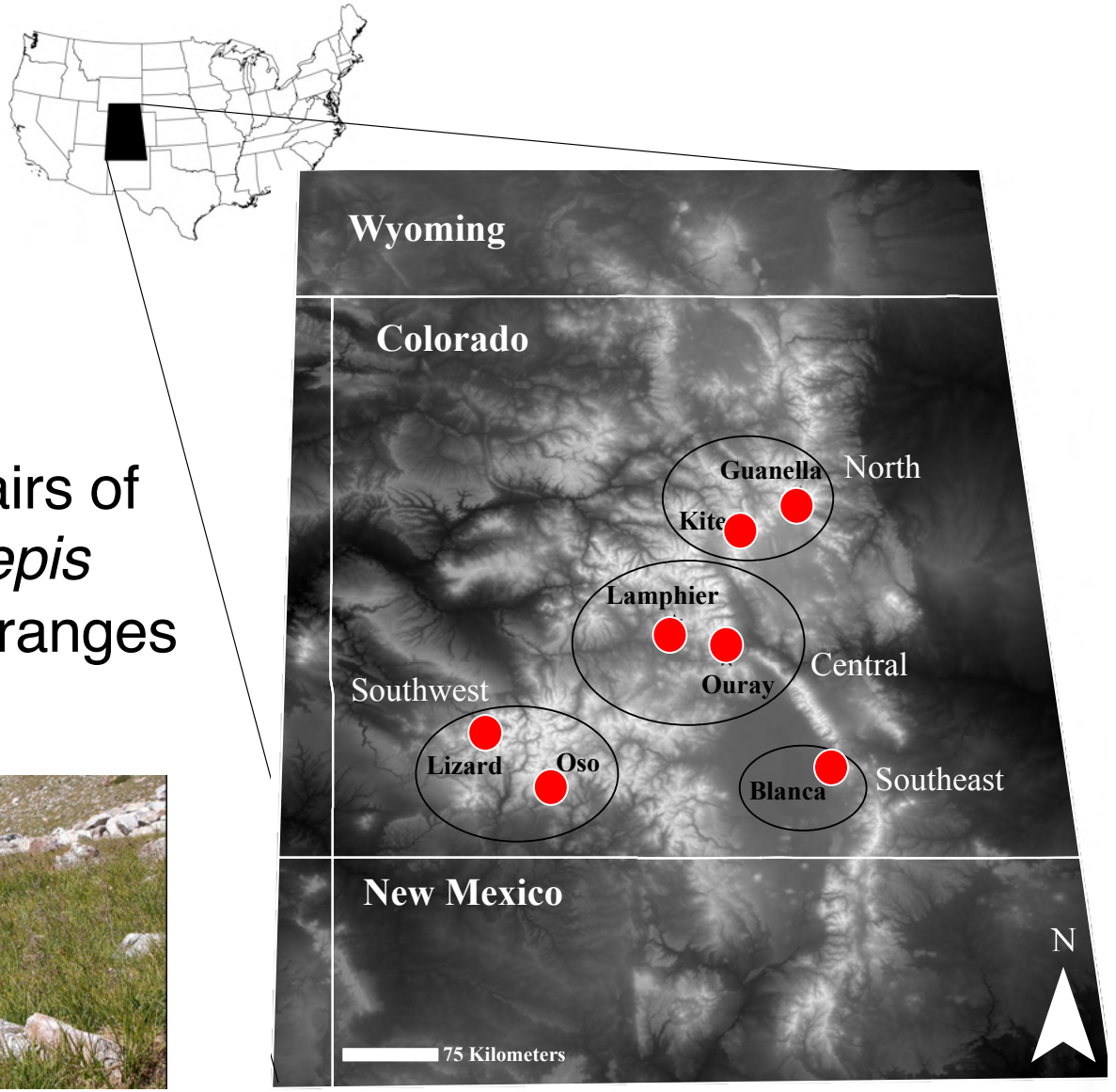
If microhabitat matters…

- distances separating populations may have been considerably greater in the past – *but only in the wetland specialist*

Sky island communities: microhabitat differences lead to differences in species responses to climate change

- SNPs from over 22,000 loci (RADseq)

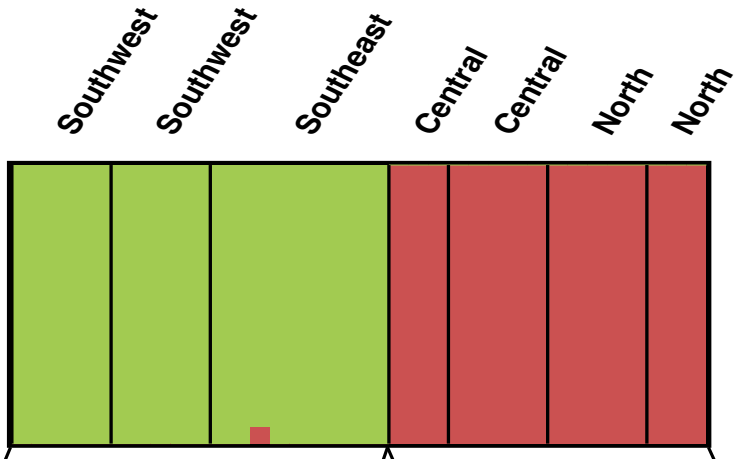
- sampled population pairs of *C. nova* and *C. chalciolepis* from different mountain ranges

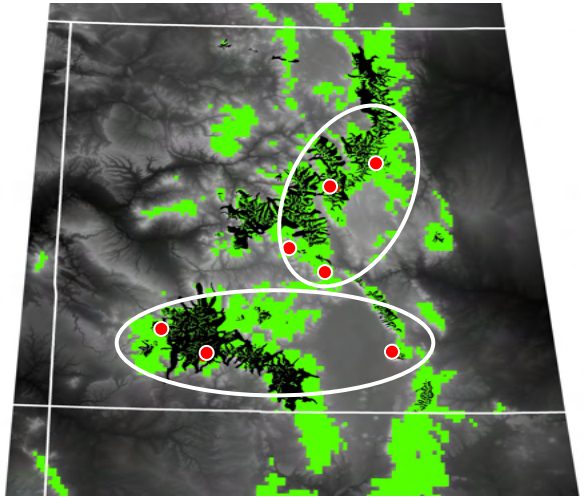


Massatti and Knowles, Evolution (in press)

C. nova

restricted to wetlands



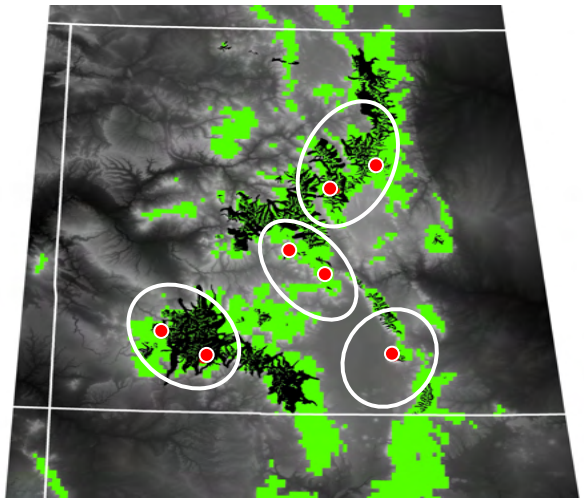


projected past distribution

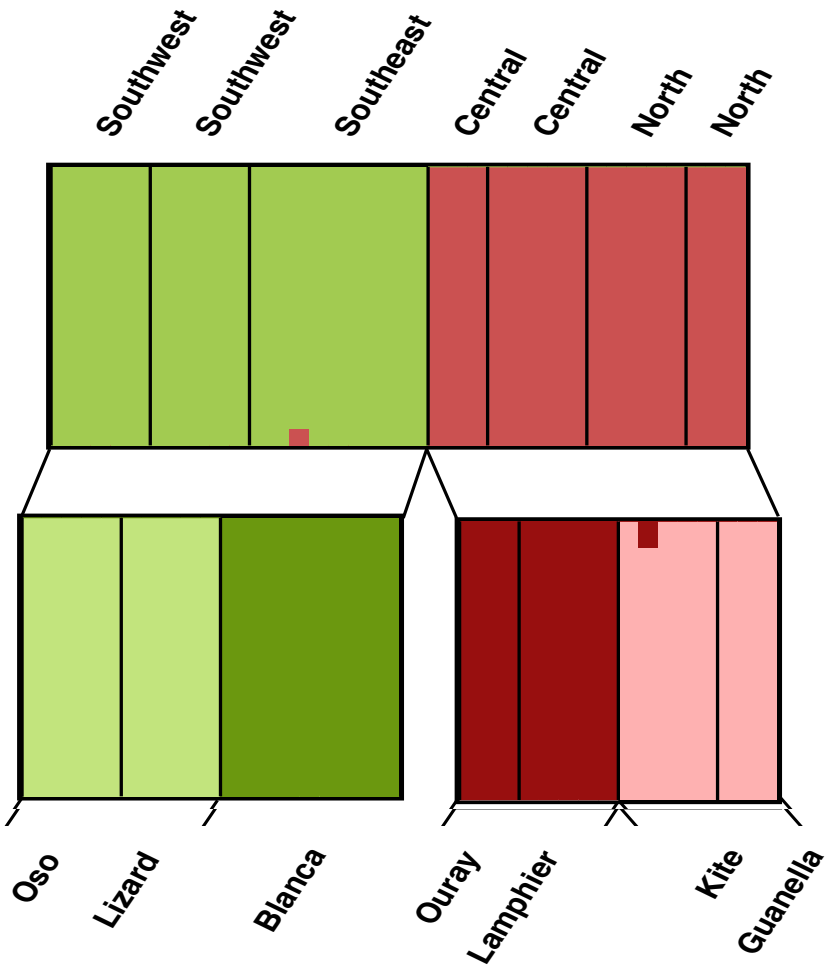
- Structure analysis of SNPs from over 22,000 loci

C. nova

restricted to wetlands



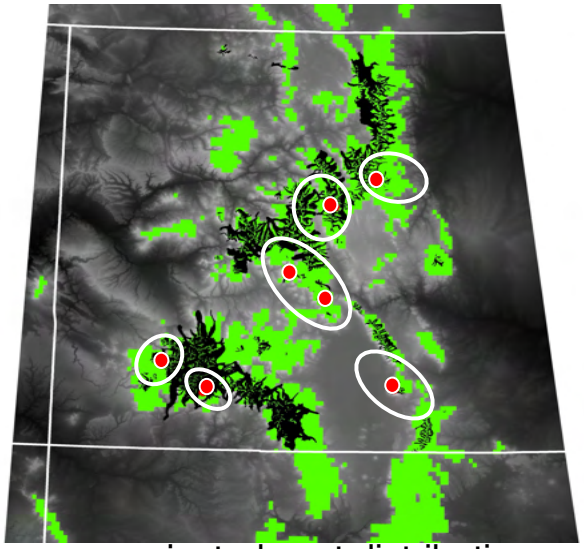
projected past distribution



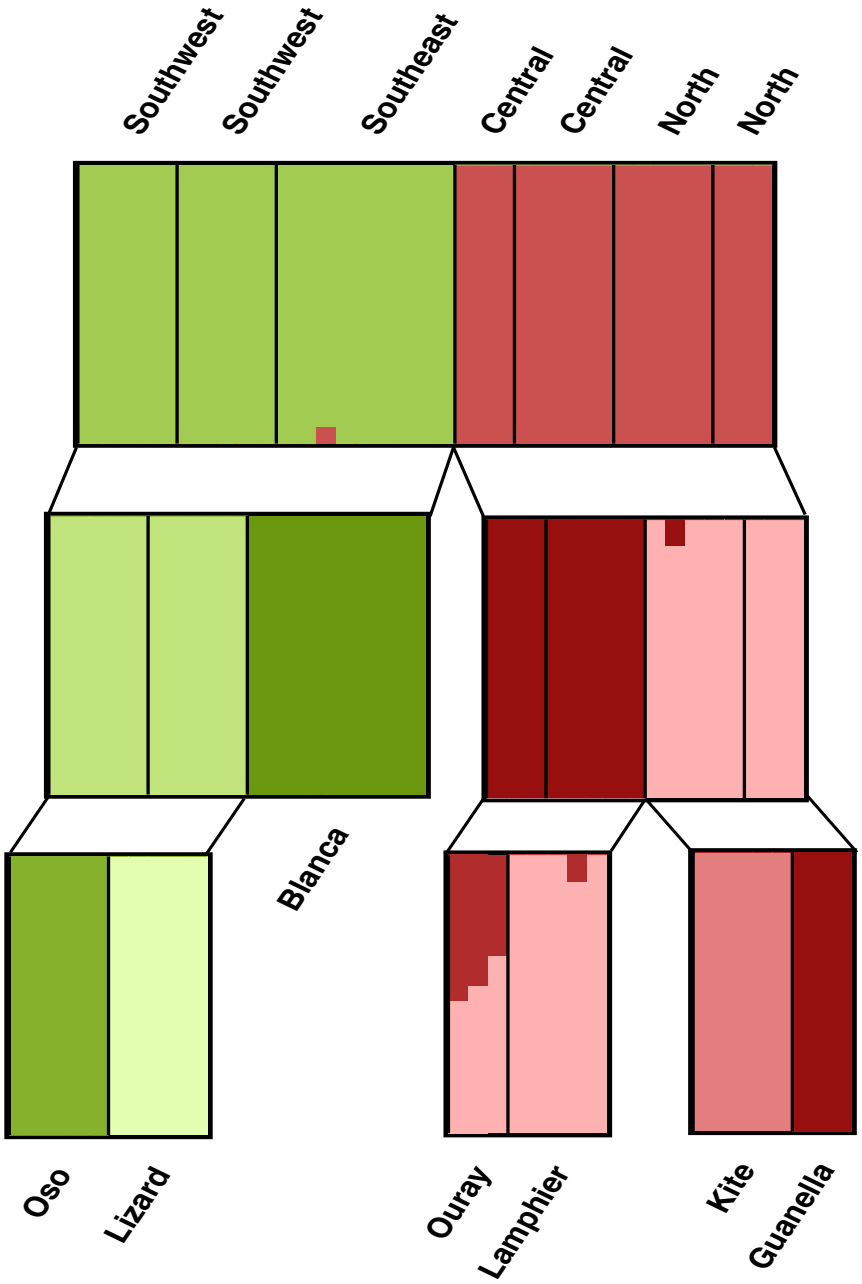
- Structure analysis of SNPs from over 22,000 loci

restricted to wetlands

C. nova



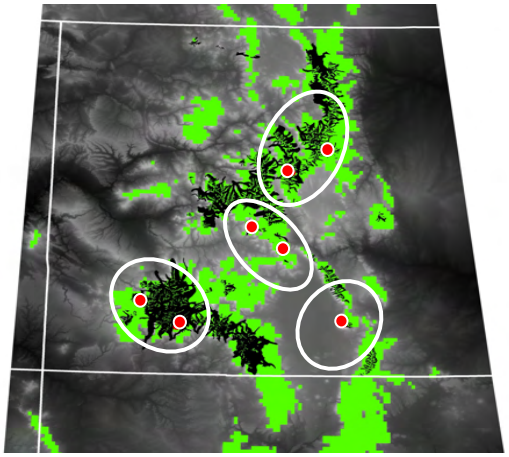
projected past distribution



- STRUCTURE analysis of SNPs from over 22,000 loci

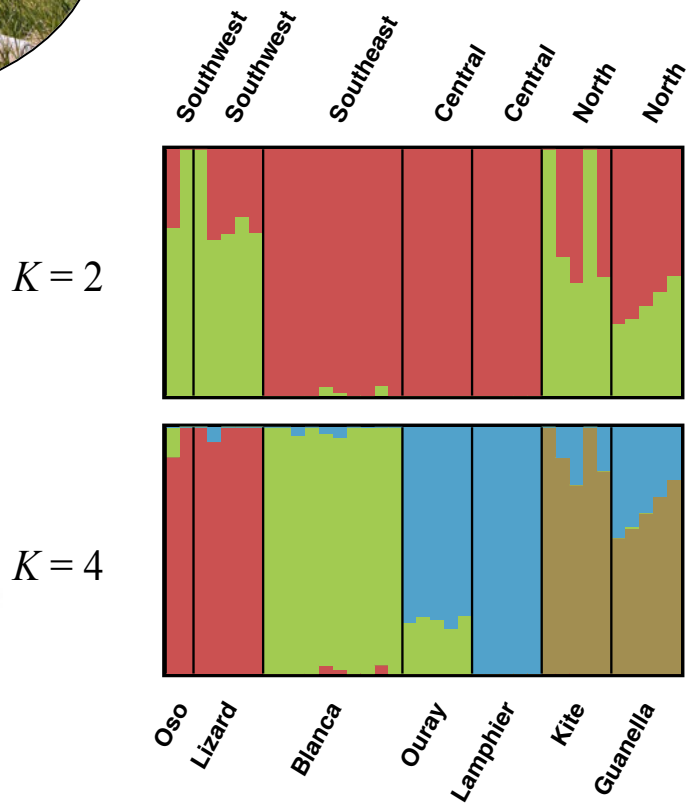
Massatti and Knowles, Evolution (in press)

inhabits slopes and ridges

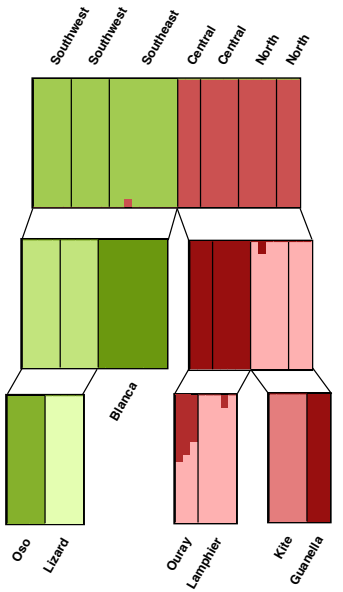


projected past distribution

C. chalciolepis



C. nova

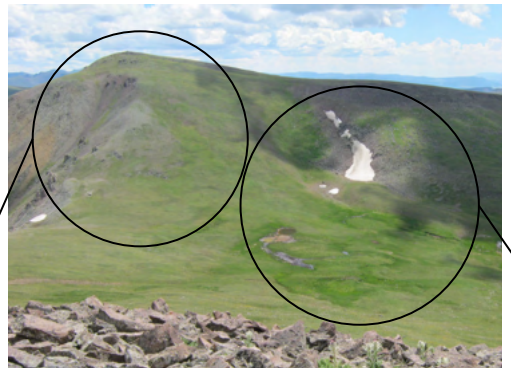


Massatti and Knowles, Evolution (in press)

Descriptive analyses support prediction that microhabitat mediates the response of species to climate change

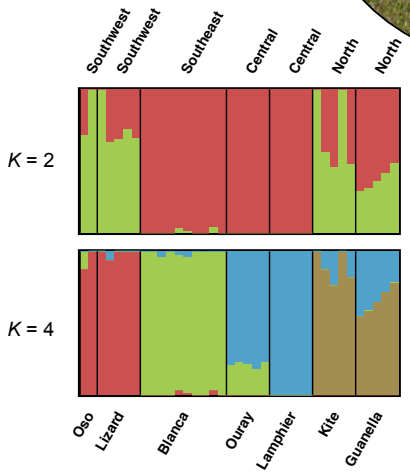
Carex chalciolepis

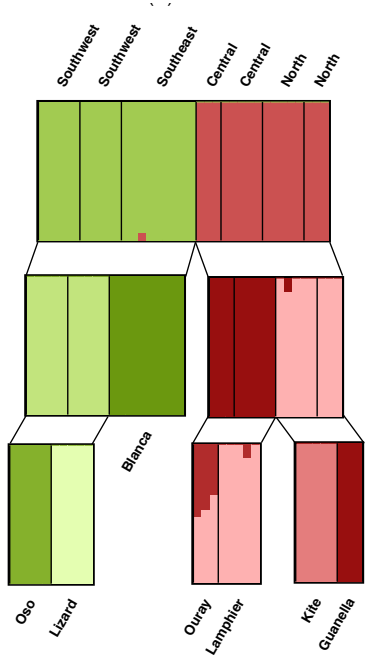
dry ridges



Carex nova

wetland specialists



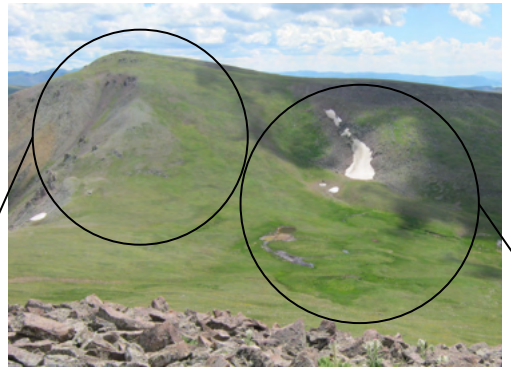


Massatti & Knowles (2014) *Evolution*

Test if observed discordant phylogeographic structure could be caused by differences in microhabitat affinity ….

Carex chalciolepis

dry ridges



Carex nova

wetland specialists

- generate species-specific expectations for patterns of genetic variation

(i.e., glaciers are barrier for movement of wetland specialists only)

iDDC: Generate species-specific expectations for patterns of genetic variation

He, Edwards & Knowles, Evolution 2013

integrative
Distributional
Demographic
Coalescent
modeling

Distributional model
(i.e., ecological niche model) with
predictions on probability of occurrence
across the landscape

⇩

Demographic model
informed by habitat
suitailities

⇩

Spatially-explicit coalescent
simulations based on
demographic model

⇩

Tests of hypotheses/models
using ABC

Habitat suitability scores

40	20	10	5
100	60	20	10
100	100	40	40
80	80	60	60

K(m)

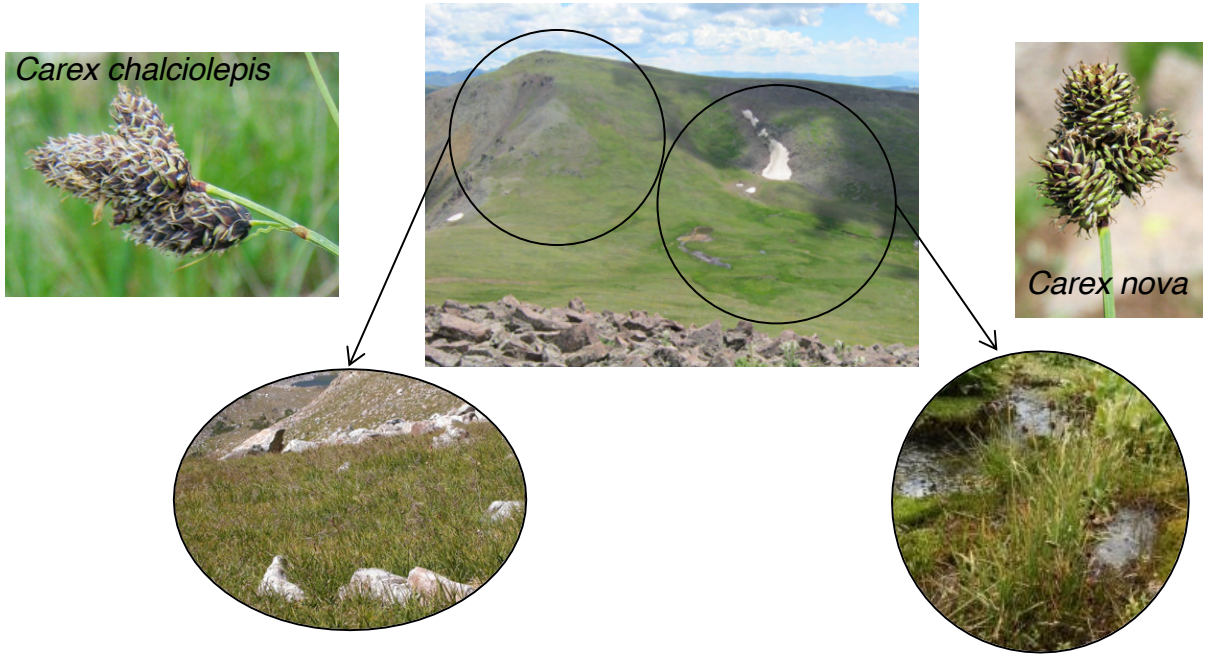
400 (40)	200 (20)	100 (10)	50 (5)
1000 (100)	600 (60)	200 (20)	100 (10)
1000 (100)	1000 (100)	400 (40)	400 (40)
800 (80)	800 (80)	600 (60)	600 (60)

Carrying capacity: k_i

Gene coalescence across the landscape

SPLATCHE2

iDDC: Generate species-specific expectations for patterns of genetic variation

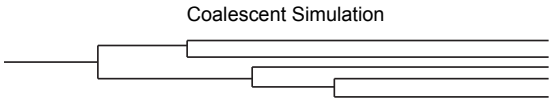


H: species-specific responses to climate change

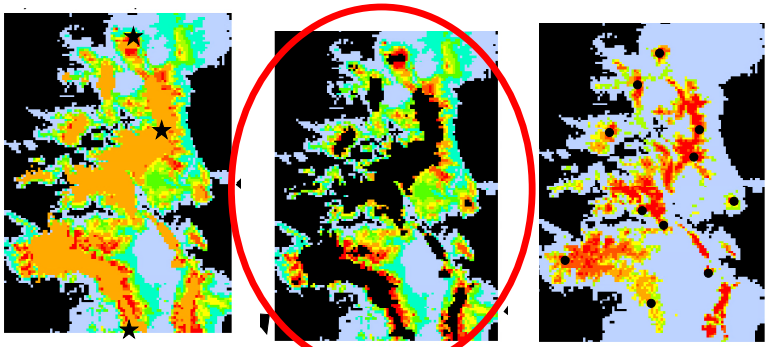
- Glaciated areas act as barriers,
  but only in wetland specialist

So genetic discord between species is not dismissed as reflecting idiosyncratic nature of history; genetic discord predicted from taxon-specific traits!

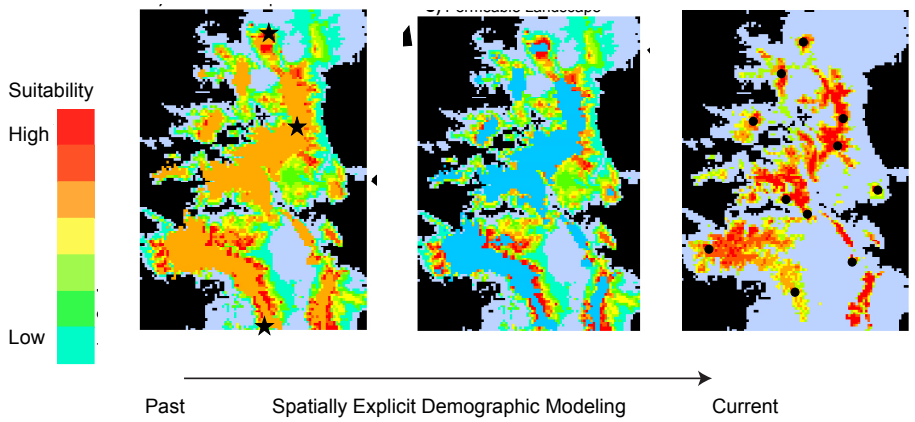
iDDC modeling:



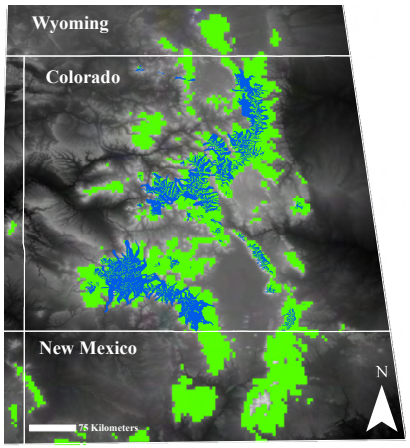
- Glaciated areas barrier



- Glaciated areas permeable



Glaciers shown in blue



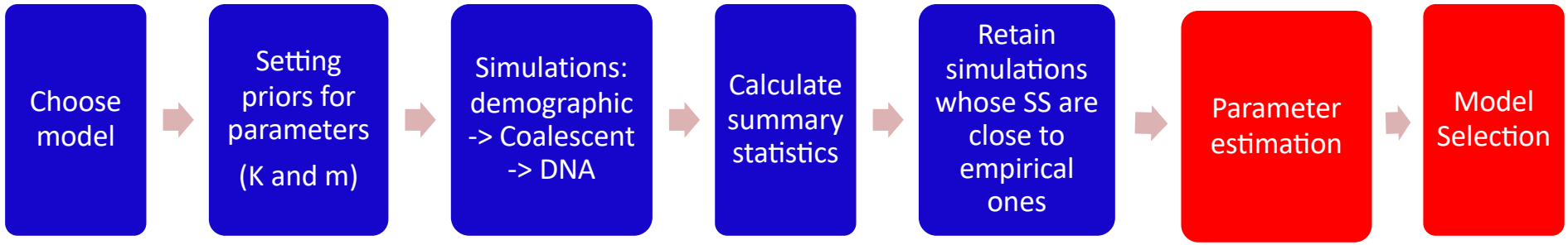
Generate lots of simulated data sets under each model

We identify sets of parameters for the models that produce simulated data that match the empirical data.

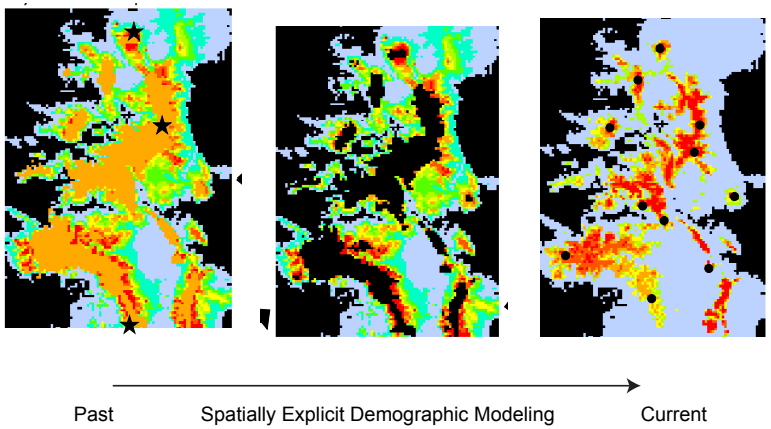
Model Selection using Approximate Bayesian Computation (ABC)

Massatti & Knowles (2016) Mol. Ecol.

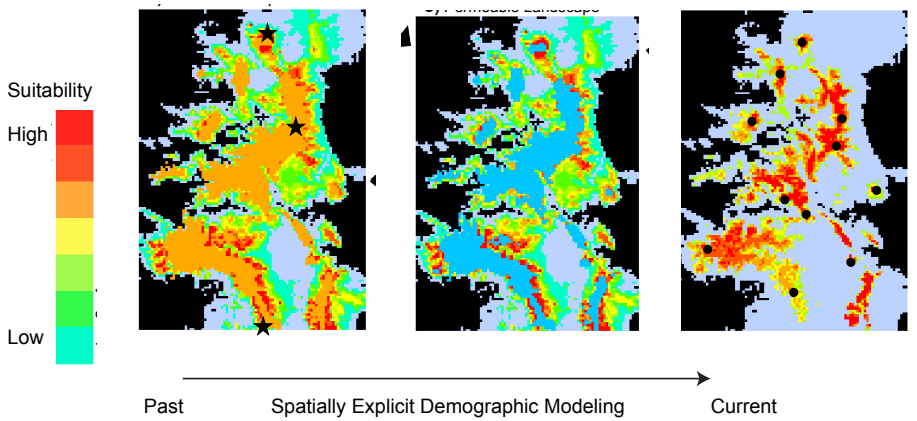
Tests of hypotheses/models using ABC



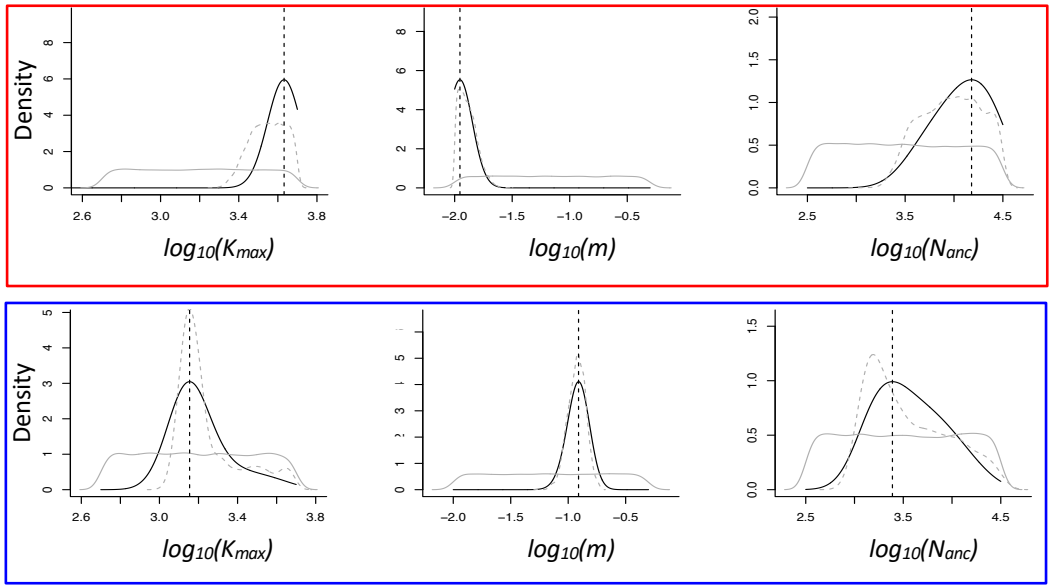
Model: Glaciated areas barrier



Model: Glaciated areas permeable



5000 simulations closest to empirical data retained for parameter estimation



Marginal densities:

Carex chalciolepis
Bayes factor ~3

Carex nova
Bayes factor ~23

	Barrier Model	Permeable Model
Carex chalciolepis	4.87×10^{-5} (0.65)	1.38×10^{-4} (0.97)
Carex nova	1.29×10^{-4} (0.84)	5.68×10^{-6} (0.08)

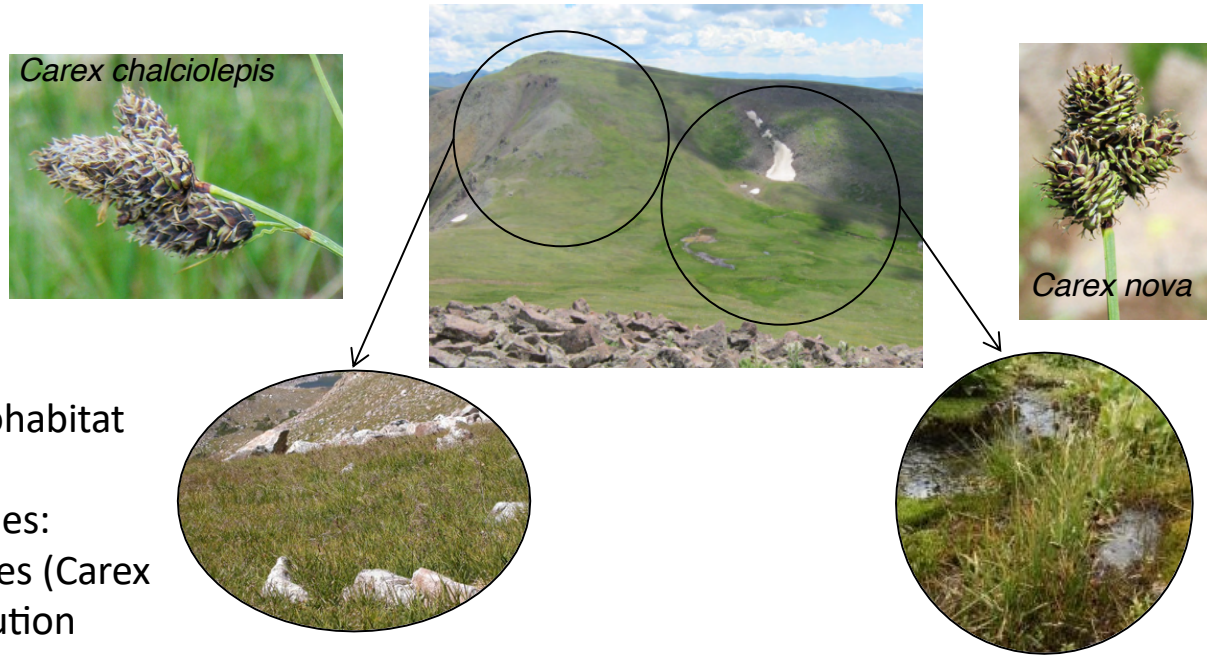
Is the most probable model capable of generating the observed data ?

(compare the L of retained simulated data sets to the L for the empirical data: *"P-value")*

Massatti & Knowles (2016) Mol. Ecol.

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



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.

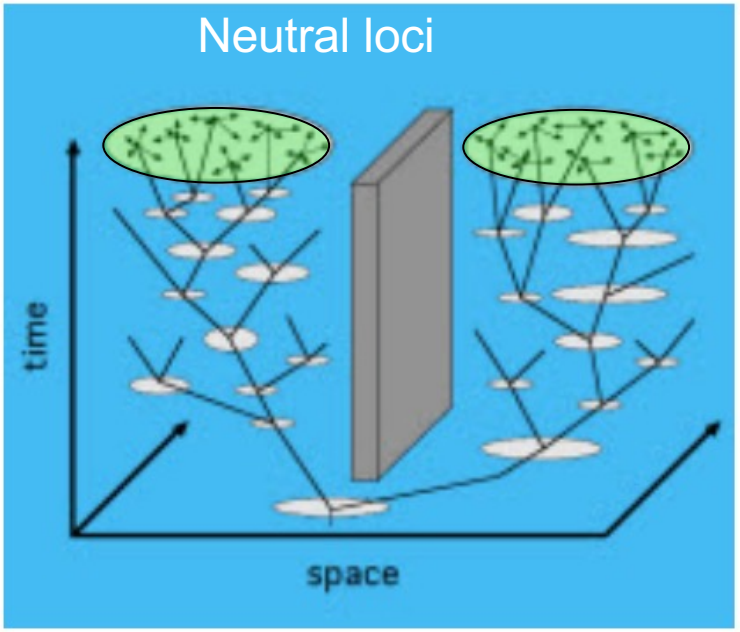
- Glaciated areas act as barriers, but only in wetland specialist

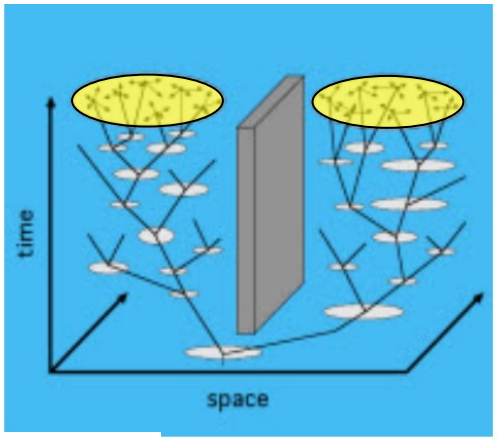

Communities may be characterized by
species-specific responses to climate change

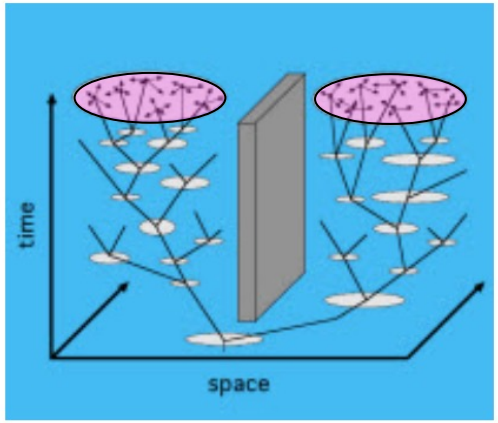
Community-level tests of similarity in history

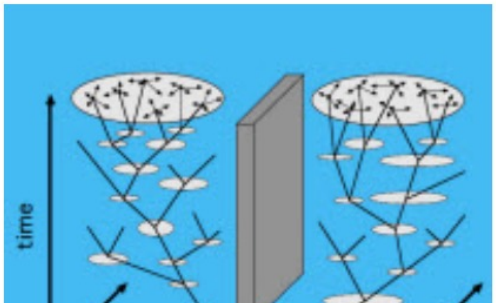
(Oaks et al., 2013; Oaks, 2014)

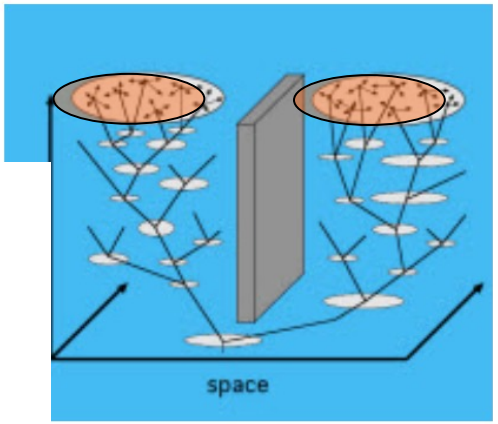
Genes and Geography Across Species





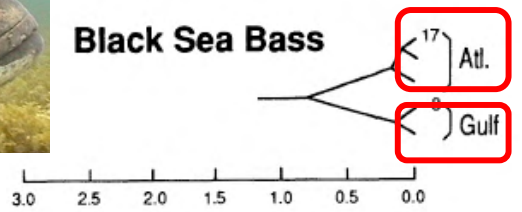


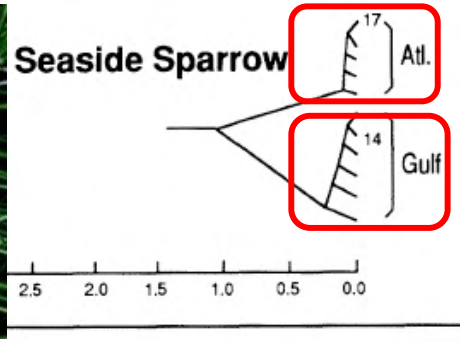


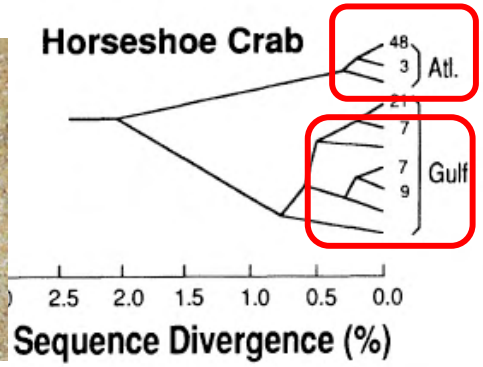


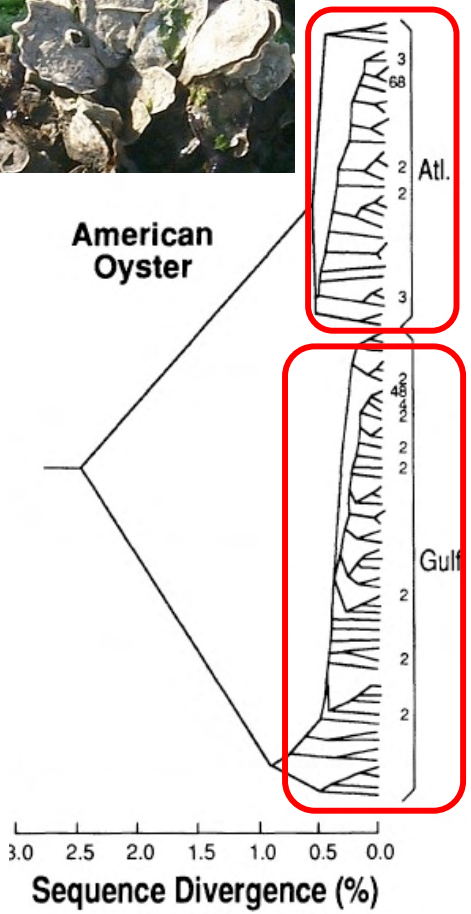
similarity of the association between genes and geography across species – CONCORDANCE – is typically used to test evolutionary hypotheses

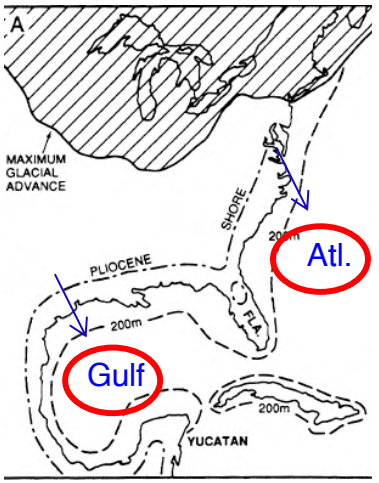
Concordance used in descriptive studies







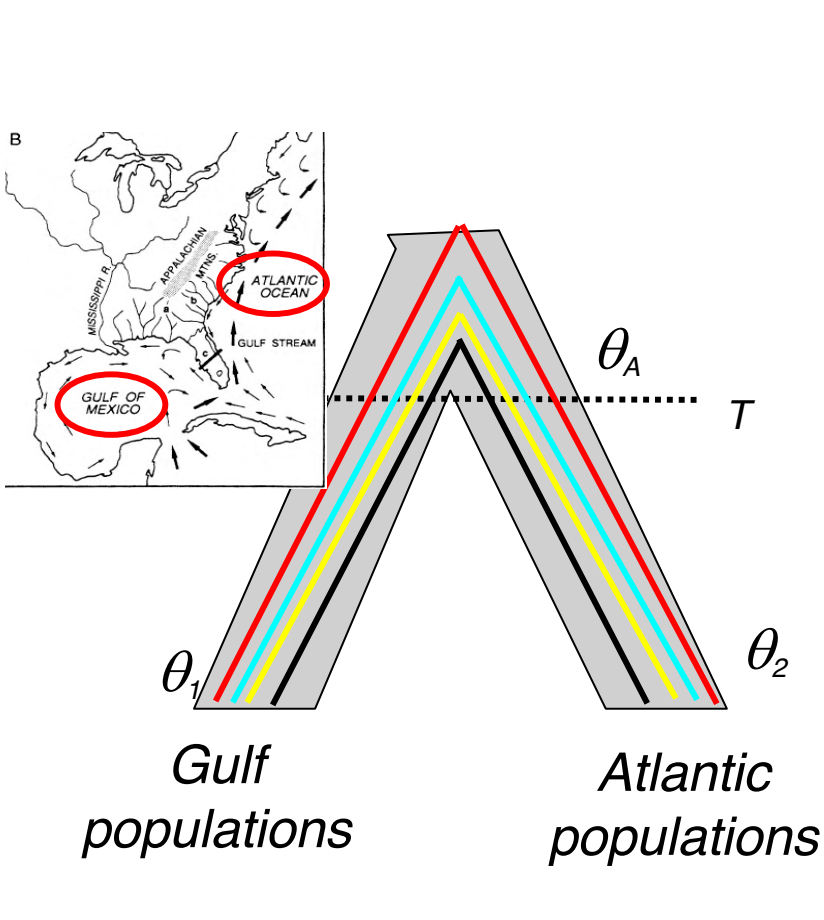


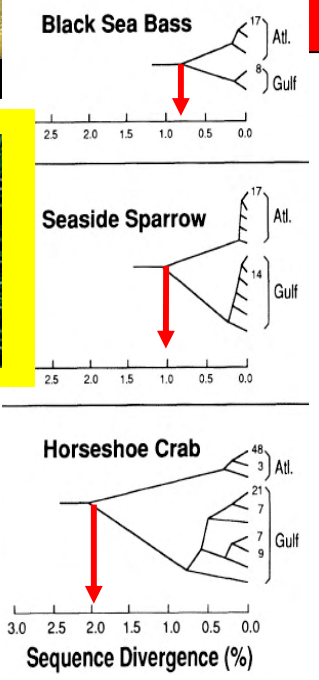


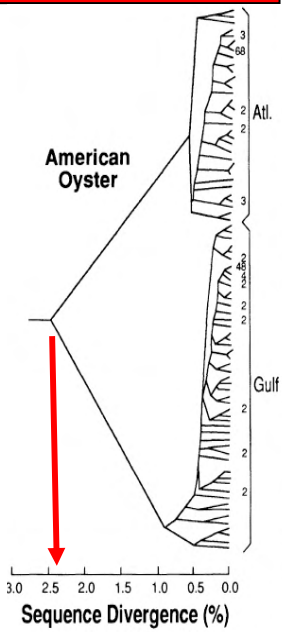
Avise 1992

Test for shared vicariant history of the coastal community:

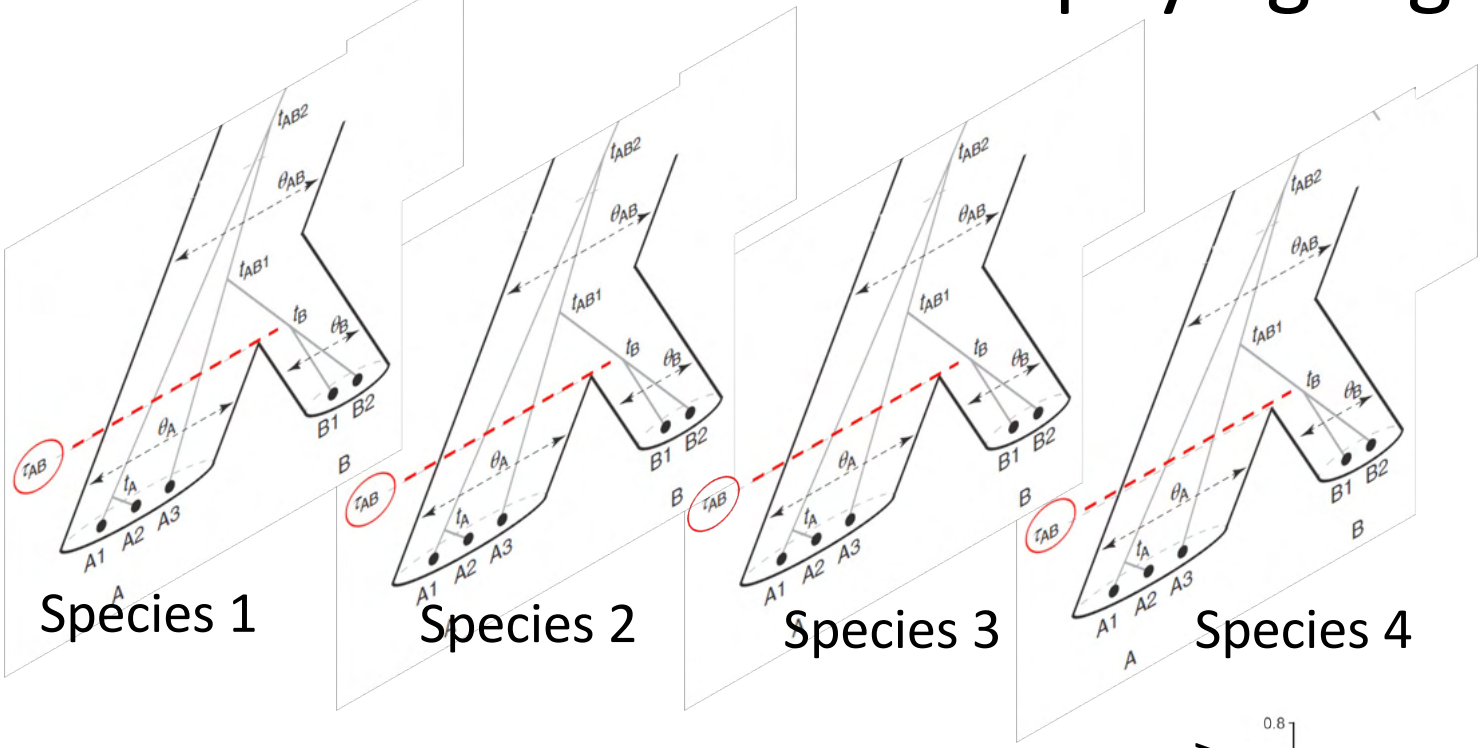
With a model we can assess statistically how much of a difference in the depths of the gene trees would still be consistent with the same divergence



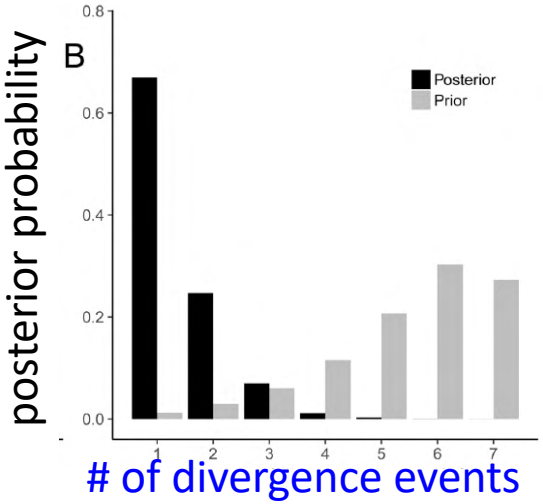




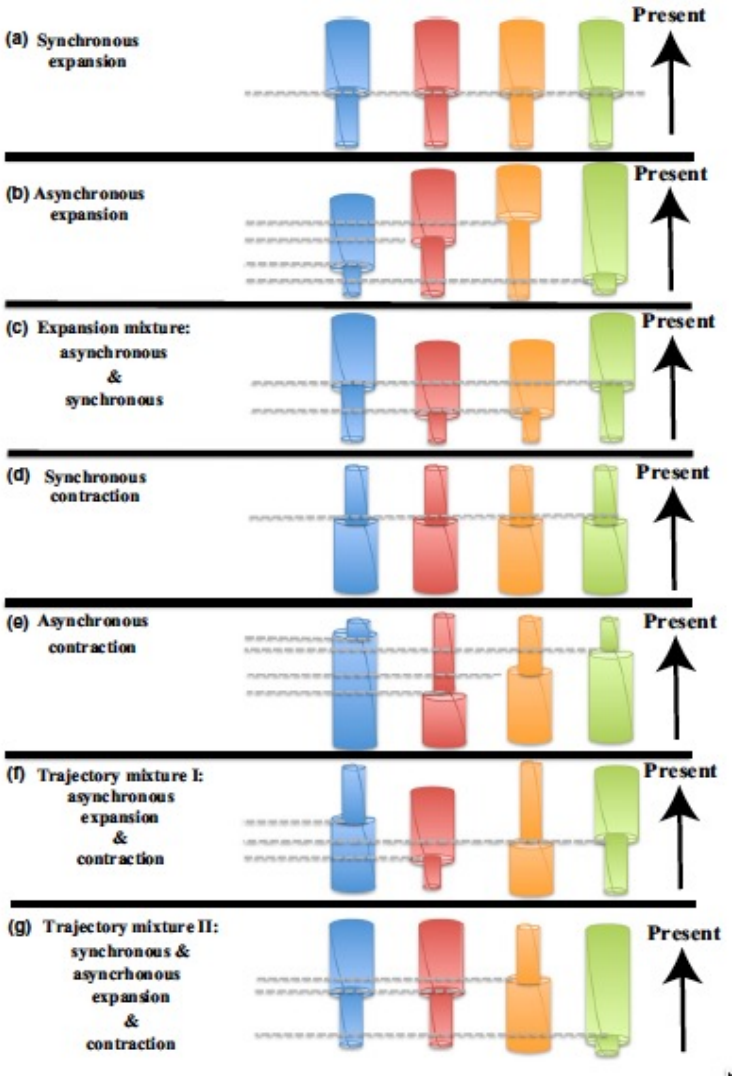
Concordance used in statistical phylogeography



Statistically evaluate a parameterized model of co-divergence among species using hierarchical Approximate Bayesian Computation (hABC), where a hyperparameter captures a community-level property (e.g., # of divergence events)



Concordance to test hypotheses of co-expansion



ECOLOGY LETTERS

Ecology Letters, (2016) **19**: 1457–1467 doi: 10.1111/ele.12695

Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates

Burbrink et al. 2016

Concordance criteria for hypothesis testing

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

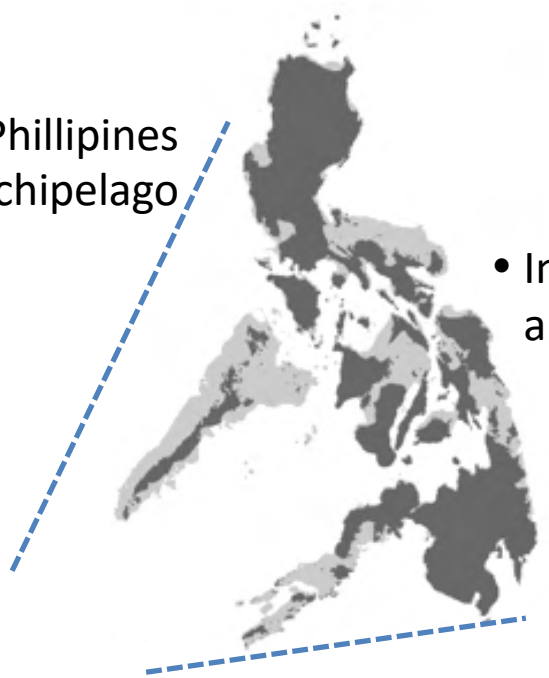
Oaks et al. (2012) *Evolution*

posterior probability

hABC approach

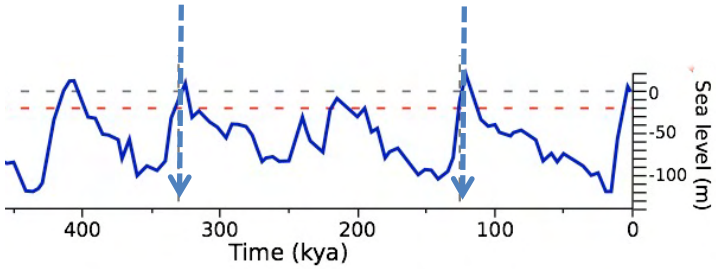
Number of divergence events

Phillipines Archipelago



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

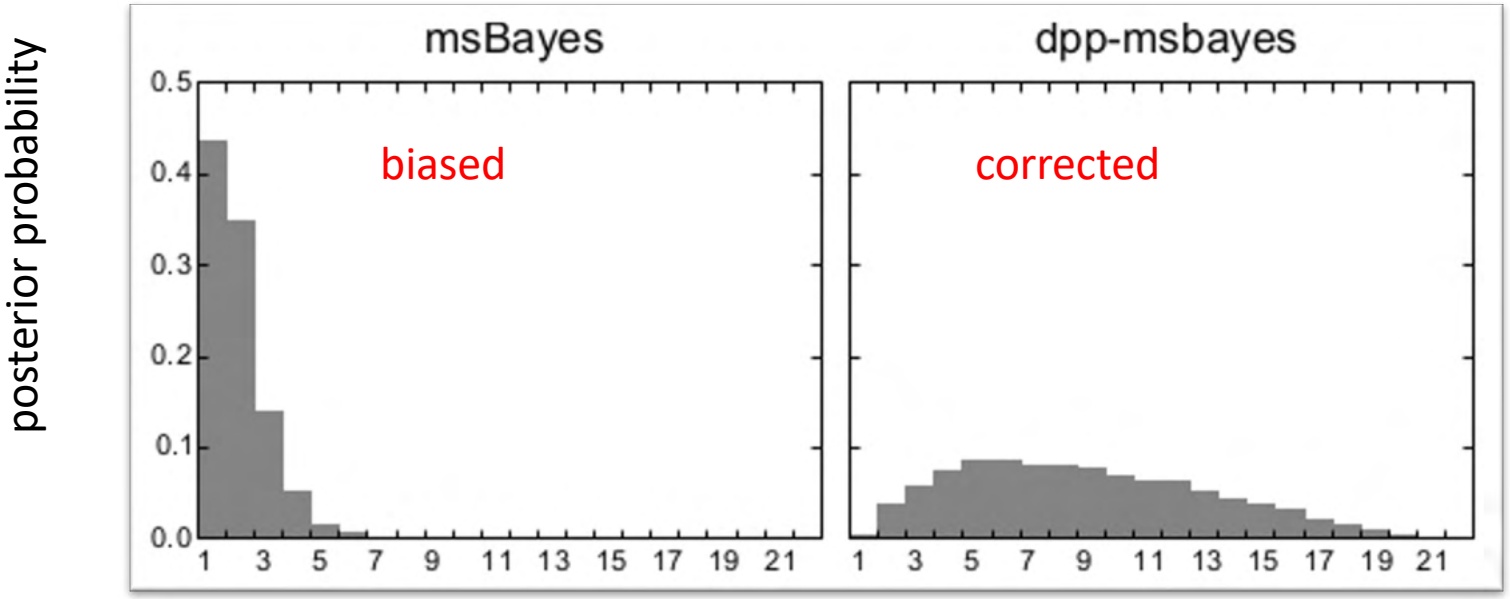
changes in connectivity/isolation of islands with sea-level changes (light versus dark grey outlines)



Concordance criteria for hypothesis testing

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

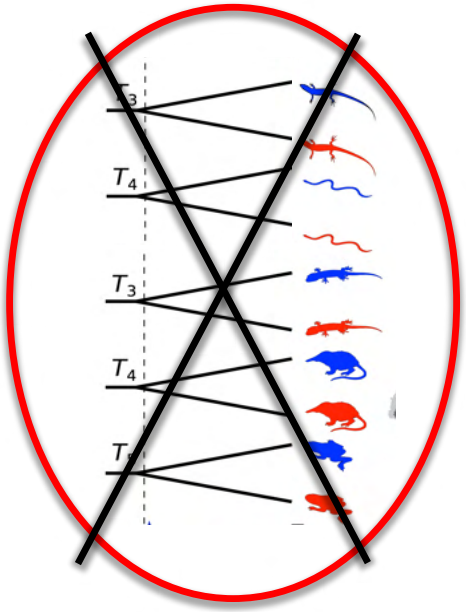
Performed a suite of simulation-based power analyses



(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

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

Concordance is arguably too generic of a hypothesis across these disparate taxa to test the “species pump” model of divergence.

Genes and Geography across species

for testing hypotheses
about evolutionary history

- potential for misleading inference by not considering both biotic and abiotic components

~~Generic~~ Refined hypothesis of phylogeographic concordance

- a study design that considers taxon attributes

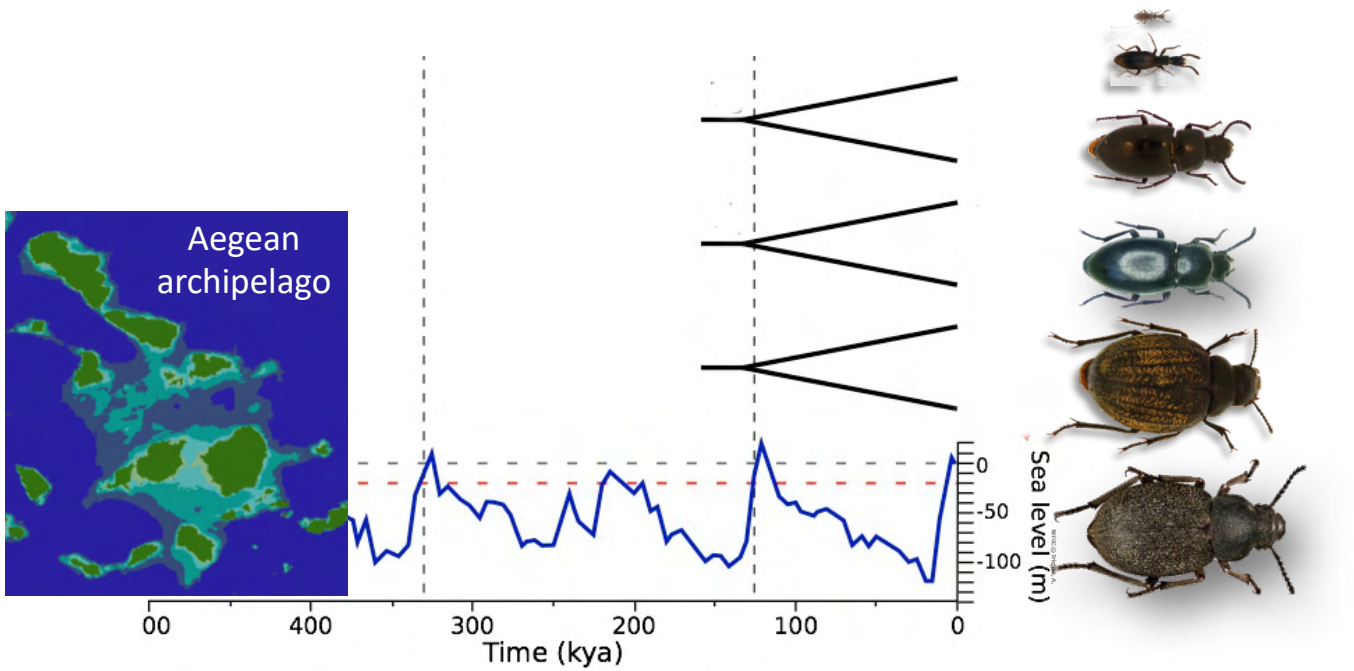
Hypothesis of simultaneous divergence to test whether sea-level oscillations during the Pleistocene caused diversification

Community-level tests of similarity in history

- Importance of considering refined-hypotheses based on taxon-specific traits when concordance is the criteria to be used to test a particular hypothesis

Refined hypotheses based on taxon-specific traits in comparative phylogeography

Soil inhabitants

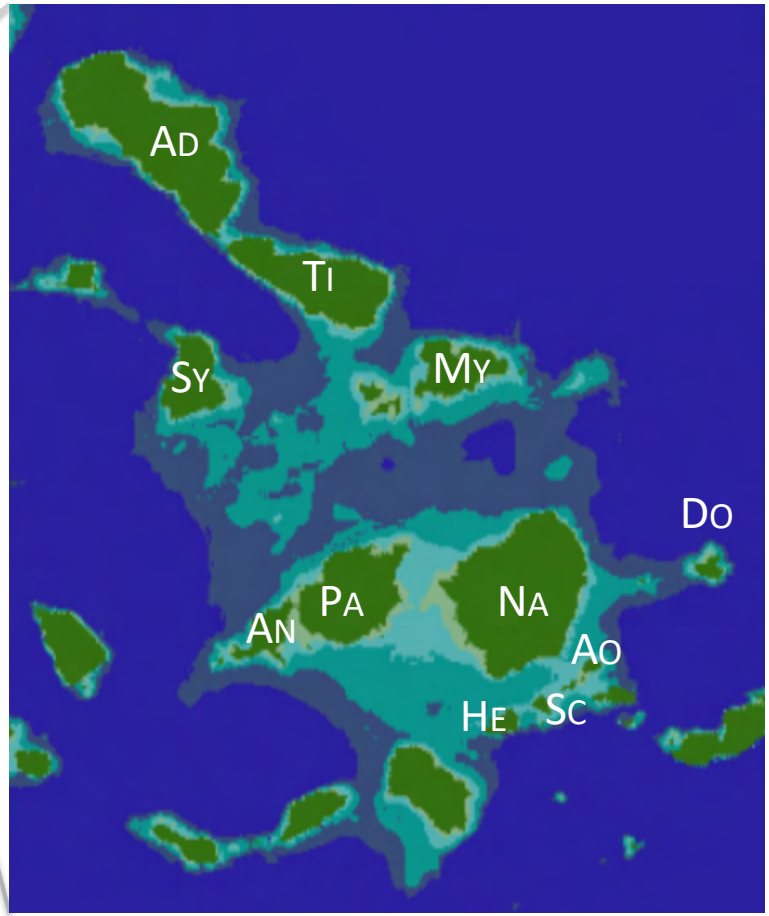


- key to avoiding null models that are “too null” and could lead to misleading inference

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

Refined models of phylogeographic concordance to test the “species pump” model

Current dry land
120 m isobath

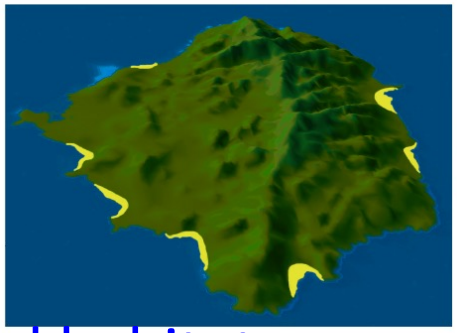


current dry land
0 to -30 m
-30 to -60 m
-60 to -90 m
-90 to -120 m
below -120 m

13 species of darkling beetles
(Coleoptera: Tenebrionidae)

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

- taxa differ in their soil associations



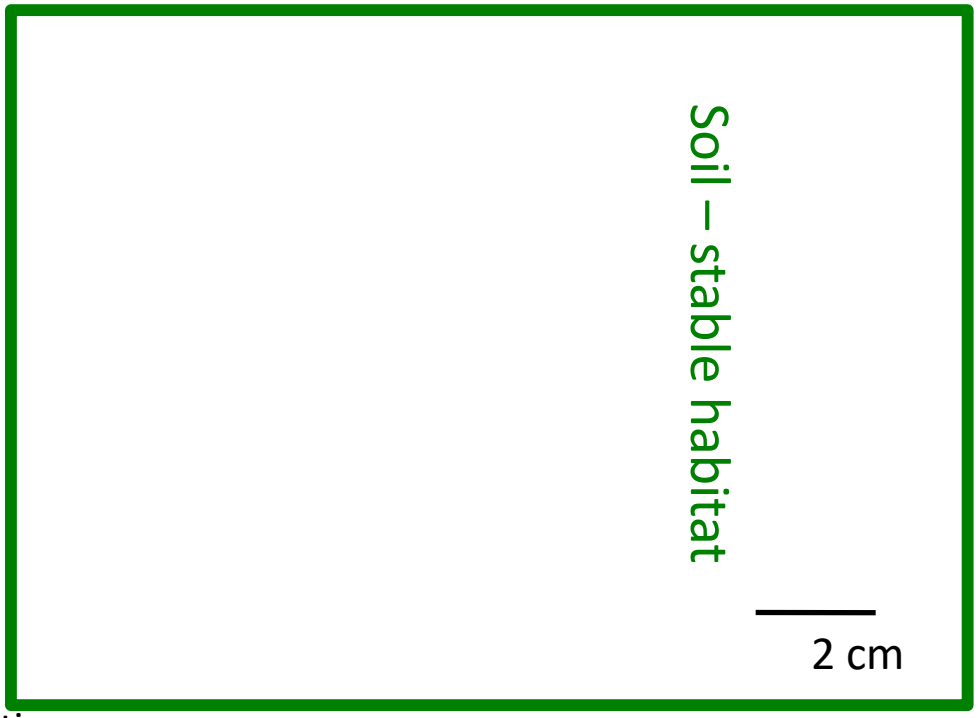
Ephemerality of sand habitats may supersede effects of sea-level connections!

- uniform trophic ecology & inherent dispersal abilities

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

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

Test of simultaneous divergence



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

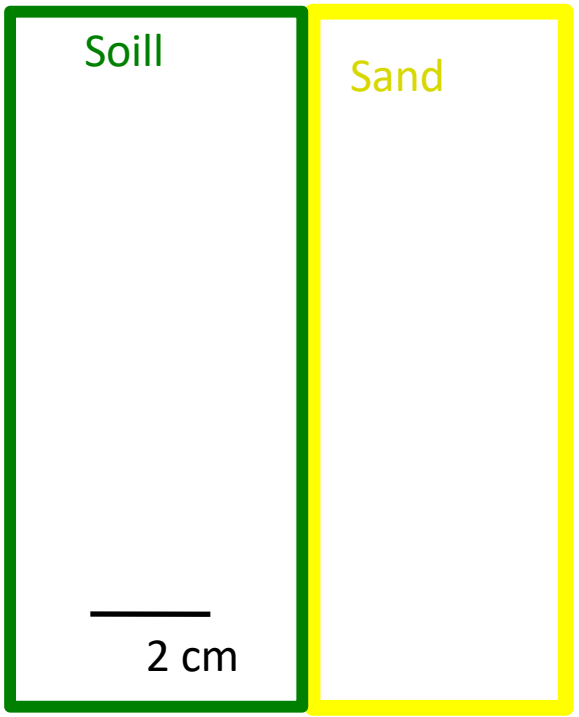
By focusing on ecologically equivalent taxa, test of concordance supported the species pump model of divergence

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

Generic hypotheses of global phylogeographic concordance

No evidence for simultaneous divergence

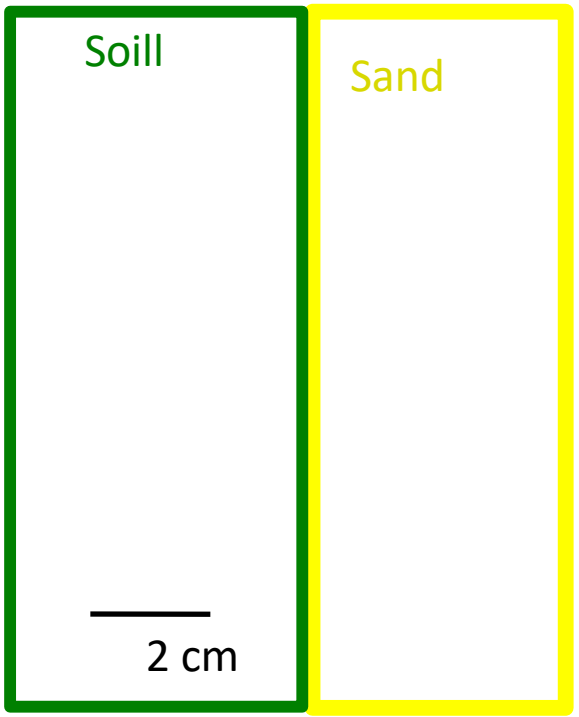
hABC implemented in
dpp-msbayes (Oaks, 2014)



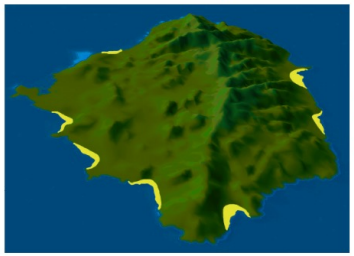
Papadopoulou & Knowles (2015) *Mol. Ecol*. 24: 4252-4268

Generic hypotheses of global phylogeographic concordance

No evidence for simultaneous divergence



Ephemerality of sand habitats!

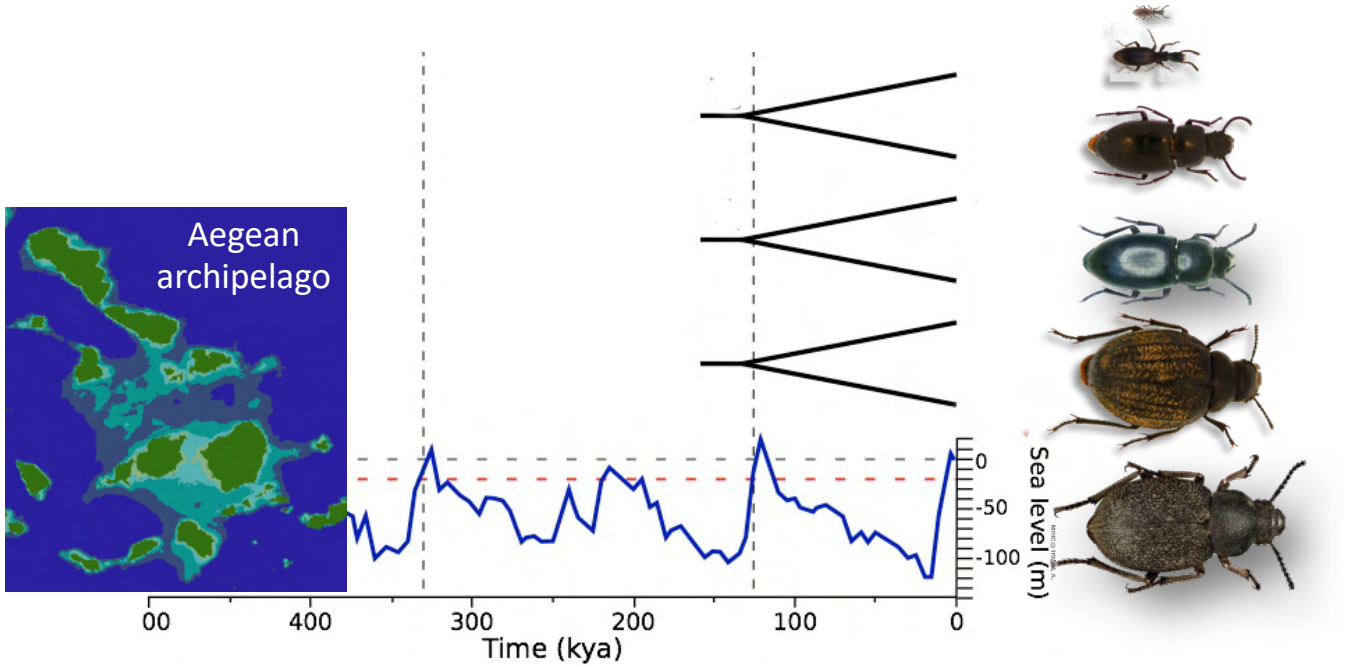


~~Lack of global concordance~~ → ~~rejection of species pump model of divergence~~ ???

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

Refined hypotheses based on taxon-specific traits in comparative phylogeography

Soil– stable habitat

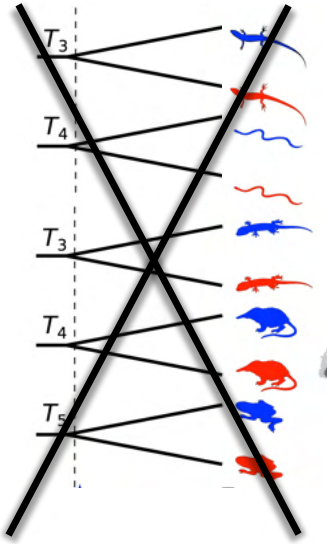


- refinement of the expectation for concordance is needed for concordance itself to be a meaningful metric

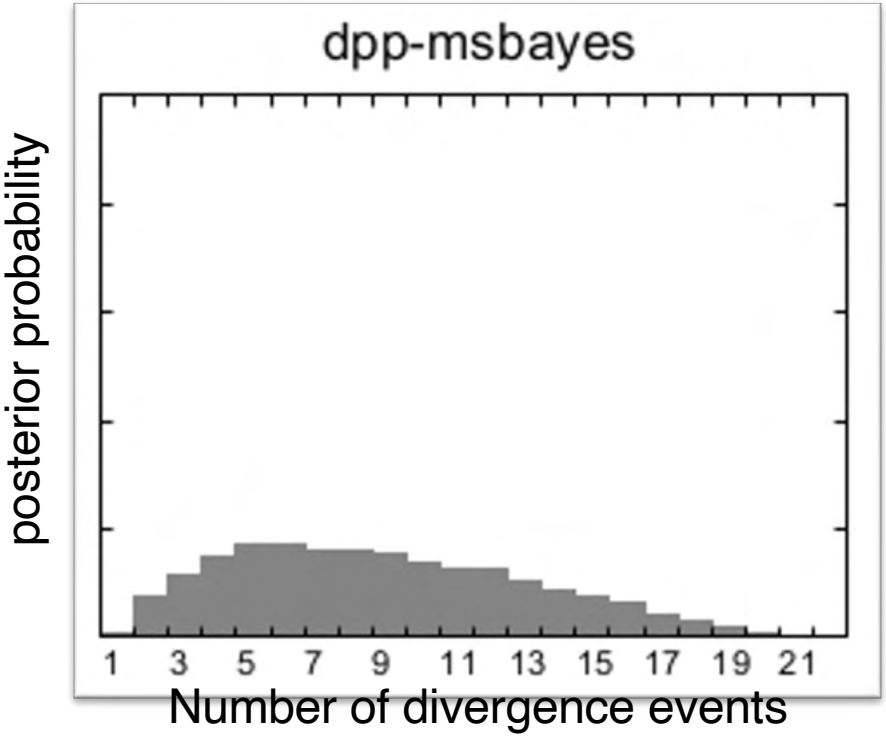
- reduced predictive power of generic hypotheses – their rejection leads to inconclusive statements that do not offer particularly meaningful insights

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!



- ad hoc interpretations of discordance

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

Papadopoulou & Knowles (2016) *PNAS*

Refined hypotheses based on taxon-specific traits in comparative phylogeography

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

Biological insights:

How we formulate hypotheses for
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

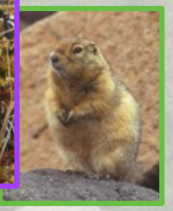
He et al (2013) *Evolution*

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.*

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

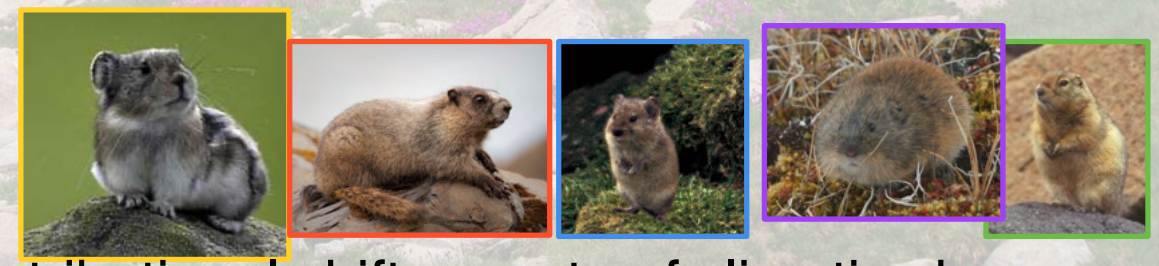
Massatti & Knowles (2014, 2016) *Evolution,* Mol. Ecol.

Role of habitat stability in structuring genetic variation

He et al (2013) *Evolution*

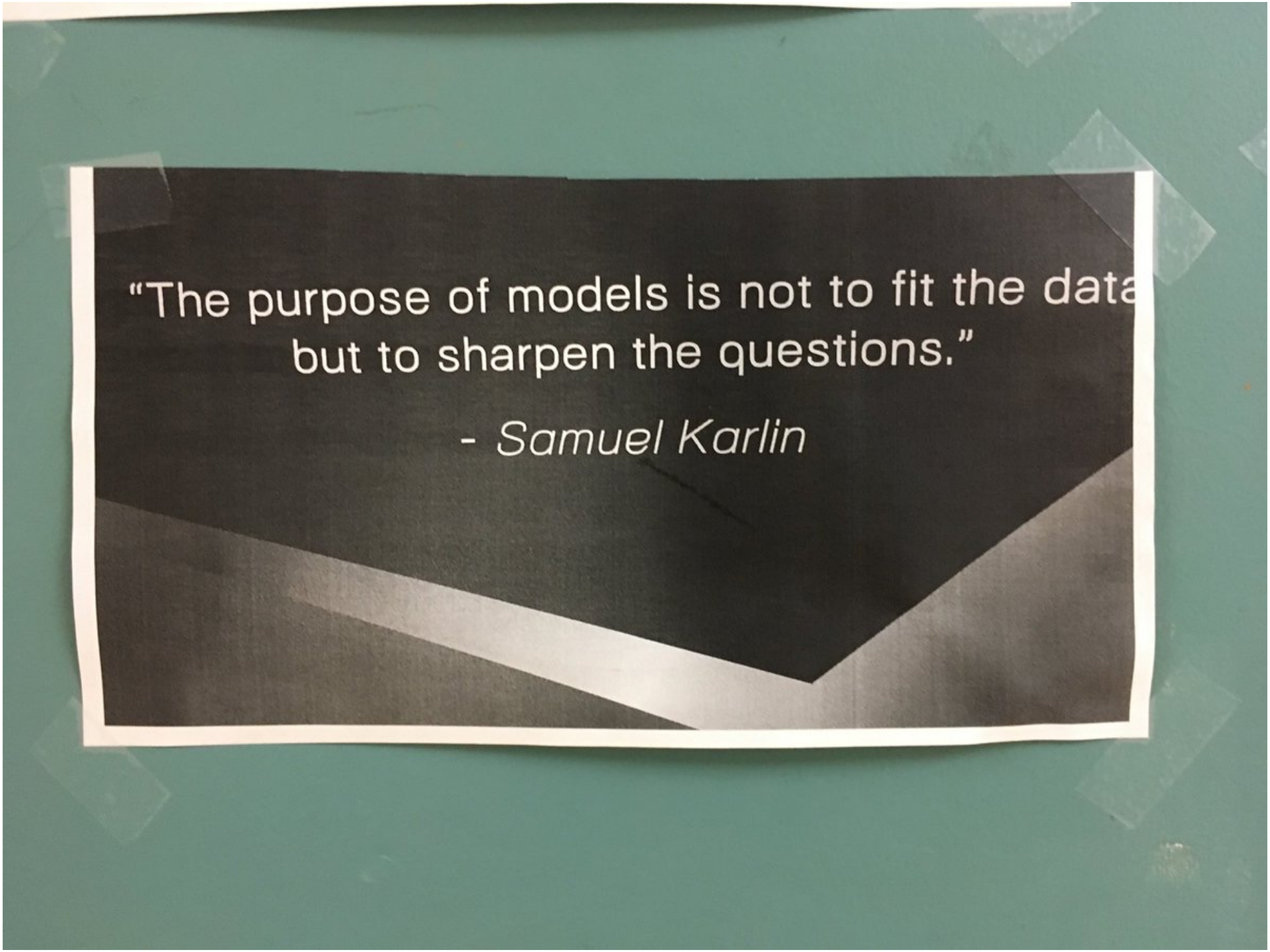
Present versus past distributions as drivers of divergence

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.*


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

Evolutionary applications of model-based analyses:

(i) Inferring species boundaries (aka species delimitation)

(ii) Phylogenetic inference (and beyond the species tree)

(iii) Biogeographic study

(iv) Phylogeography

(v) Adaptive evolution

Myotis lucifugus

Little brown bats are widespread in North America and were the most abundant species in the eastern US prior to white nose syndrome (WNS), which is caused by introduced fungal pathogen

Little brown bats decimated by white nose syndrome (WNS)

Population declines > 90% since introduction of fungal pathogen that causes WNS

Dead bats in underground hibernation sites (shown here on the floor of a mine)

Others leave hibernating sites prematurely, like these dead bats on the outer screen of a house < 1 km from a hibernation site (note the snowy landscape).

Myotis lucifugus

Survival of the species may ultimately depend upon its capacity for adaptive change

- Compare the genetic makeup of wild survivors and non-survivors of WNS to tests for adaptive change

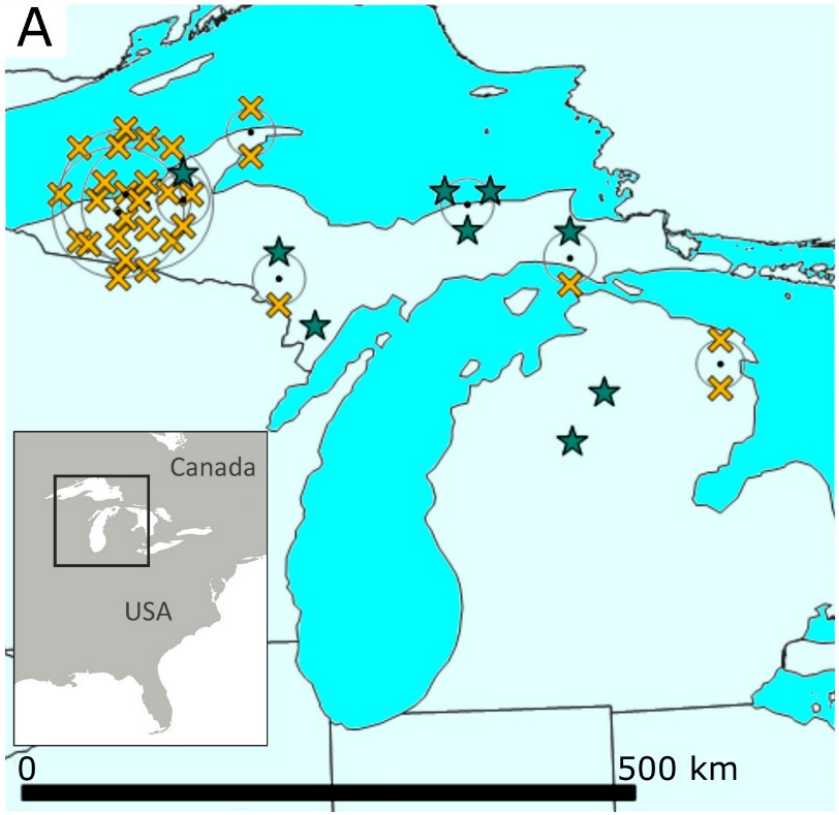
Auteri GG, Knowles LL (2020) Decimated little brown bat population show potential for adaptive change. *Scientific Reports.* 10:3023. doi.org/10.1038/s41598-020-59797-4

Giorgia G. Auteri

UNIVERSITY OF MICHIGAN

Studied geographically isolated population of little brown bats

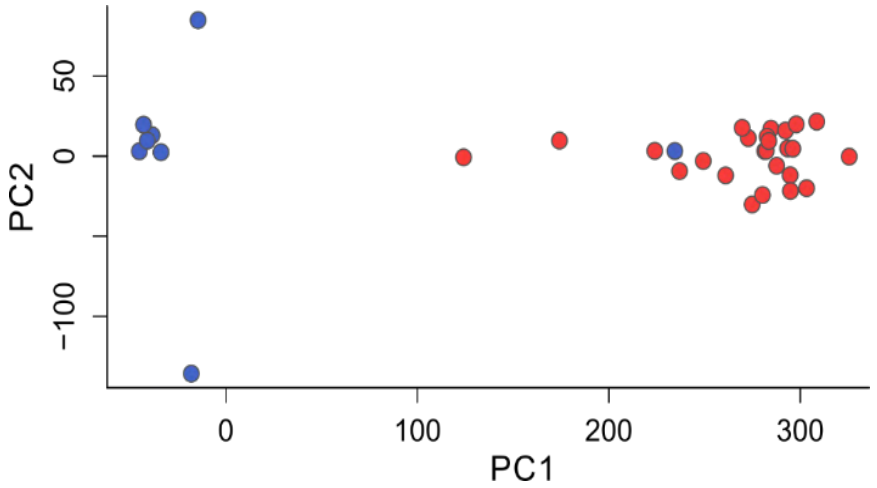
non-survivor
survivor



WNS arrived in 2014

- RADseq: 14,345 loci , 19,797 SNPs

Evidence of strong genetic drift caused by the massive population losses in little brown bats.

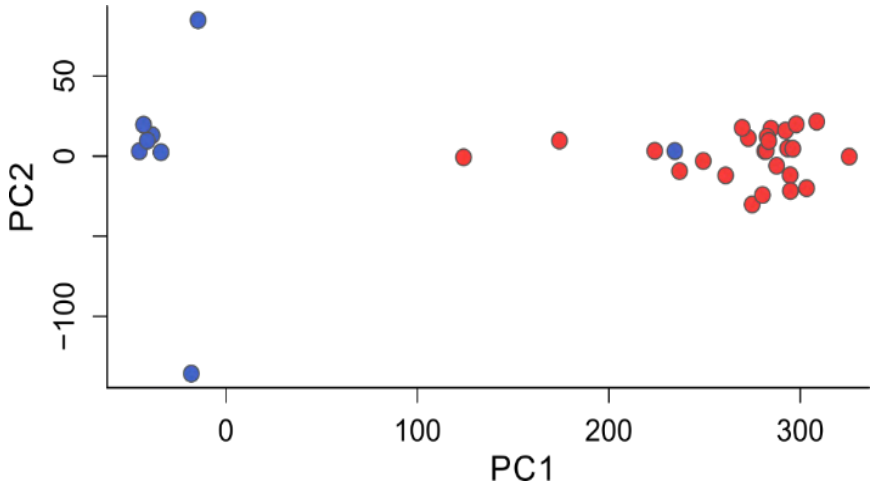


PCA of survivors of WNS (in blue) with non-survivors (in red) projected onto the PC axes

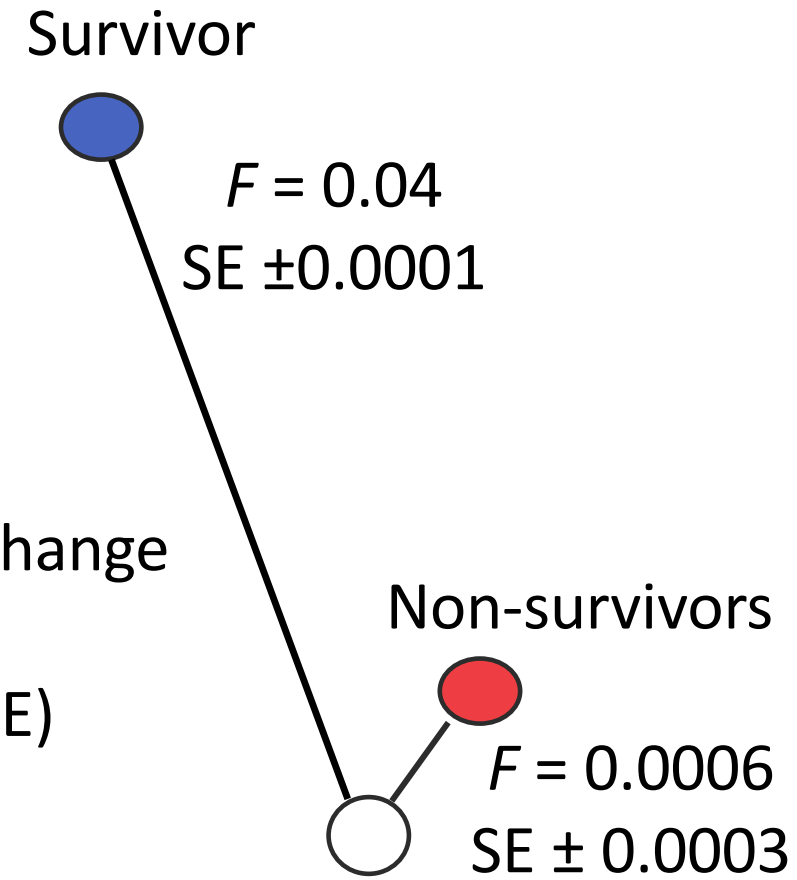
- 14,345 SNPs and 33 individuals

Auteri GG, Knowles LL (2020) Decimated little brown bat population show potential for adaptive change. *Scientific Reports.* 10:3023. doi.org/10.1038/s41598-020-59797-4

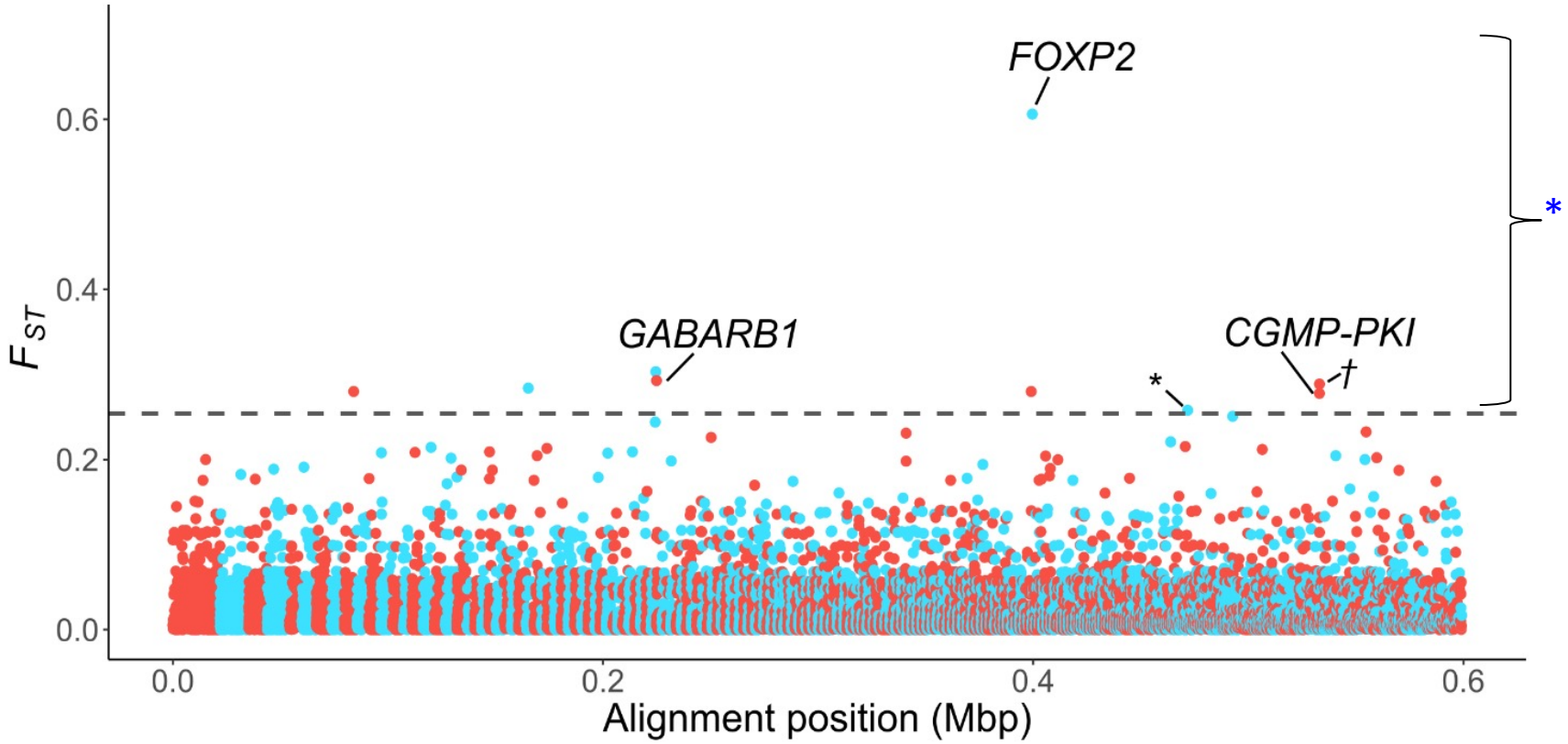
Evidence of strong genetic drift caused by the massive population losses in little brown bats.



Quantified rate of evolutionary change from inferred ancestor (using F-model in STRUCTURE)



To identify genetic changes among individuals that might have contributed to their survival of WNS, as opposed to changes due to strong genetic drift, used an F_{ST}-outlier approach

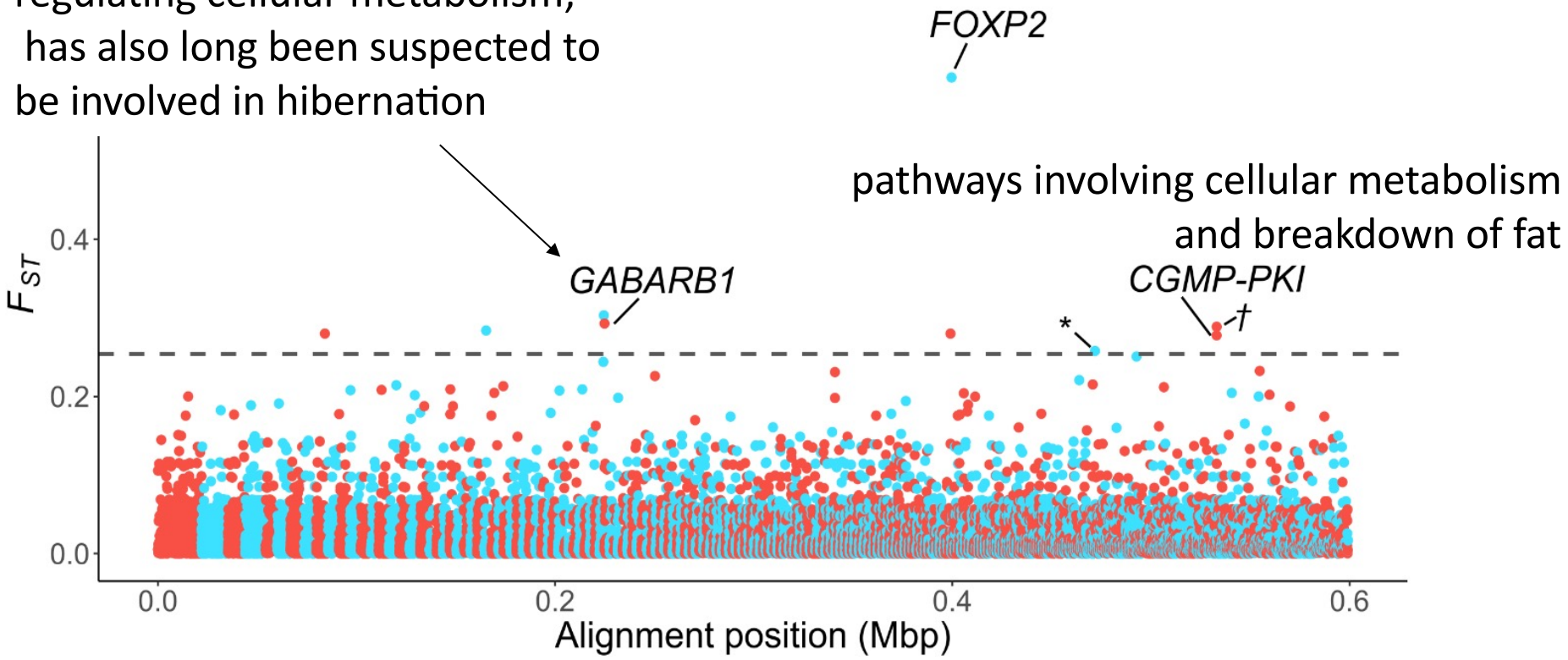


- alternating red and blue mark different genomic scaffolds

*signature of selection can be detected by levels of genetic differentiation at a gene that exceeds background levels across the genome

Links between metabolic demands and survival

regulating cellular metabolism;
has also long been suspected to
be involved in hibernation

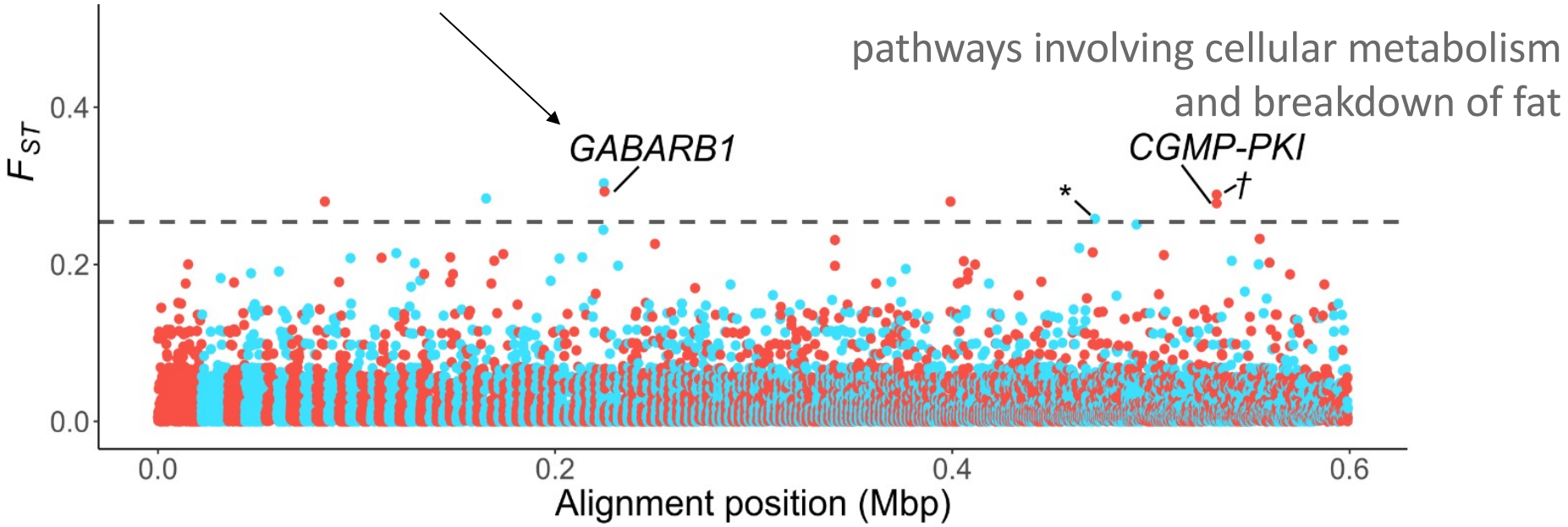


Physiological functions that make immediate sense in an adaptive context—deaths from the WNS fungus are a result of too frequent arousals from hibernation that causes starvation.

Links between metabolic demands and survival

associated with vocalizations, and echolocation in bats

regulating cellular metabolism; has also long been suspected to be involved in hibernation



Variation in calls is closely associated with type of prey and the habitat bats navigate, suggesting potentially adaptive shifts might result from selective pressures related to proficient hunting or prey preferences

Too soon to claim that the species will be “saved” via an evolutionary rescue effect.

Evidence of potentially adaptive evolution in the survivors of little brown bats is particularly notable on several fronts:

- We detected selectively driven divergence, despite strong genetic drift caused by the massive population losses in little brown bats.

- These evolutionary changes were detected in less than three generations since exposure to WNS

- Putatively selected loci and their potential adaptive functions point to multifaceted nature of selection (i.e., genes linked to physiological and behavioral traits, whose roles vary across habitats of highly seasonal environments)

Evolutionary applications of model-based analyses:

(i) Inferring species boundaries (aka species delimitation)

(ii) Phylogenetic inference (and beyond the species tree)

(iii) Biogeographic study

(iv) Phylogeography

(v) Adaptive evolution

Species delimitation (discovery)

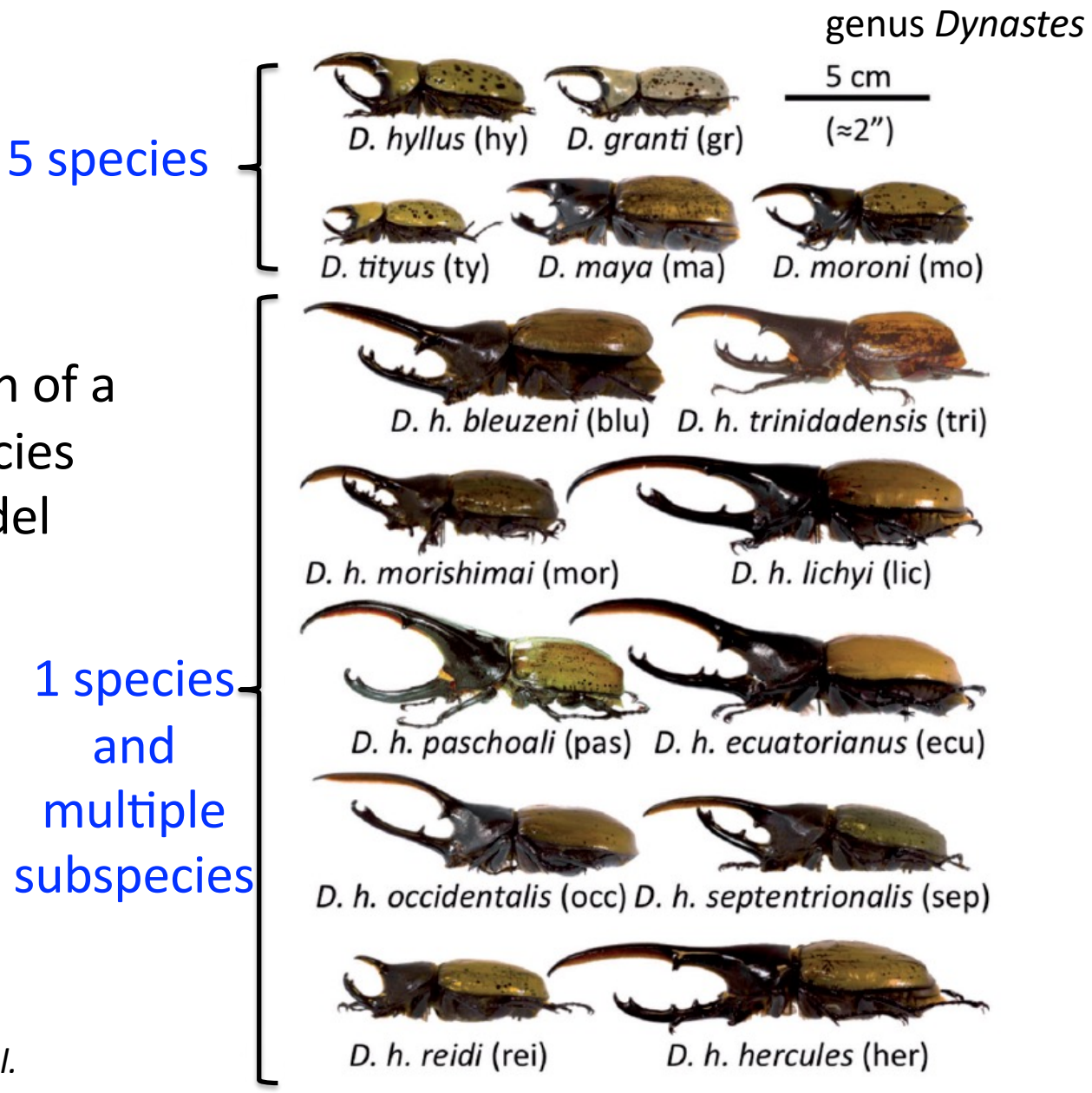
Learning goals:

- Describe applications of the multispecies coalescent (MSC) to species delimitation
- Explain the merit/limitations of the multispecies coalescent (MSC) to delimitation
- Describe (i) how over-estimation of species numbers might occur with applications based on the MSC (ii) what determines the degree of overestimation
- Explain the relevance of the speciation process to delimitation approaches

Hypotheses about species boundaries

Model-based inference of species boundaries

Statistical evaluation of a hypothesized species delimitation model

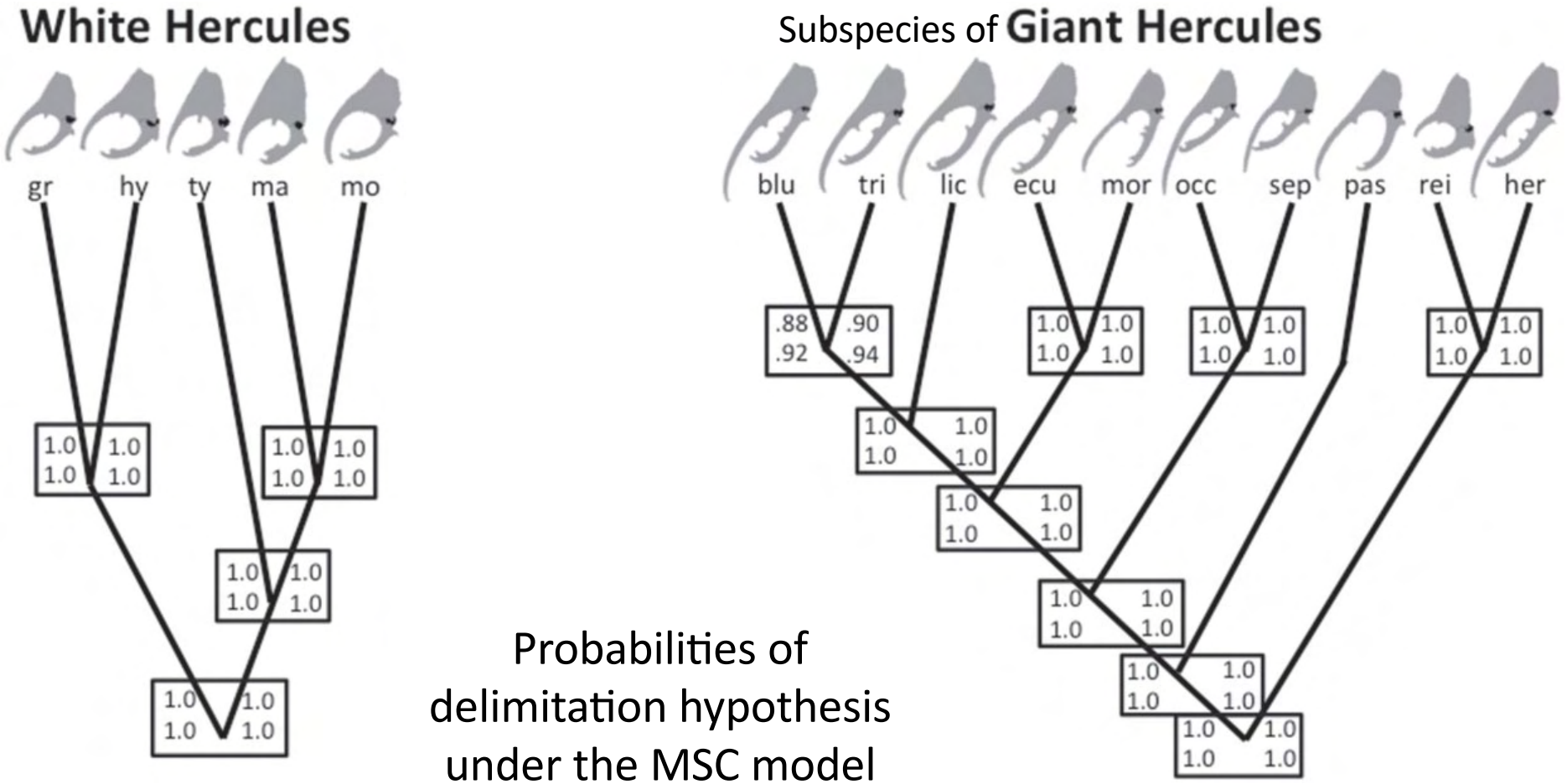


Huang & Knowles (2016) *Syst. Biol.*

Model-based inference of species boundaries

5 recognized species
In North America

1 recognized species
In South America



Probabilities of
delimitation hypothesis
under the MSC model

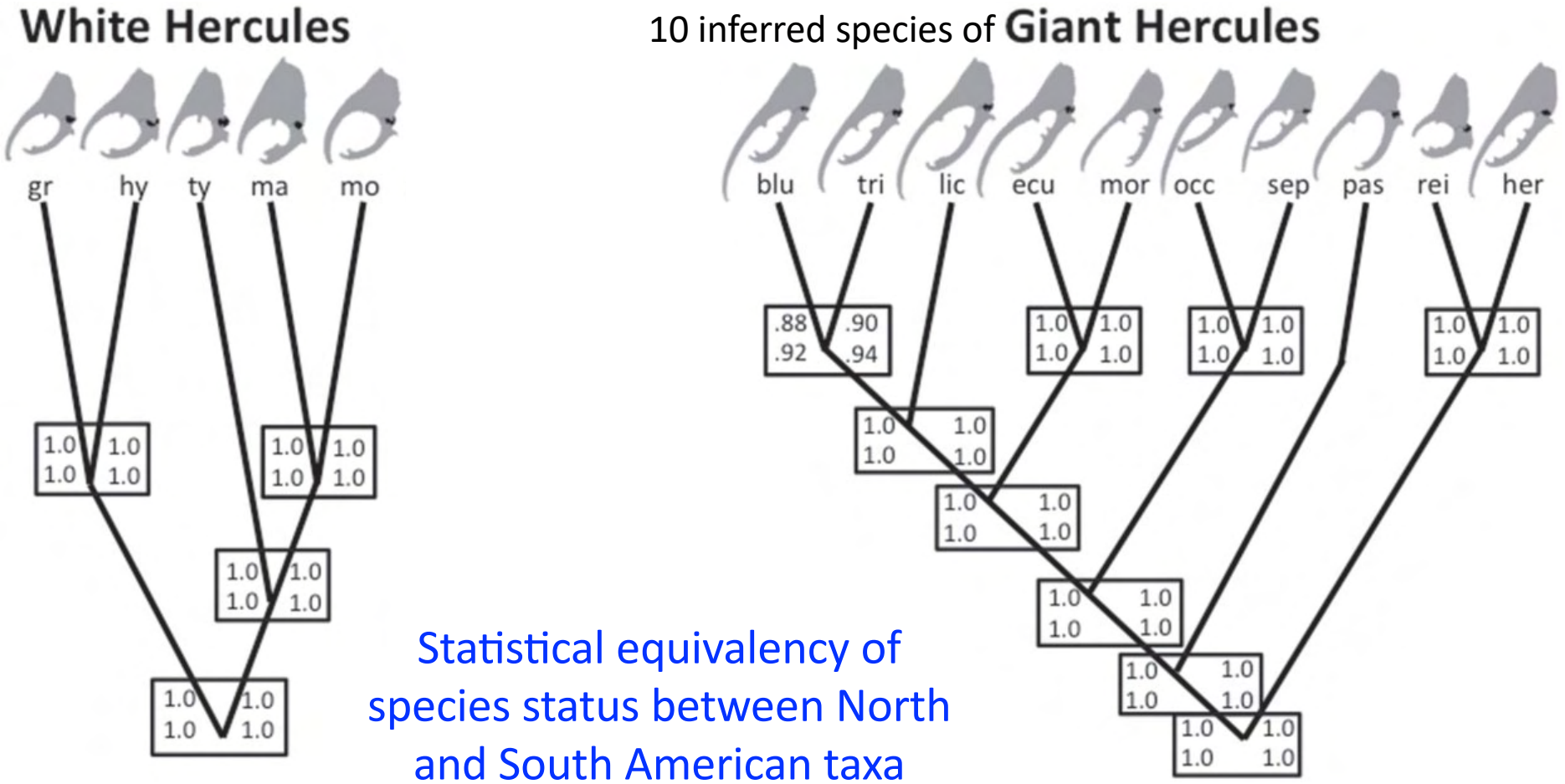
genus *Dynastes*

Huang & Knowles (2016) *Syst. Biol.*

Model-based inference of species boundaries

5 recognized species
In North America

1 recognized species
In South America



Statistical equivalency of species status between North and South American taxa

genus *Dynastes*

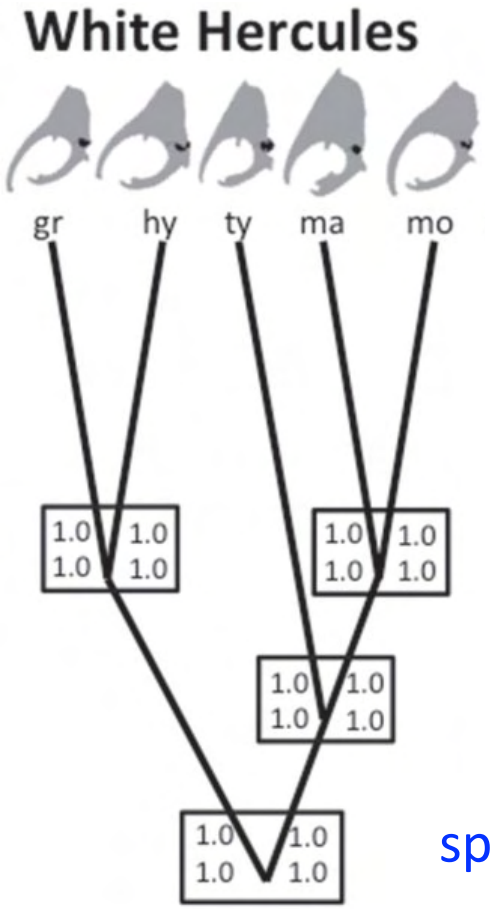
Huang & Knowles (2016) *Syst. Biol.*

Model-based inference of species boundaries

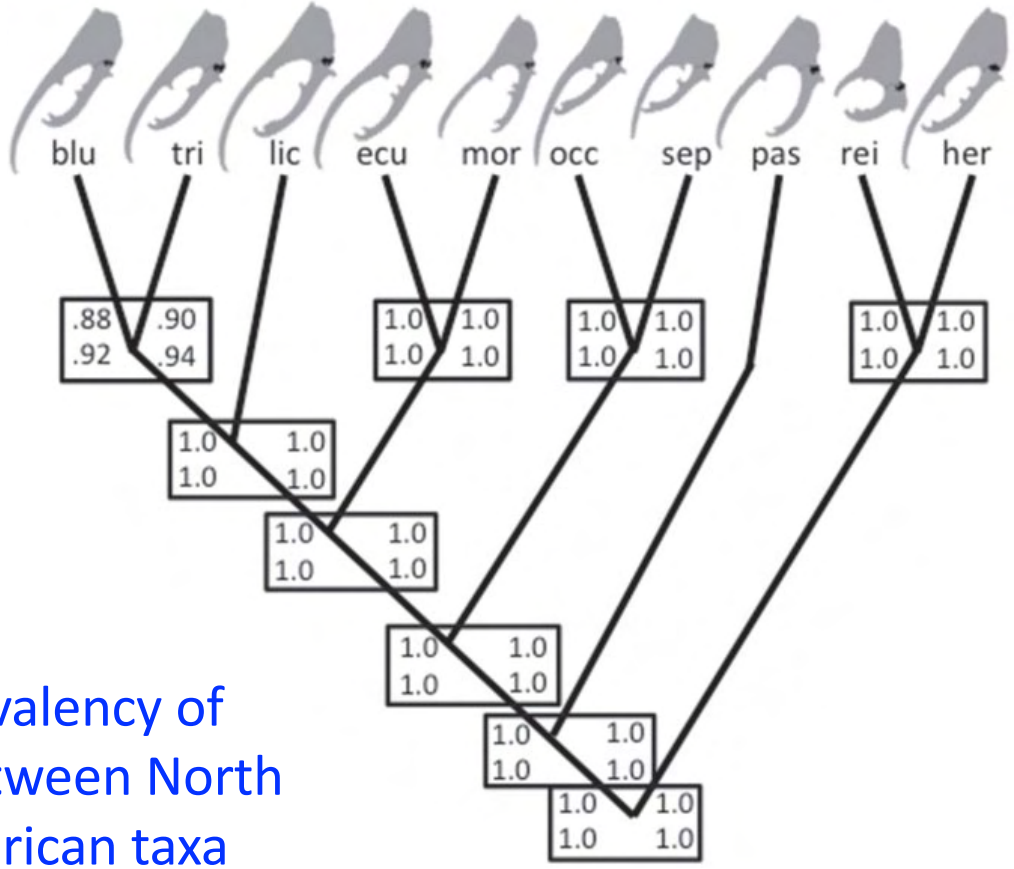
5 recognized species
In North America

Differences in recognized species is due to differences in taxonomic practices.

1 recognized species
In South America



10 inferred species of **Giant Hercules**



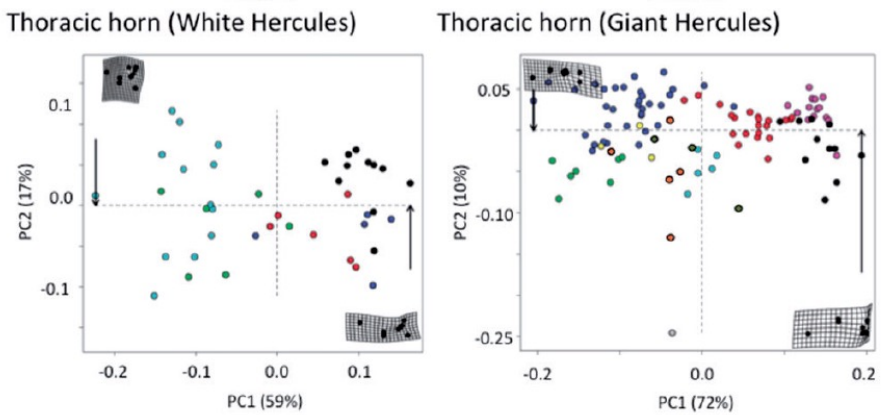
Statistical equivalency of species status between North and South American taxa

genus *Dynastes*

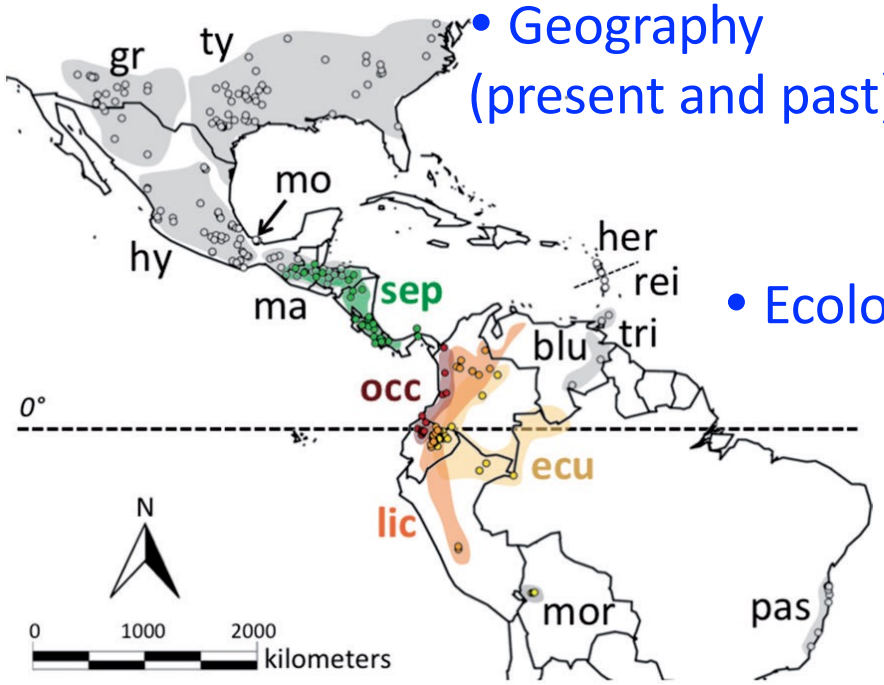
Huang & Knowles (2016) *Syst. Biol.*

Integration across data types to corroborate delimited taxa

- Quantification of phenotype

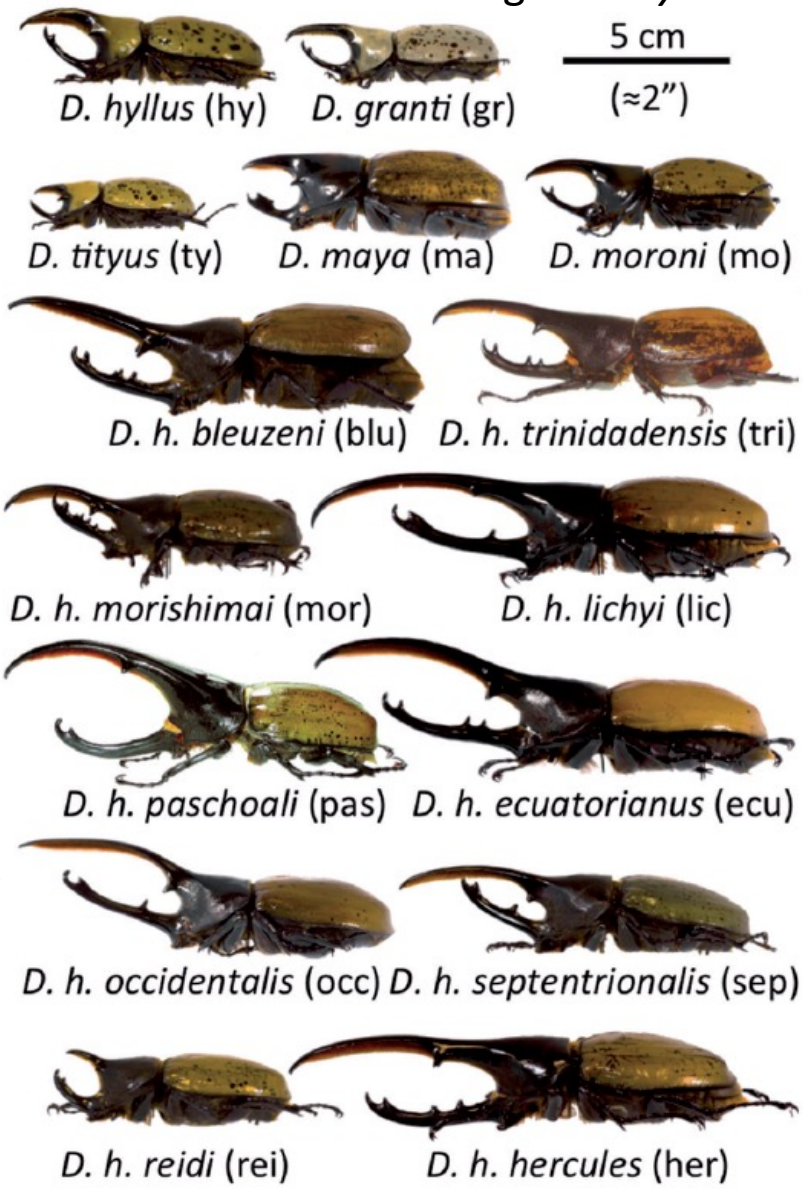


- Geography (present and past)



- Ecology

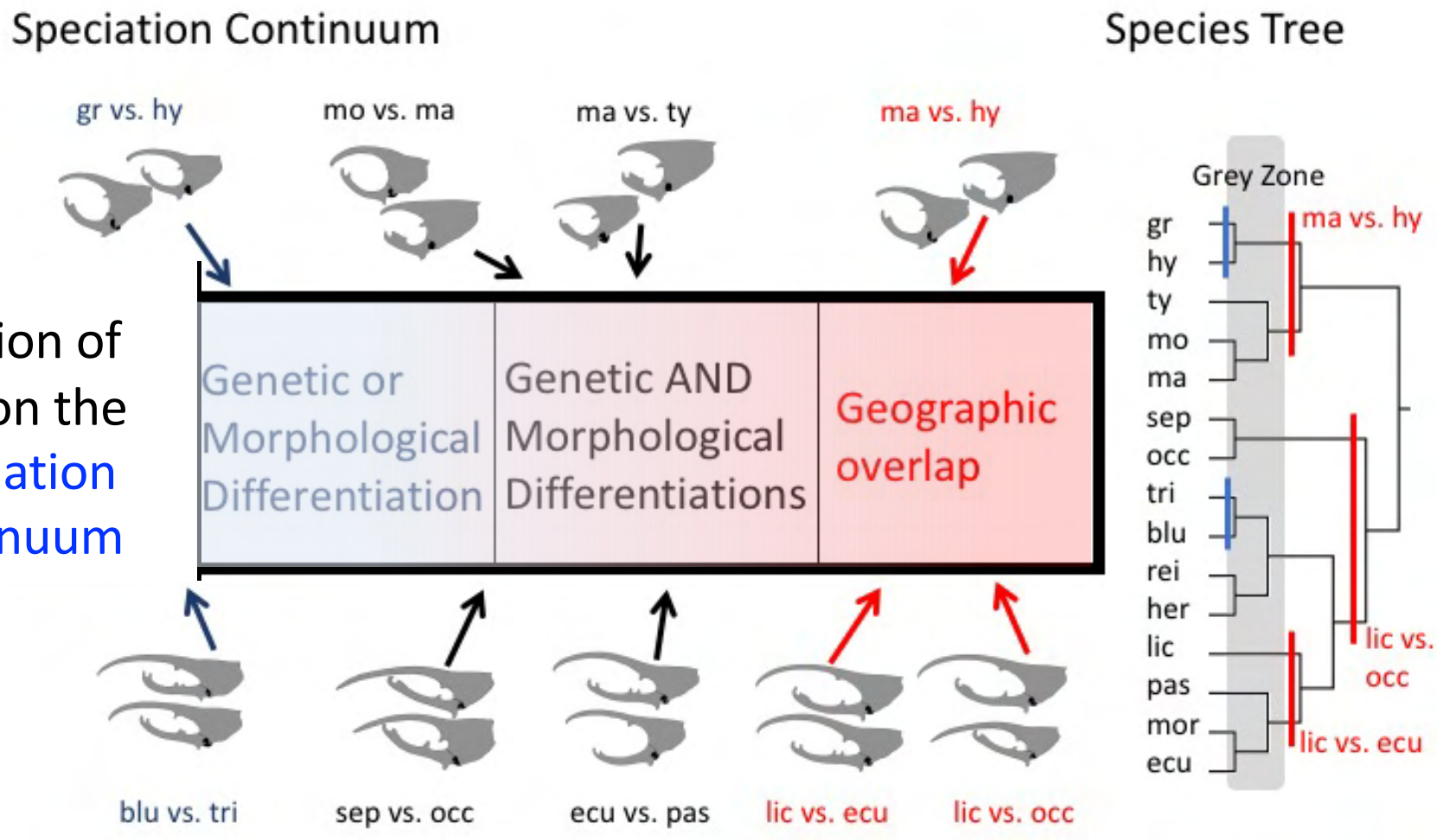
genus *Dynastes*



Huang & Knowles (2016) *Syst. Biol.*

Integrative data also provides insights into the divergence process

Position of taxa on the speciation continuum



Huang & Knowles (2016) *Syst. Biol.*

Transformative potential of model-based analyses:

- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation based on genetic data alone
- Demographic inference

....models are how we communicate
our knowledge to a statistical apparatus

Transformative potential of model-based analyses:

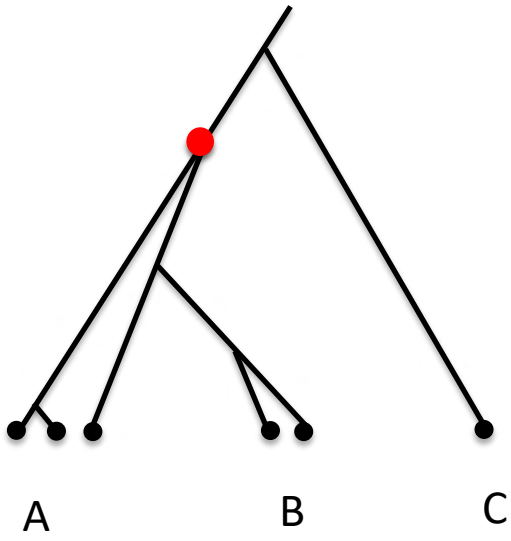
- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation based on genetic data alone
- Demographic inference

- All models are flawed..., some are more or less useful

....models are how we communicate
our knowledge to a statistical apparatus

Transformative potential of model-based analyses:

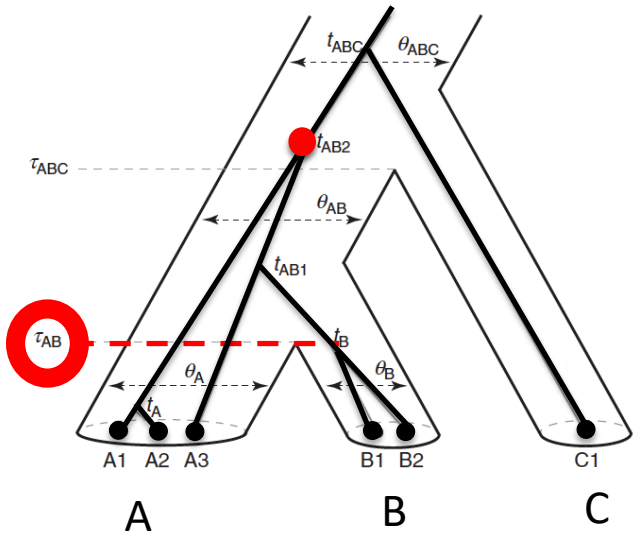
- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation
- Demographic inference
  (e.g., estimate divergence between population A and B)



Model of gene lineage divergence under an assumption of a molecular clock

Transformative potential of model-based analyses:

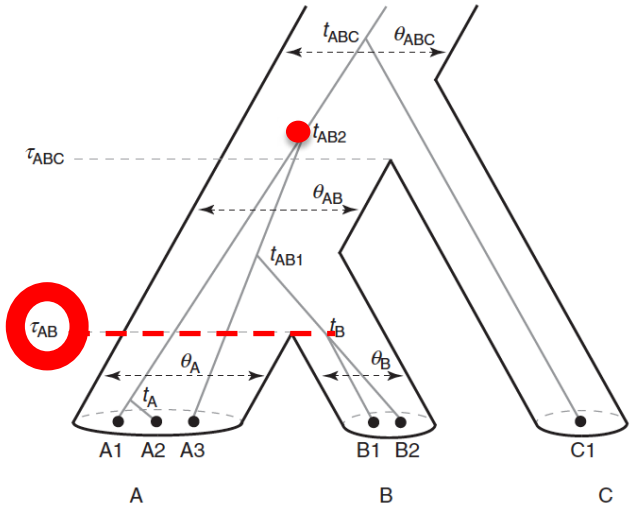
- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation
- Demographic inference
  (e.g., estimate divergence between population A and B)



Coalescent model of
gene lineage sorting process

Transformative potential of model-based analyses:

- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation
- Demographic inference
  (e.g., time of divergence)

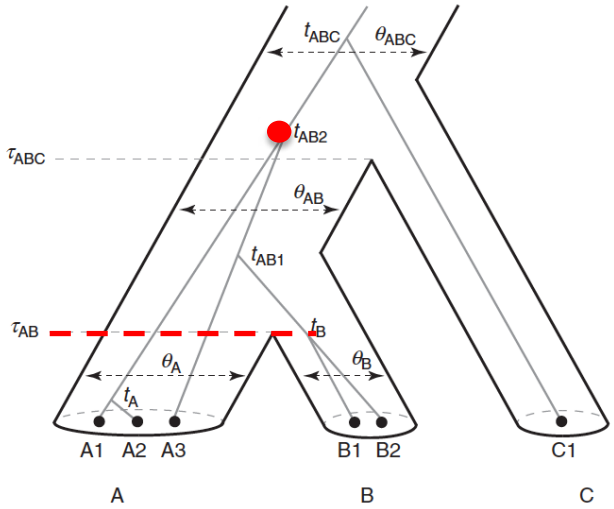


- All models are flawed..., some are more or less useful

....depending upon how effectively they represent our expert knowledge of evolution

Transformative potential of model-based analyses:

- Codon substitution and analysis of natural selection
- Adaptive molecular evolution
- Divergence time estimation and biogeographic analysis
- Phylogenetic inference
- Species delimitation
- Demographic inference
  (e.g., time of divergence)



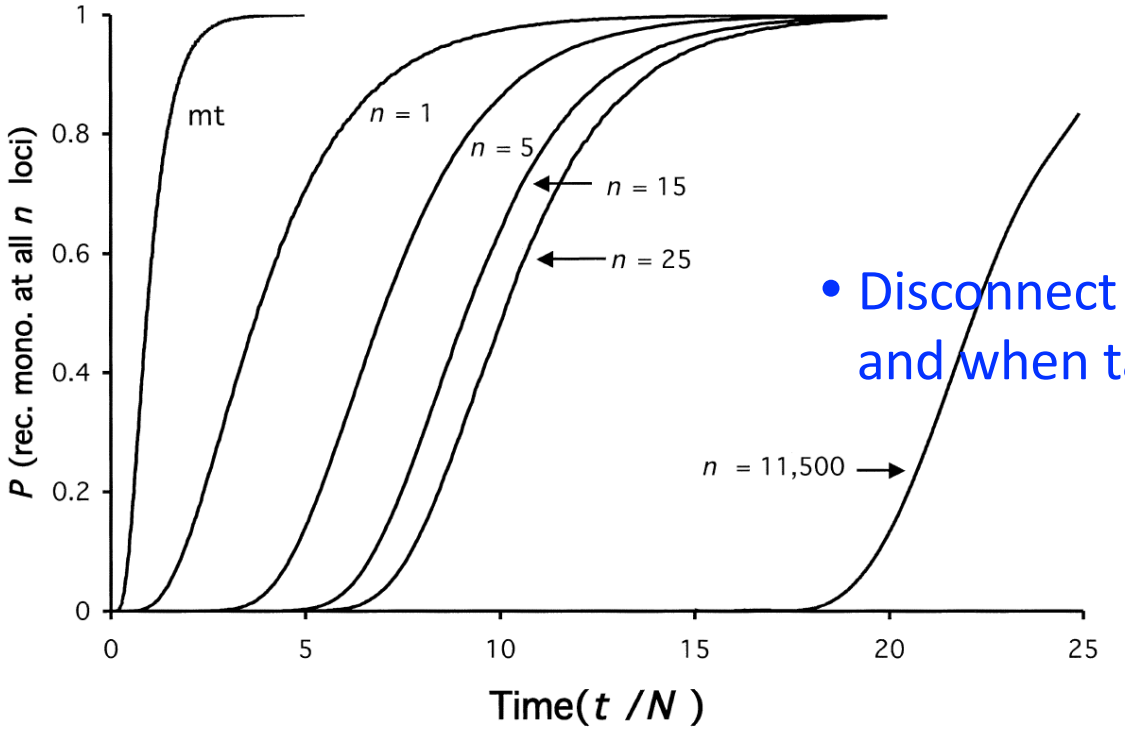
- All models are flawed..., some are more or less useful

....depending upon how effectively they represent our expert knowledge of evolution

Isolation is the property that allows species to be recognized genetically

- Exclusivity criteria (e.g., reciprocal monophyly)

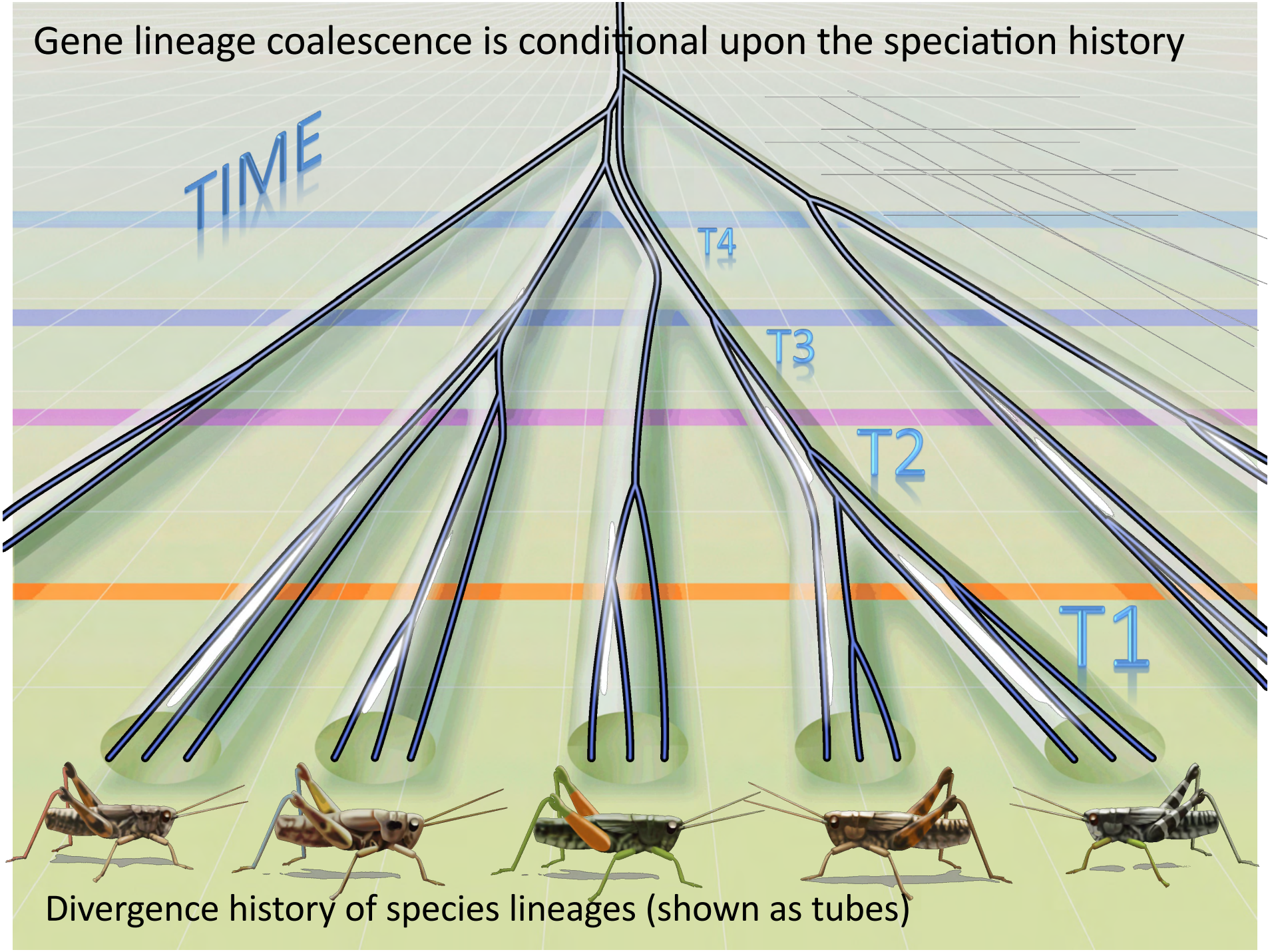
  "A group of organisms is exclusive if their loci coalesce more recently within the group than between any member of the group and any organisms outside the group''
  (Baum & Shaw 1995, p. 296).



- Disconnect between the time of speciation and when taxa reach reciprocal monophyly

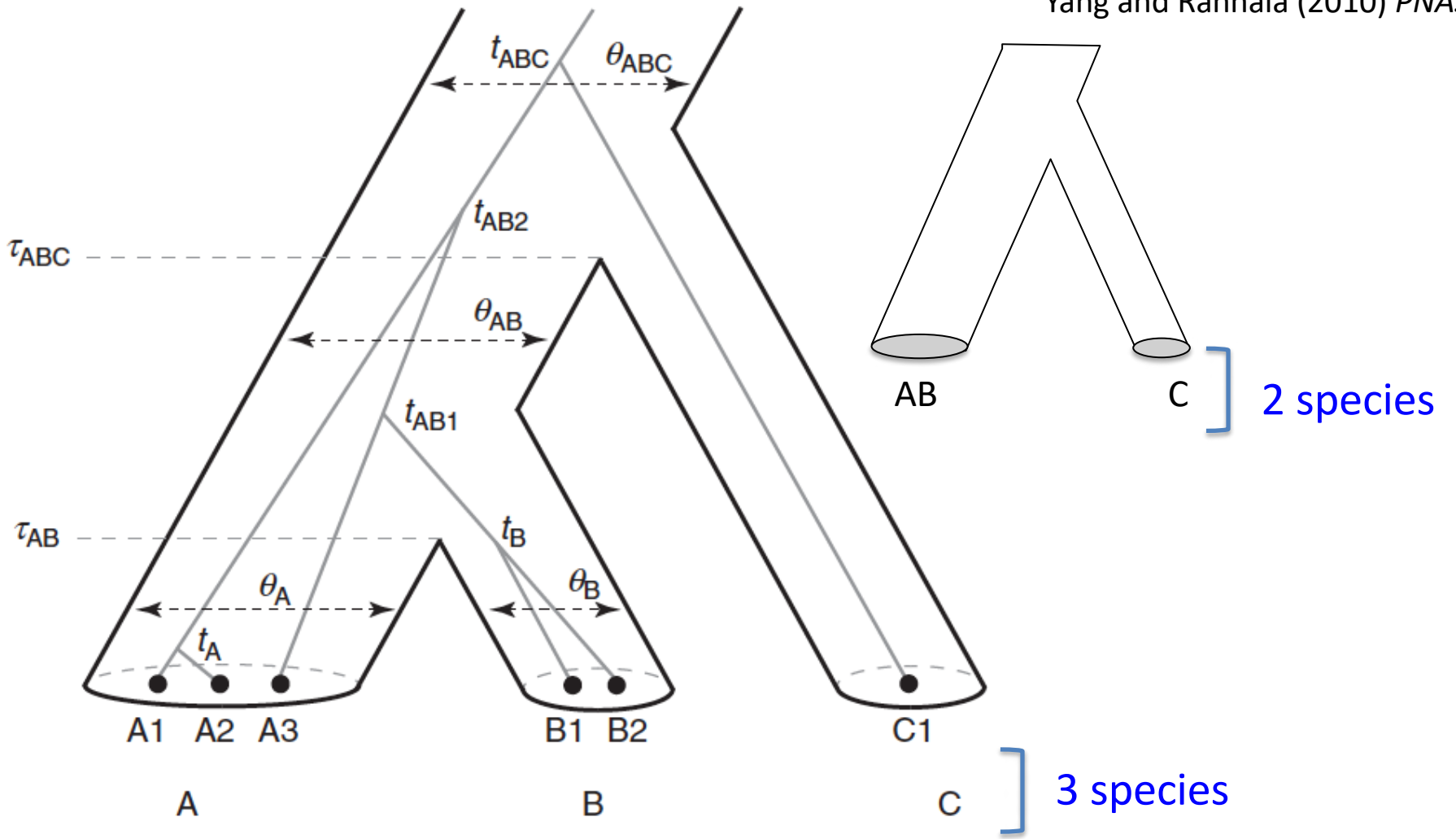
FIG. 1. Probabilities of observing reciprocal monophyly with time for populations that are genetically isolated. Curves are shown for a single mitochondrial locus and for samples of different numbers of nuclear loci.

Hudson and Coyne (2002) Evolution 56:1557–1565


Gene lineage coalescence is conditional upon the speciation history
TIME
T4
T3
T2
T1
Divergence history of species lineages (shown as tubes)

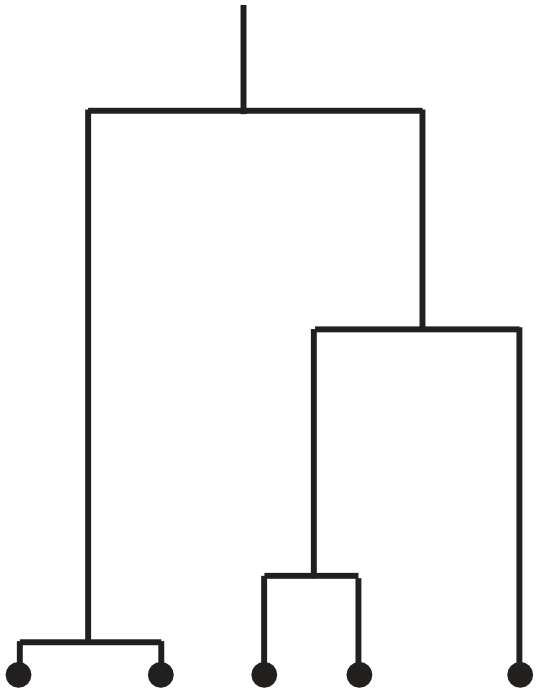
Multispecies coalescent (MSC) model used to evaluate different species delimitation hypotheses

Yang and Rannala (2010) *PNAS*



Different species delimitation hypotheses are formulated as competing statistical models and inferred from genetic data through Bayesian model selection (i.e., through calculation of posterior probabilities of a model), as in the popular program bpp

Delimitation with the Coalescent

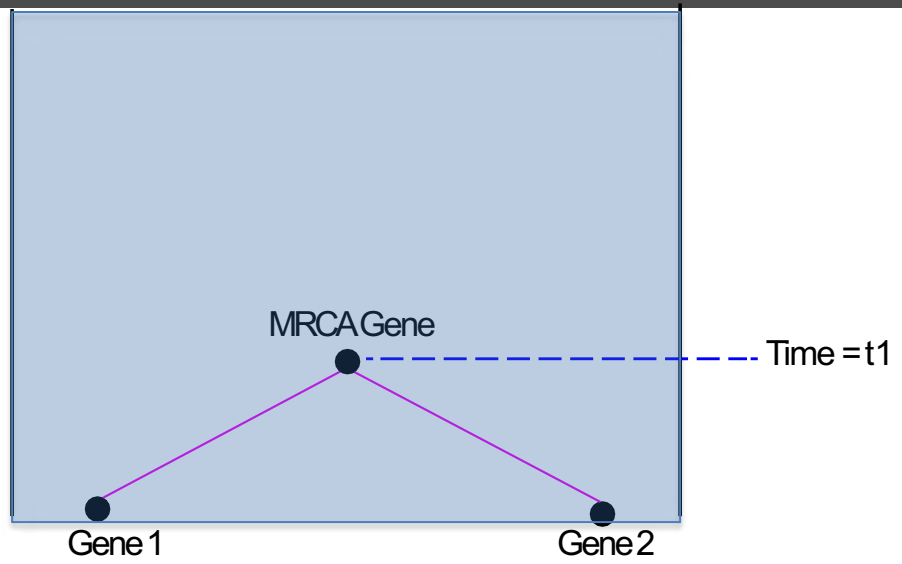


- Have a gene tree

Coalescent Theory Applications in a Nutshell

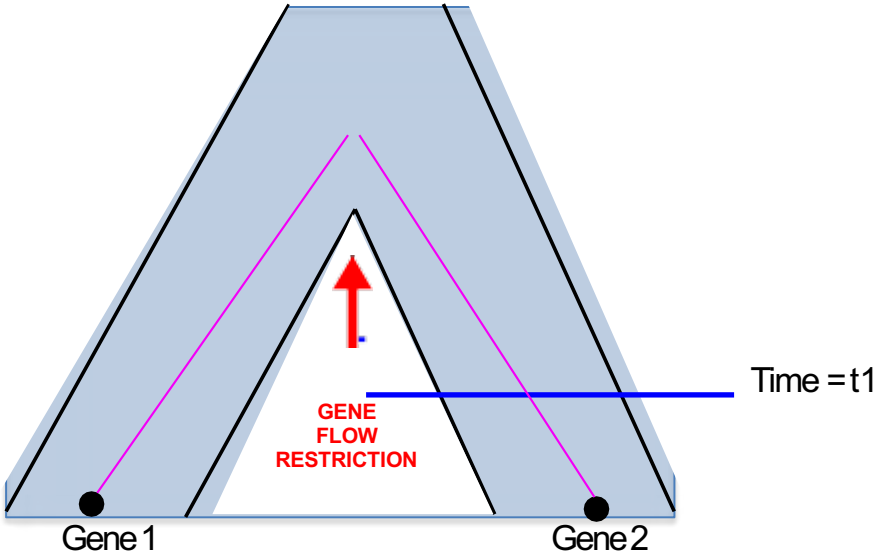
- Makes predictions about the *waiting time* between coalescence events based on population size and sample size.
- "coalescence events" (backward-time) = = "divergence events" (forward-time)
- Predictions are based on assumptions of particular properties of the population that the genes (or individuals having those genes) are evolving.
- Deviances in observed waiting times from that predicted can be used to make inferences about deviances in actual population properties from assumed Wright-Fisher panmictic population

How Does Structuring Change the Coalescent Times?



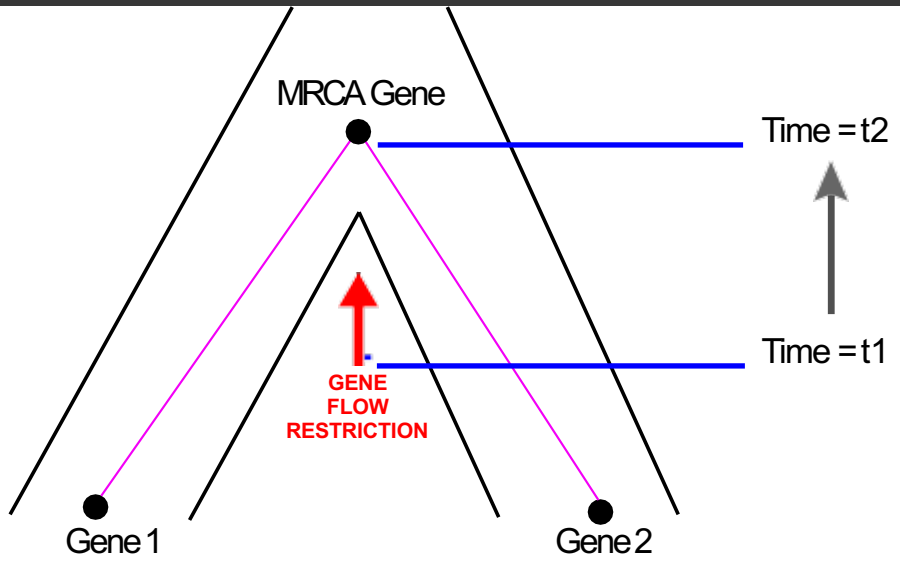
- Recall that the coalescent makes predictions about the timings to coalescence for genes sampled at random from a panmictic population.

How Does Structuring Change the Coalescent Times?



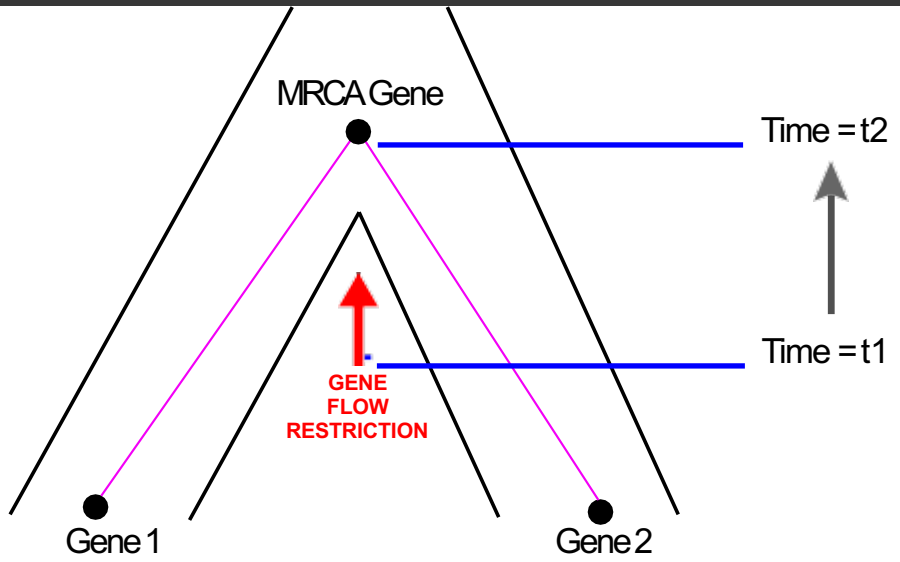
- Recall that the coalescent makes predictions about the timings to coalescence for genes sampled at random from a panmictic population.
- What happens if there are restrictions to panmixia?

How Does Structuring Change the Coalescent Times?



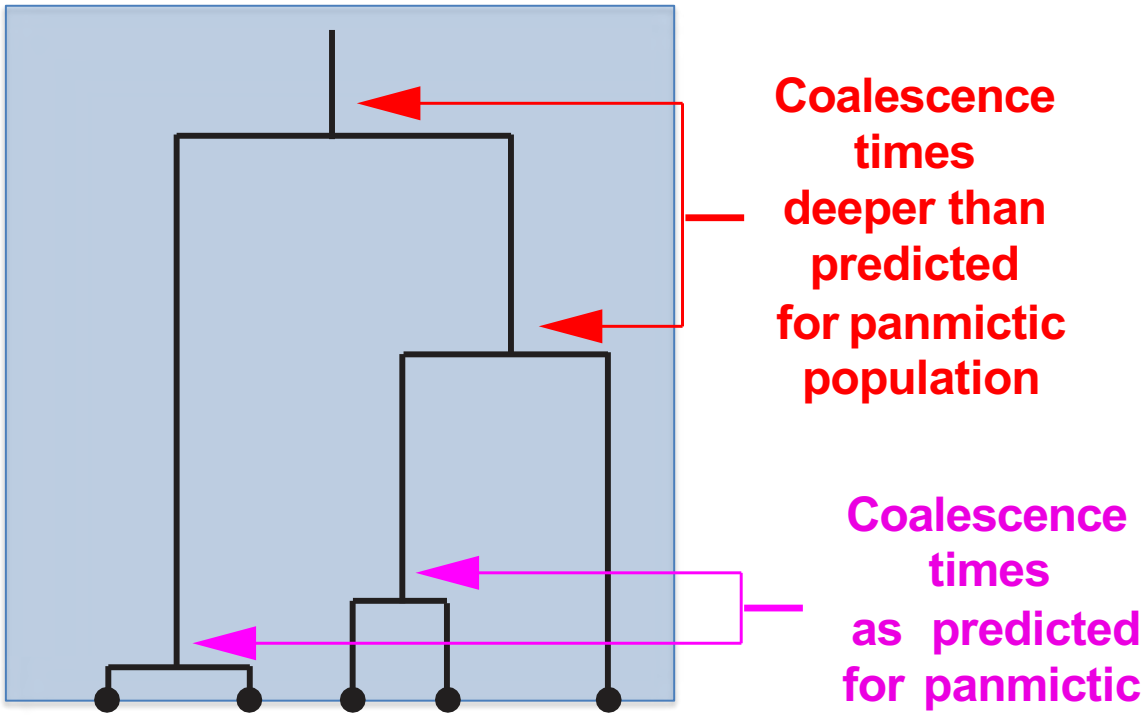
- Recall that the coalescent makes predictions about the timings to coalescence for genes sampled at random from a panmictic population.
- What happens if there are restrictions to panmixia?
- Then the timings to coalescent get *extended*

How Does Structuring Change the Coalescent Times?

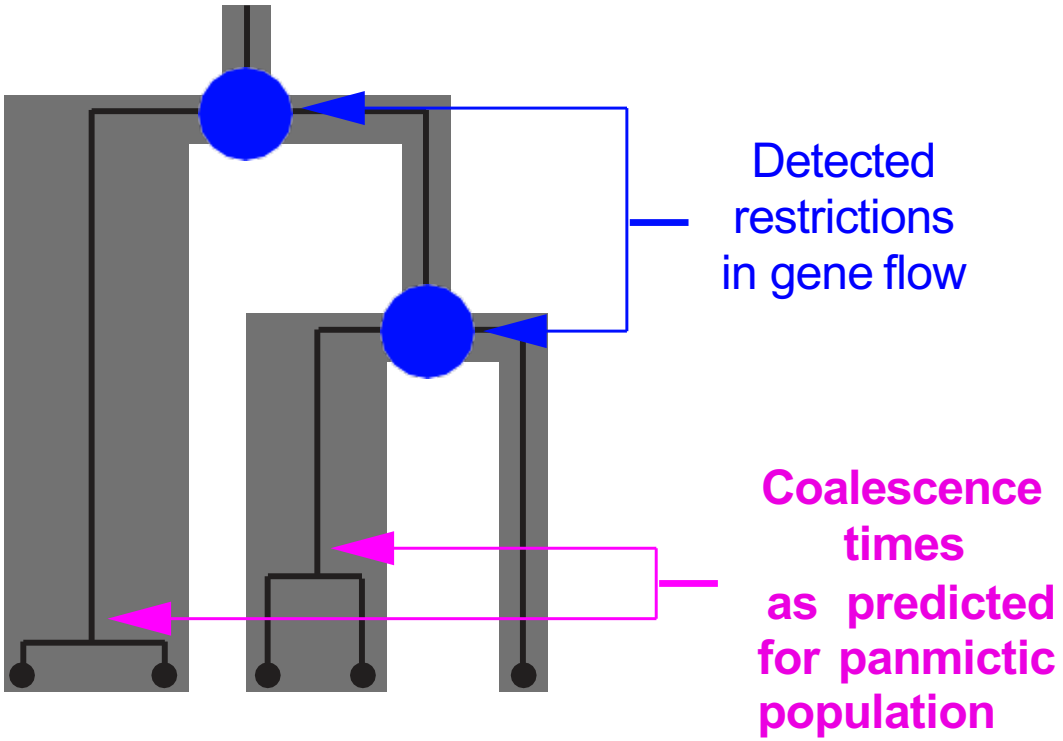


- Recall that the coalescent makes predictions about the timings to coalescence for genes sampled at random from a panmictic population.
- What happens if there are restrictions to panmixia?
- Then the timings to coalescent get *extended*
- This is the basis of the multispecies coalescent, MSC

Delimiting Units with the MSC



Delimiting Units with the MSC



- the MSC models the extensions in timings of coalescent events as disruptions of Wright-Fisher panmixia.
- It fits a "containing tree" to these disruptions (i.e., 3 species in this example)

Explosion of applications using the MSC for delimitation

Received: 15 September 2017 | Revised: 30 March 2018 | Accepted: 3 April 2018
DOI: 10.1111/1755-0998.12887

Bayesian species delimitation usin

Received: 28 July 2017 | Revised: 12 December 2017 | Accepted: 13 December 2017
DOI: 10.1111/mec.14486

WILEY MOLECULAR ECOLOGY RESOURCES

chine learning method for n genetic data

Yufeng Wu[1]

INVITED REVIEWS AND SYNTHESES

WILEY MOLECULAR ECOLOGY

Cryptic species as a window into the paradigm shift of the species concept

Cene Fišer[1] | Christopher T. Robinson[2,3] | Florian Malard[4]

EMPIRICAL EXAMPLE WITH LIZARDS OF THE *LIOLAEMUS DARWINII* COMPLEX (SQUAMATA: LIOLAEMIDAE)

Arley Camargo,[1,2] Mariana Morando,[3] Luciano J. Avila,[3] and Jack W. Sites, Jr.[1]
[1]Department of Biology & Monte L. Bean Museum, Brigham Young University, Provo, Utah 84602
[2]E-mail: arley.camargo@gmail.com
[3]CONICET-CENPAT, Boulevard Almirante Brown 2915, U9120ACD, Puerto Madryn, Chubut, Argentina

Bayesian species identification under the multispecies coalescent provides significant improvements to DNA barcoding analyses

ZIHENG YANG*† and BRUCE RANNALA†‡
*Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK, ‡College of Life Sciences, Beijing Normal University, Beijing 100875, China, ‡Department of Evolution and Ecology, University

Syst. Biol. 0(0):1–13, 2018

DOI:10.1093/sysbio/syy011

Comparison of Methods for Molecular Species Delimitation Across a Range of Speciation Scenarios

ARONG LUO[1,2,*], CHENG LING[3], SIMON Y. W. HO[2], AND CHAO-DONG ZHU[1,4]

[1]*Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China;* [2]*School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales 2006, Australia;* [3]*Department of Computer Science and Technology, College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China; and* [4]*College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China*

**Correspondence to be sent to: Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China; E-mail: luoar@ioz.ac.cn*

Simon Y. W. Ho and Chao-Dong Zhu contributed equally to this article.

E-mail: jacksonN@njhealth.org.

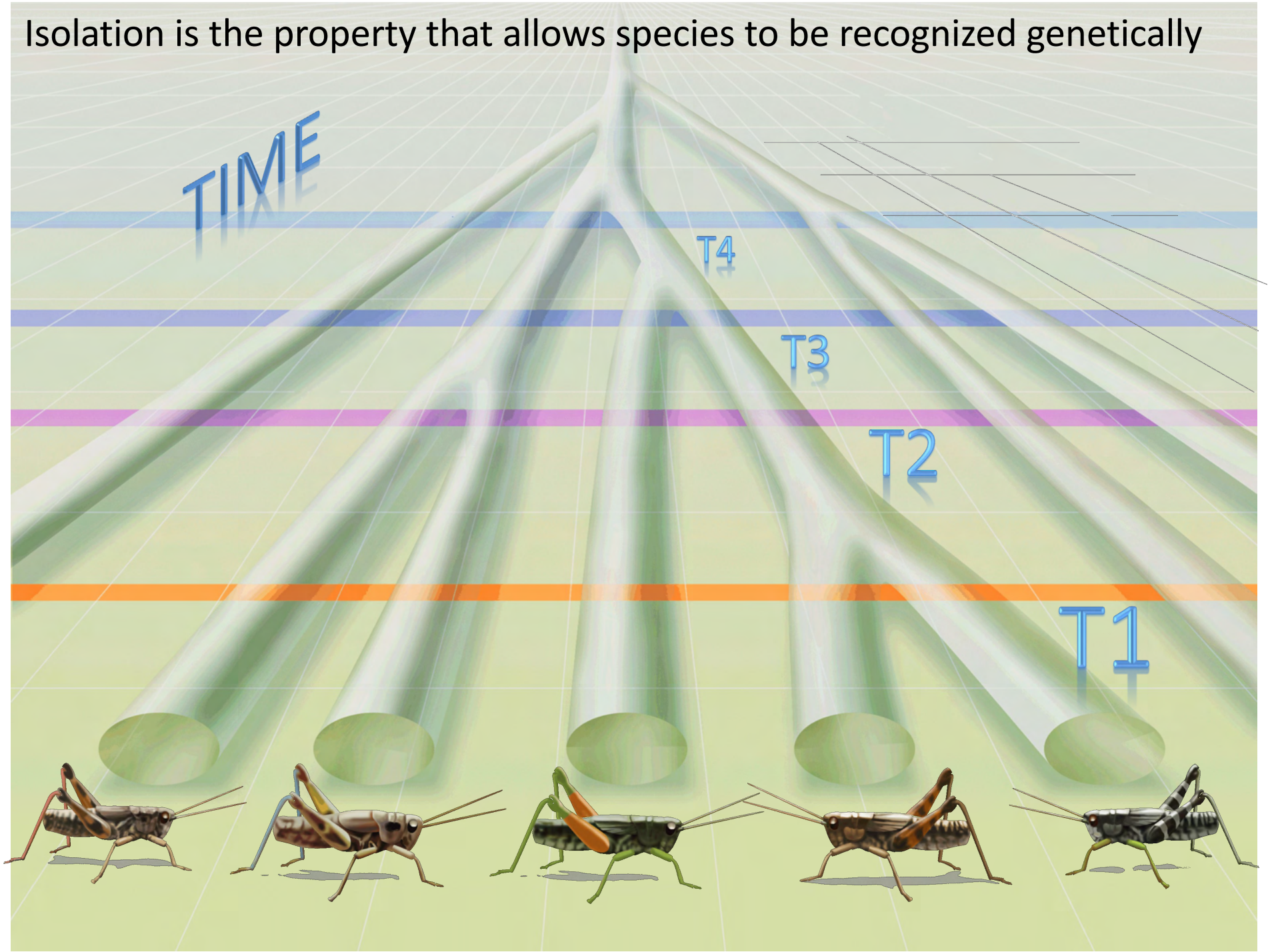
998
770
Advance Access Publication Date: 23 November 2014
Original Paper

ıment-free Bayesian for species delimitation ecies coalescent

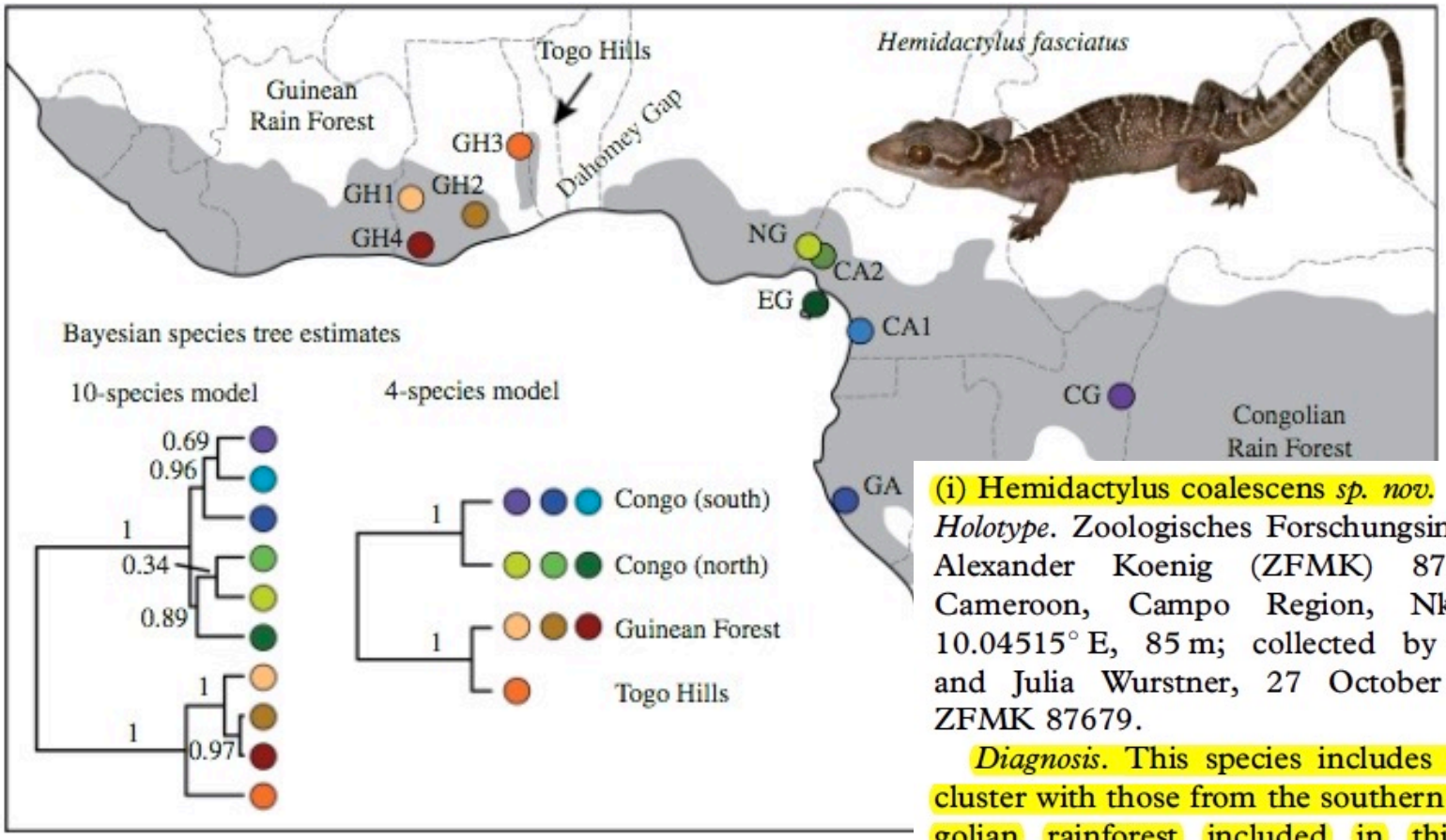
lin[1,2] and Bengt Oxelman[1,*]

al Sciences, University of Gothenburg, Box 461, SE 405 30 Göteborg, of Sciences, University of Dicle, 21280 Diyarbakir, Turkey

Isolation is the property that allows species to be recognized genetically



Model-based inference: probability of different hypotheses about species boundaries based on genetic data alone!



Leache & Fujita (2010) *Proc. R. Soc. B.*

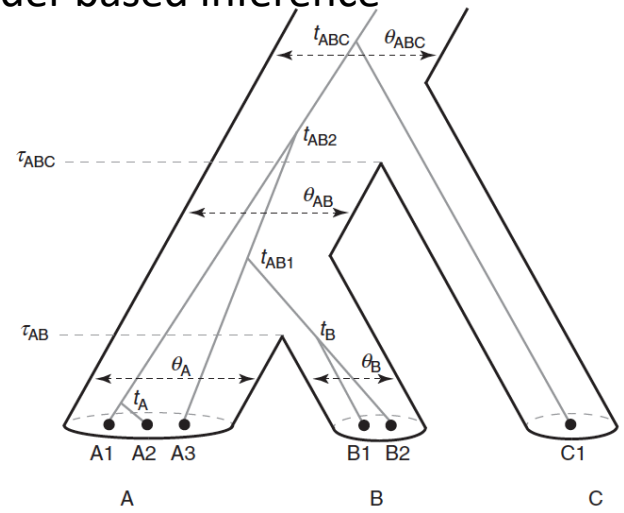
(i) Hemidactylus coalescens *sp. nov.*
Holotype. Zoologisches Forschungsinstitut und Museum Alexander Koenig (ZFMK) 87680, adult male; Cameroon, Campo Region, Nkoelon, 2.3972° N, 10.04515° E, 85 m; collected by Michael F. Barej and Julia Wurstner, 27 October 2007. Paratype = ZFMK 87679.

Diagnosis. This species includes all populations that cluster with those from the southern portion of the Congolian rainforest included in this study (southern Cameroon, Gabon and Congo), with strong support in the Bayesian species delimitation model.

Etymology. This species is named after the coalescent process used to delimit the species.

Pros of species delimitation under MSC

Model-based inference



- Can delimit species before reciprocal monophyly of alleles or fixed differences

  Knowles & Carstens (2007) *Syst. Biol.*

- Still detects lineages under low gene flow

  Zhang et al. (2011) *Syst. Biol.*

- Accuracy of species delimitation to sampling can be evaluated (i.e., will more data change status)

- De facto standardization for objectively delimiting taxa (i.e., data treated equally among all living things and avoid subjectiveness of what characters to measure) Fujita et al. (2012) *TREE*

- Can take into account uncertainty in gene trees

  Yang & Rannala 2010

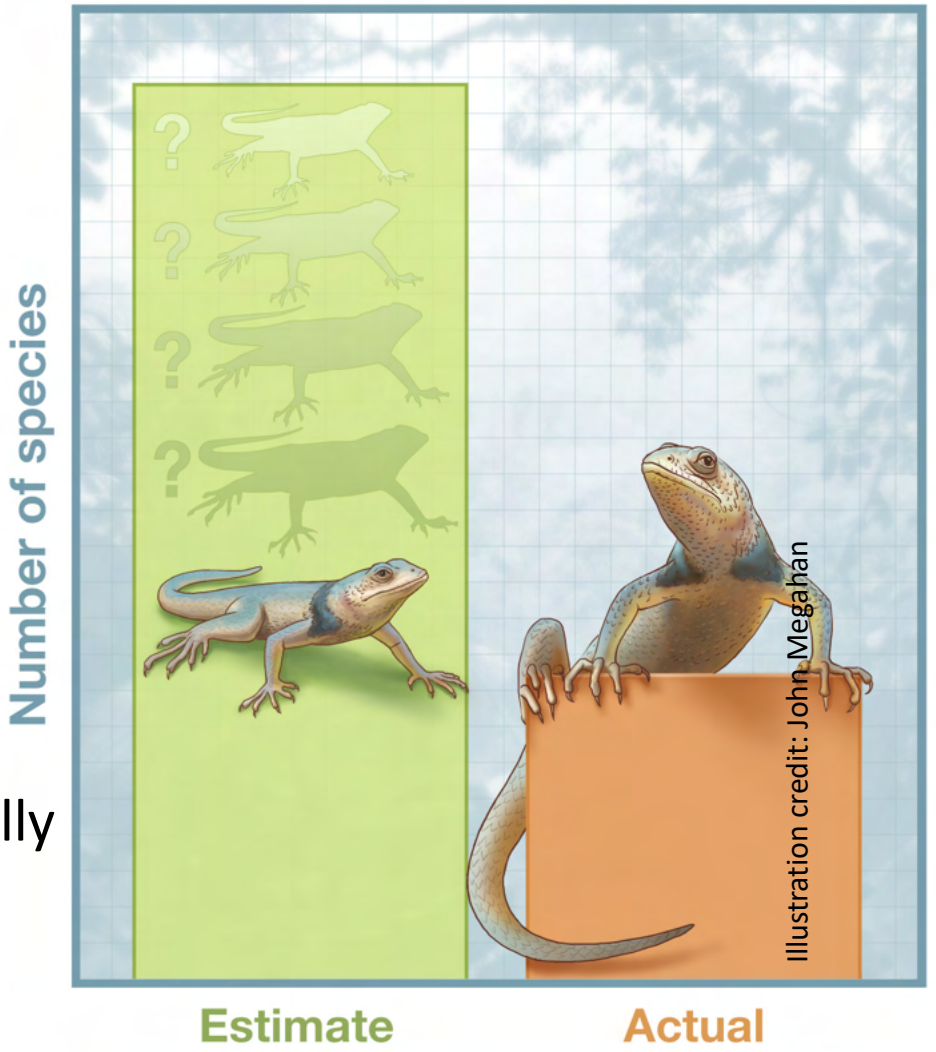
the multispecies coalescent for delimiting species boundaries

- All models are flawed...

some are more or less useful.

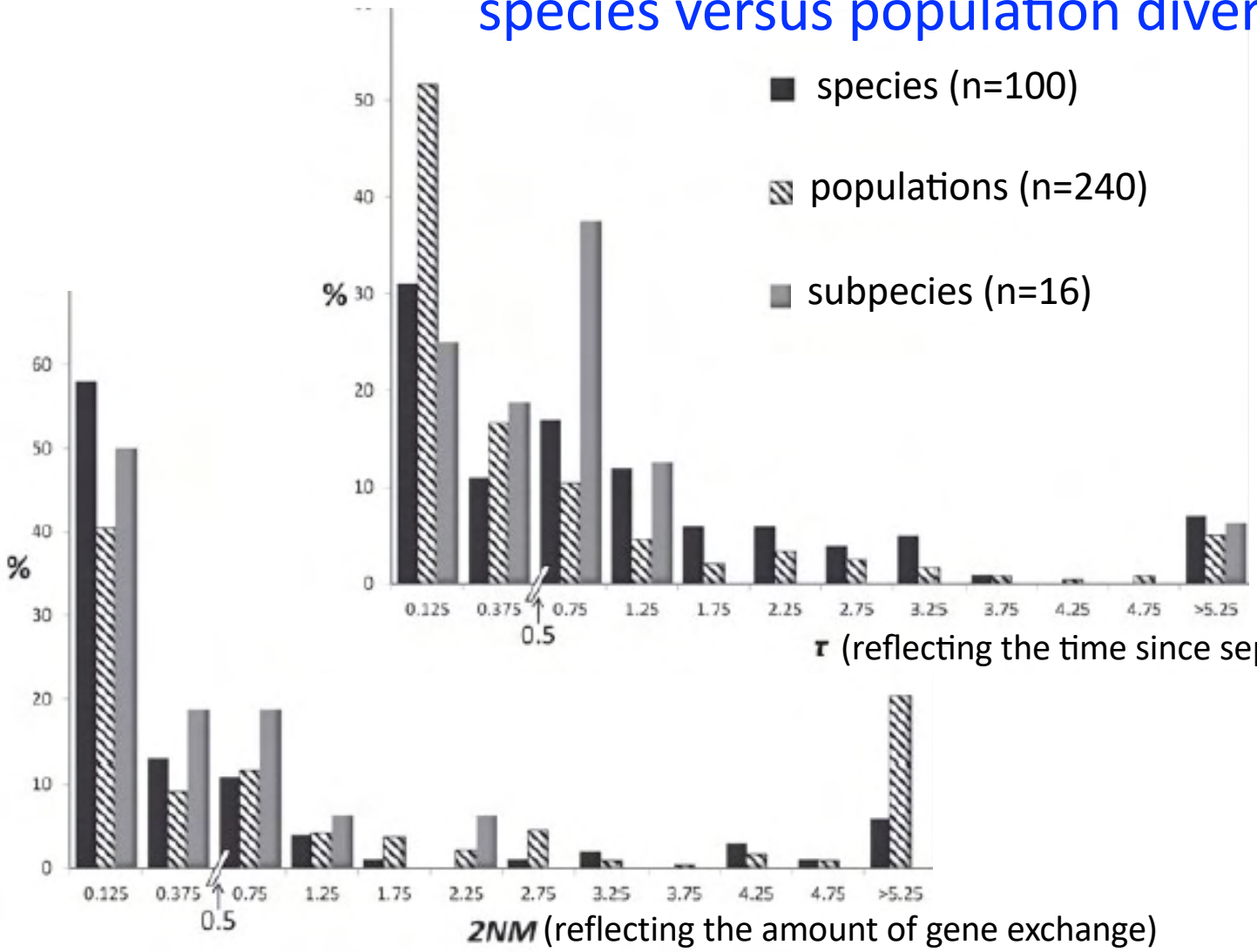
Isolation is the property that allows species to be recognized genetically

Isolation is the property that allows populations to be recognized genetically



Sukumaran & Knowles (2017) *PNAS*

No genetic distinction between species versus population divergence



τ (reflecting the time since separation)

$2NM$ (reflecting the amount of gene exchange)

Pinho and Hey(2010) *Evolution*

The MSC dominates the field, but…

How bad is the confounding of population verus species divergence?

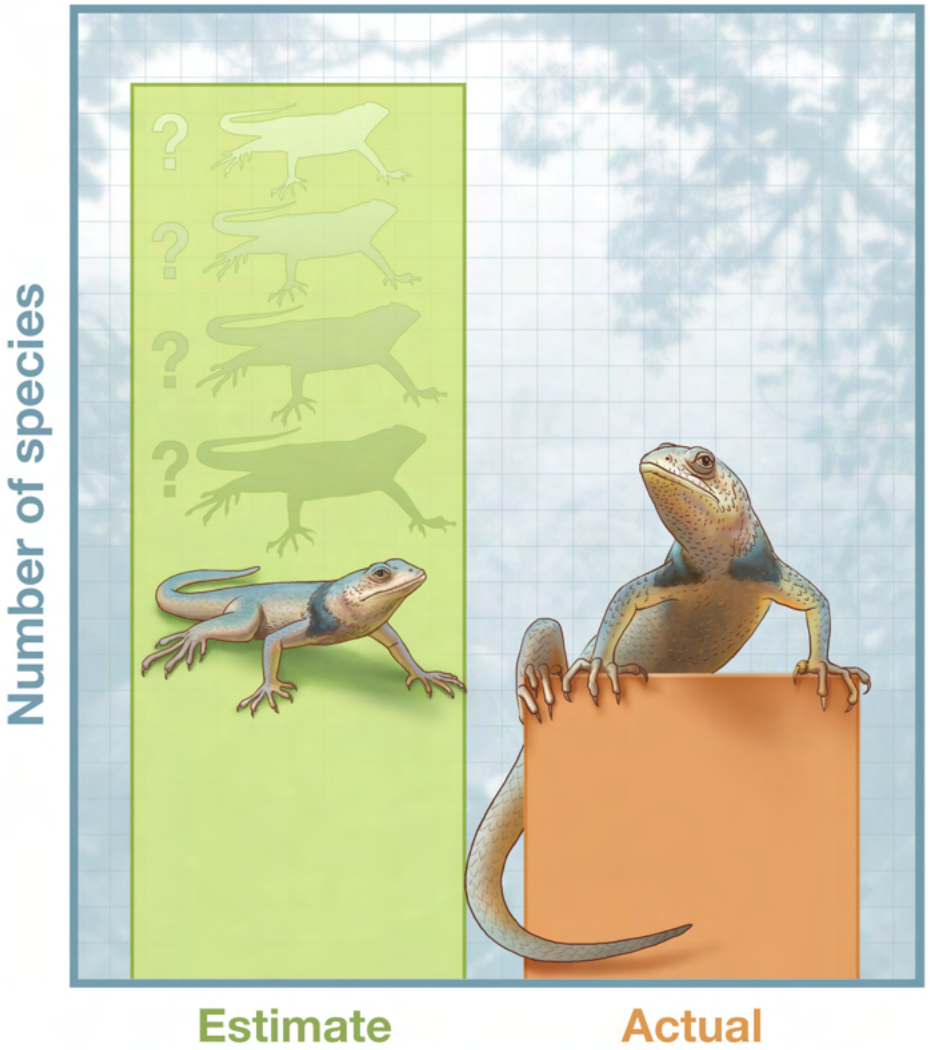


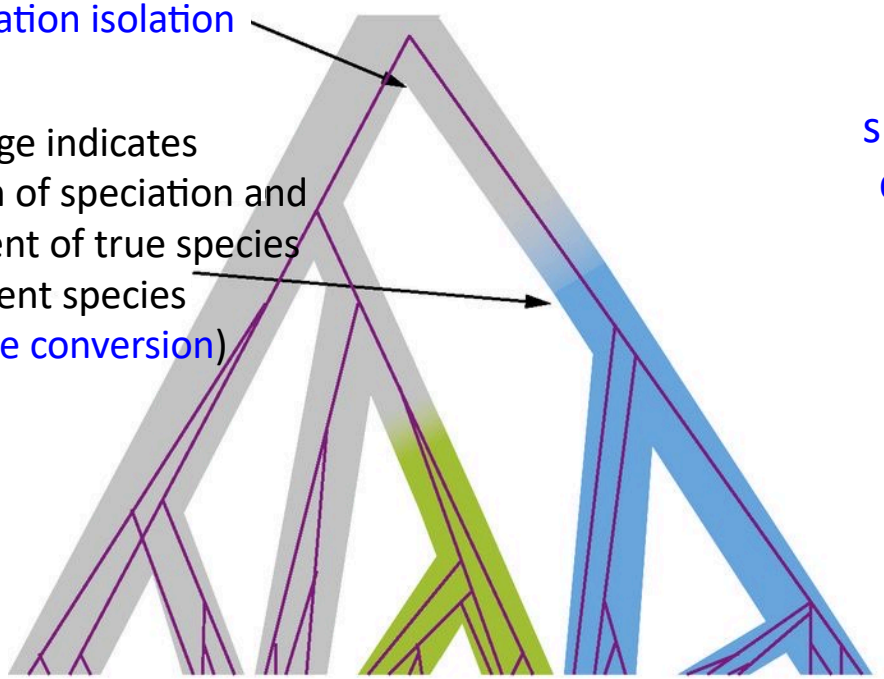
Illustration credit: John Megahan

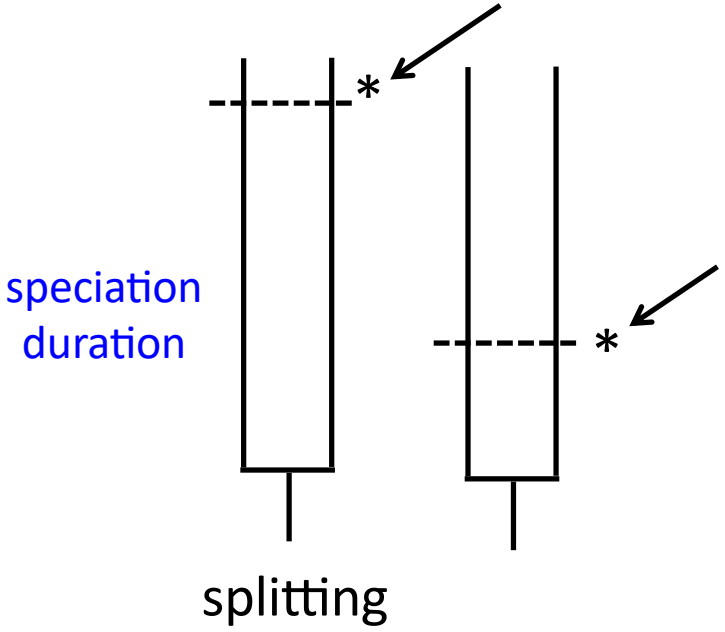
Sukumaran & Knowles (2017) *PNAS*

Simulate data to account for differences in speciation duration (i.e., speciation is a protracted process; it is not instantaneous)

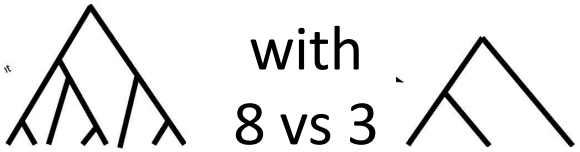
Splitting events such as this are initiation of speciation through, e.g., population isolation

Color change indicates completion of speciation and development of true species from incipient species (i.e., lineage conversion)



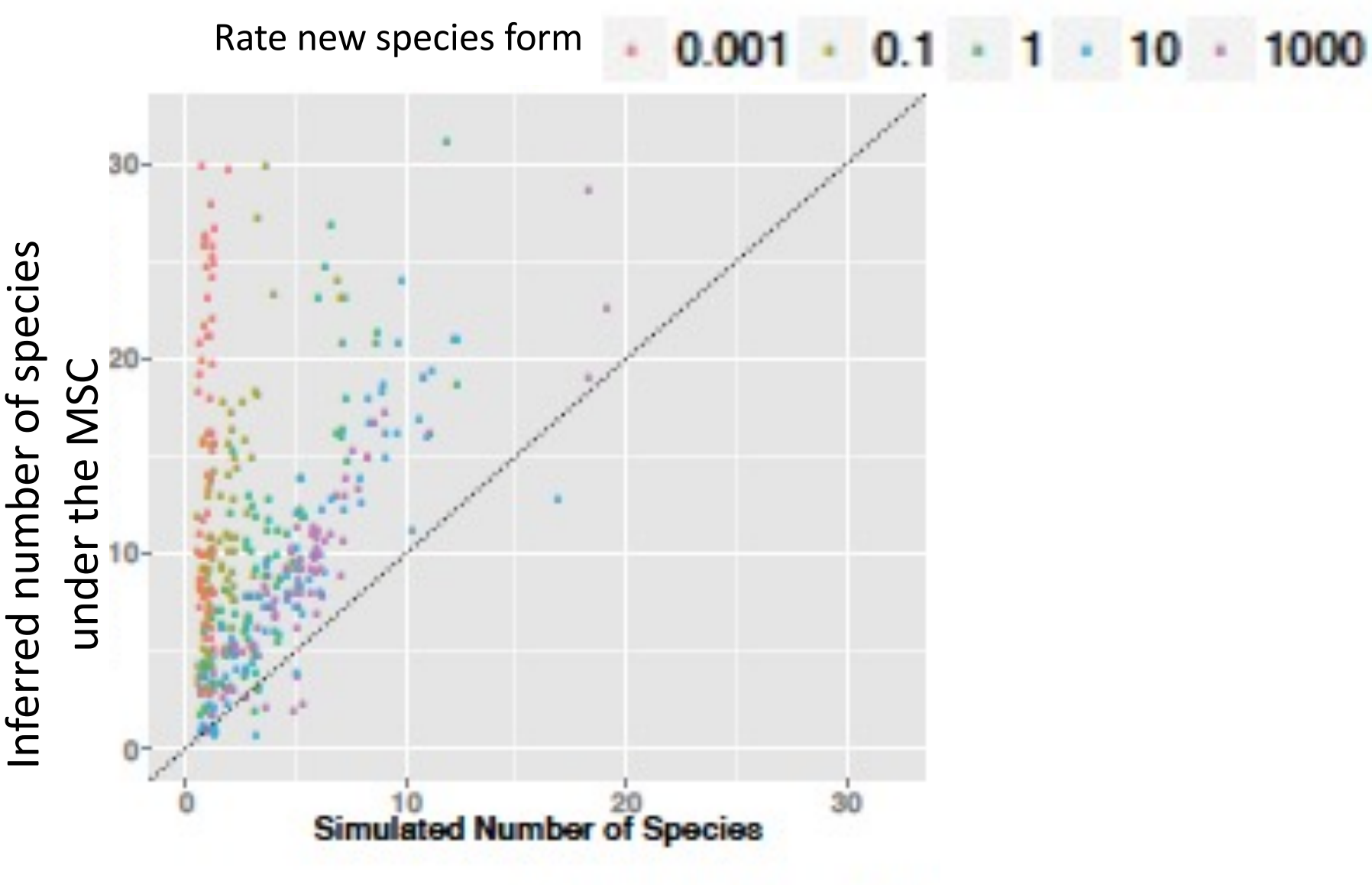


Does the MSC accurately delimit species?



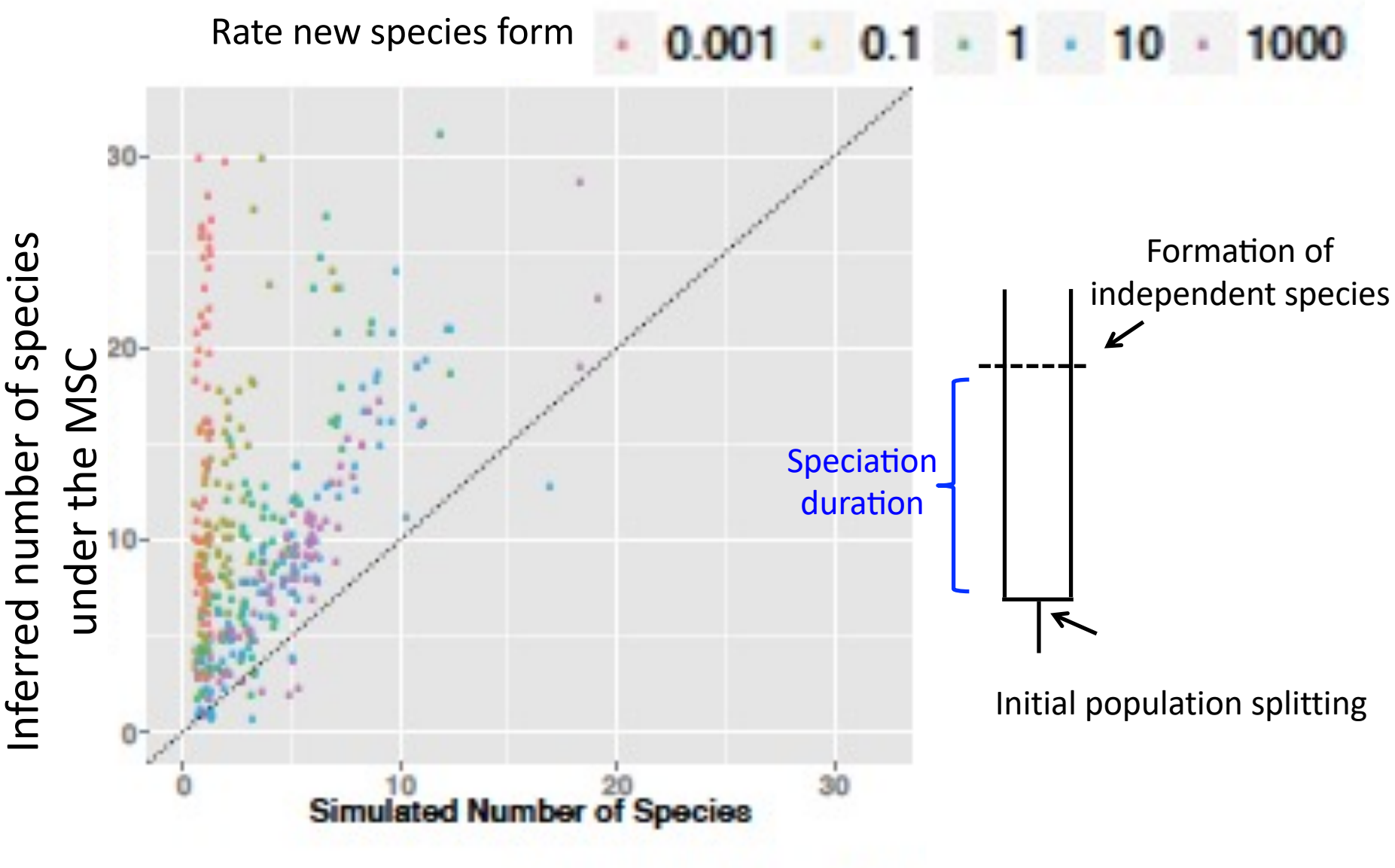
Sukumaran & Knowles (2017) *PNAS*

overestimation of species richness under the MSC



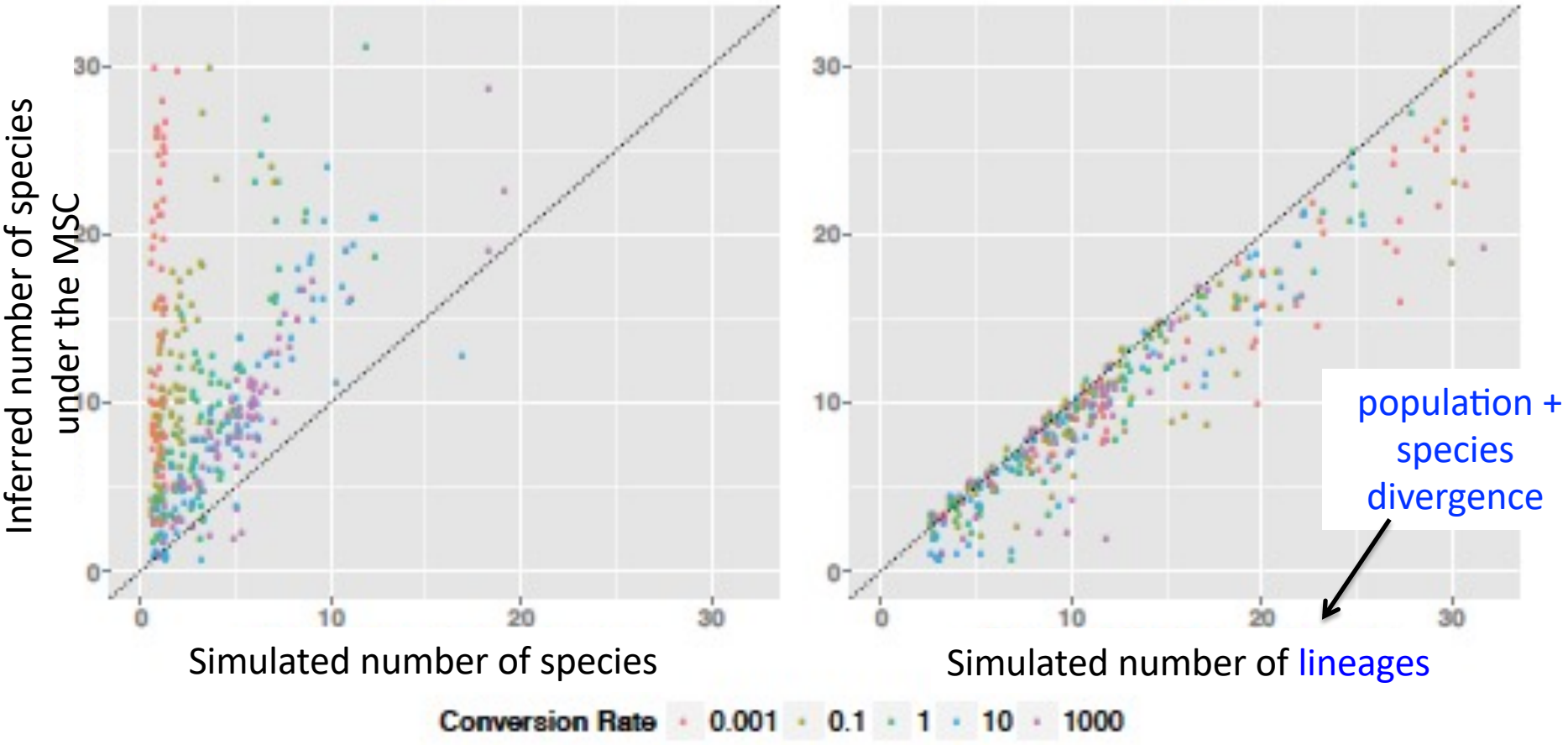
Sukumaran & Knowles (2017) *PNAS*

Degree of overestimation of species richness under the MSC depends on the speciation duration



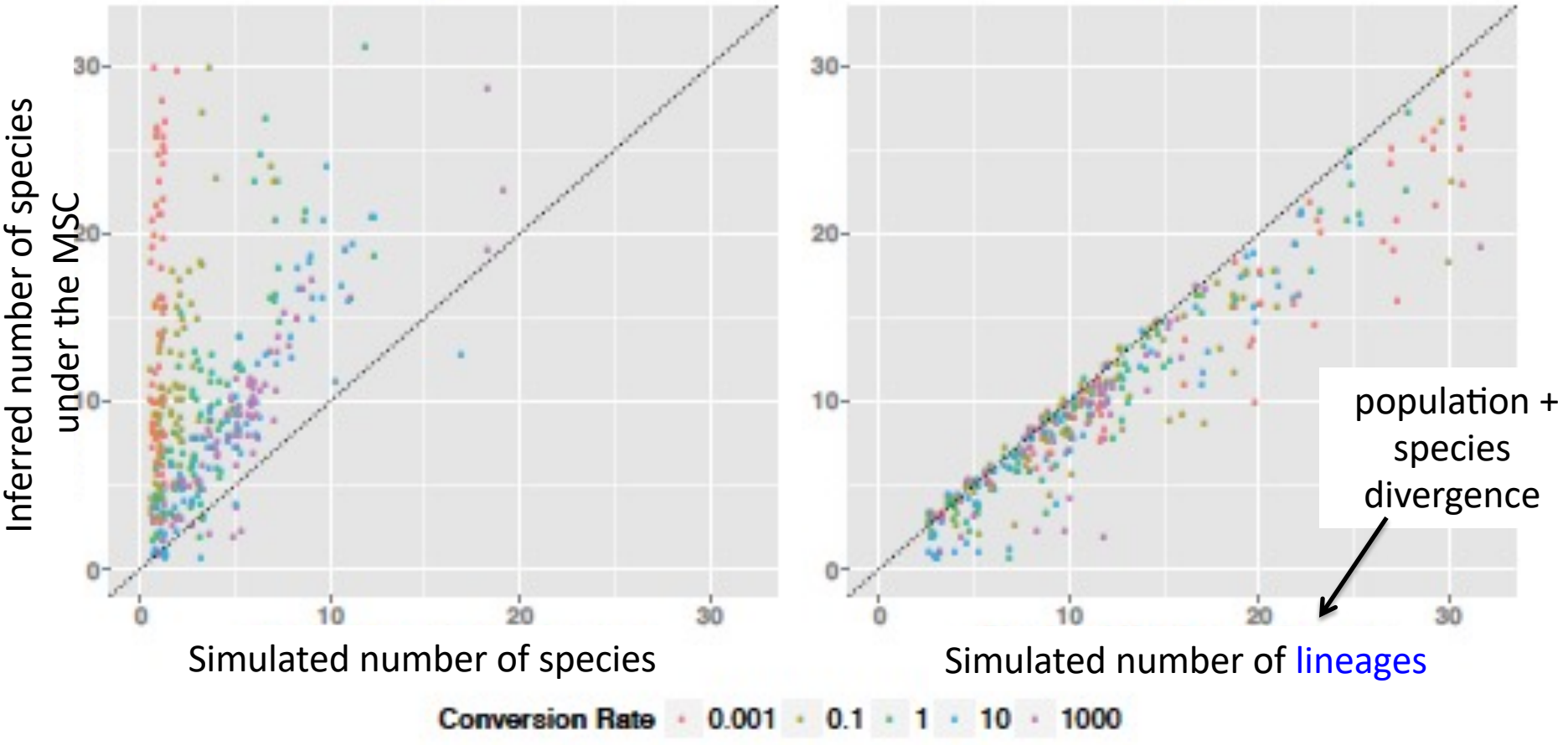
Sukumaran & Knowles (2017) *PNAS*

MSC powerful model for detecting genetic structure



Sukumaran & Knowles (2017) *PNAS*

MSC powerful model for detecting genetic structure



HOWEVER, the MSC is not capable of distinguishing genetic structure due to population versus species divergence

Sukumaran & Knowles (2017) *PNAS*

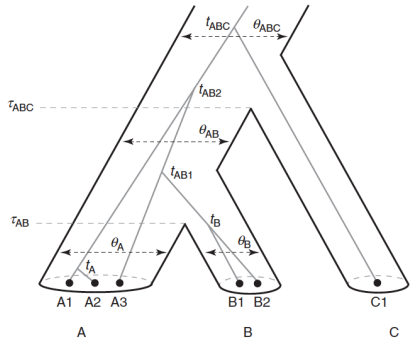
Problems with species delimitation under the MSC

- MSC detects structure – not species

Sukumaran & Knowles (2017) *PNAS*

(different statistical delimitation methods all based on the MSC, which also means seeking consensus across methods is not a good way to fail)

See Rannala (2015) Current Zoology 61, 846-853



- "Robustness" to lineage detection with low levels of gene flow is not the same as accurate species delimitation

- Sensitivity to sampling (e.g., sparse geographic coverage over-splits species)

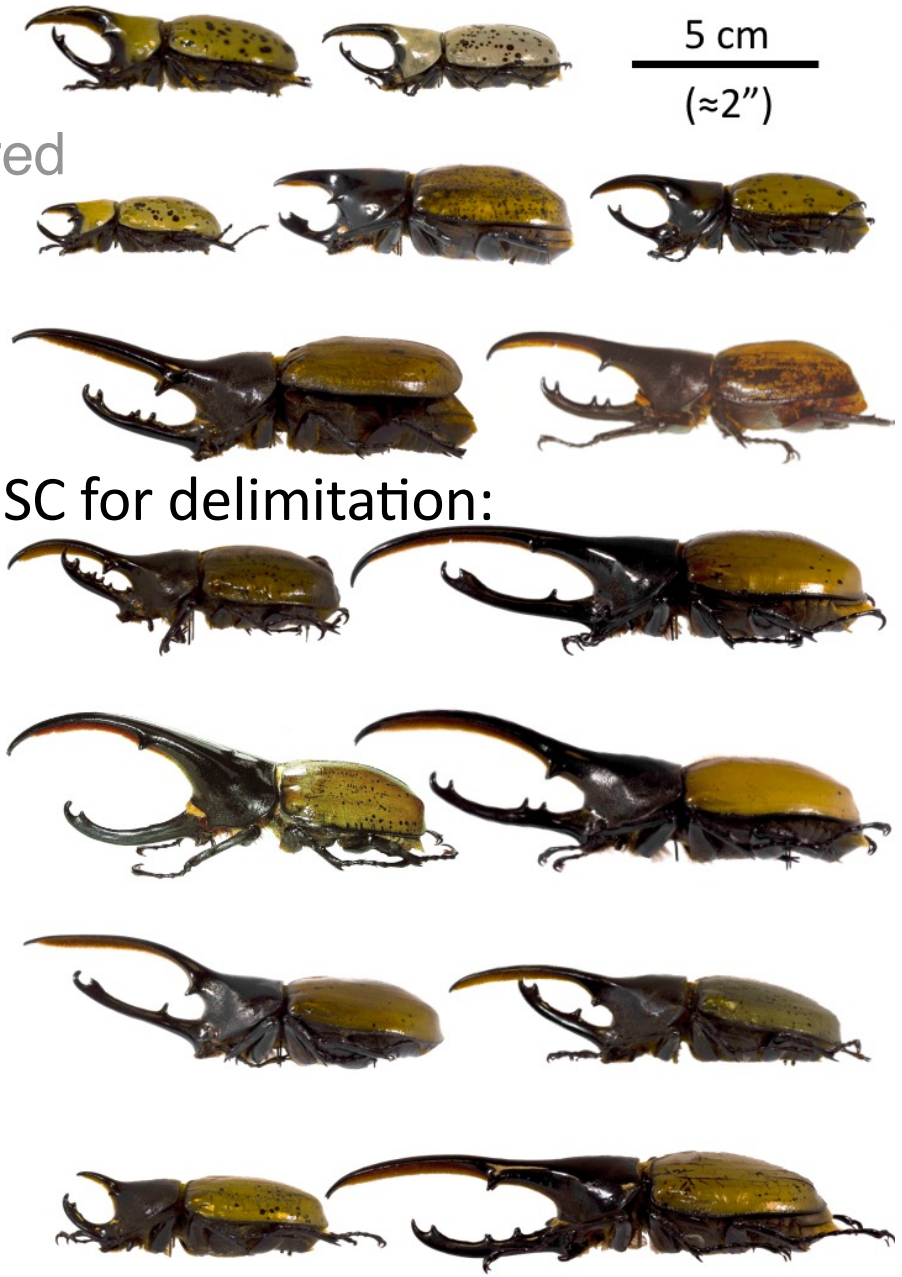
Chambers & Hillis (2020) *Syst. Biol.*

- MSC is not a de facto standardization for delimiting taxa: degree of over estimation varies depending on speciation process

Sukumaran & Knowles (2017) *PNAS*

Model-based delimitation based on the MSC:

- Erroneous species boundaries inferred

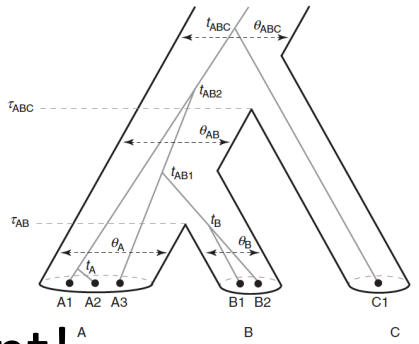


Strong assumption of relying on the MSC for delimitation:

- All genetic structure = species

Accurate species delimitation cannot be achieved with the MSC alone

Delimitation under the MSC:
genetic structure = species



- Don't run MSC and add a caveat – what's the point!
- STOP reporting about all this "cryptic" diversity

Ad hoc heuristics to interpret results from MSC-based models for delimitation

- Genealogical sorting index[1]: $2\tau/\theta$
  (i.e., population divergence time relative to the population size)
  Cummings et al. (2008) Evolution 62-9: 2411–2422

- use population divergence parameters (e.g., thresholds for divergence levels[2] or lots of migration[1])

[1]Jackson et al. (2018) Syst. Biol.

[1]Leache et al. (2018) Syst. Biol.

[2] **SPEEDEMON** Jordan Douglas and Remco Bouckaert. Quantitatively defining species boundaries with more efficiency and more biological realism. Communications Biology 5, 755 (2022). doi:10.1038/s42003-022-03723-z

These heuristics do not validate the MSC itself for species delimitation

Using diverse sources of data for inferring species boundaries has a long systematic tradition, but not with model-based inference.

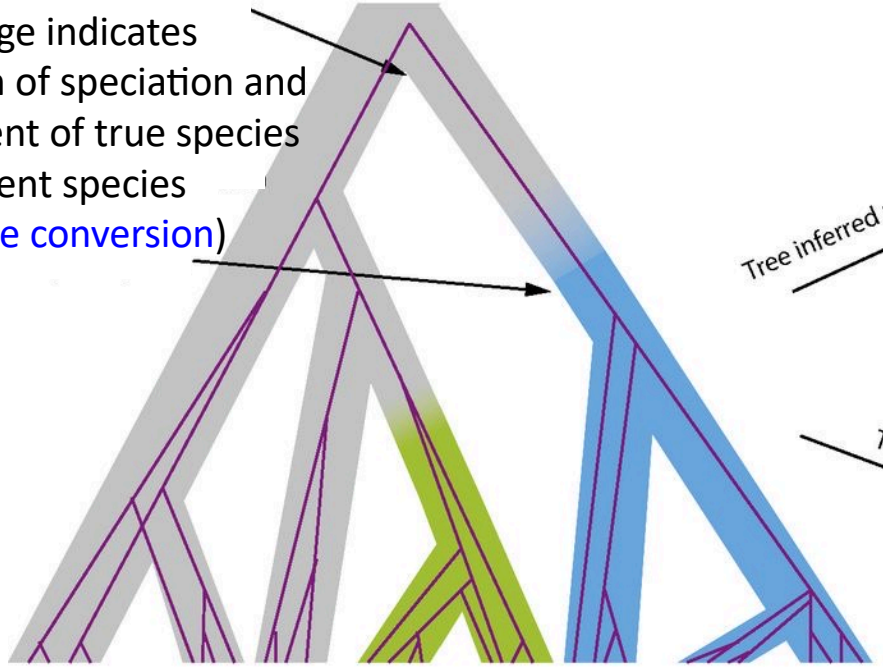
Joint analysis of morphology and genetic data!

Solis-Lemus C, Knowles LL, Ané C (2014) Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492-507.

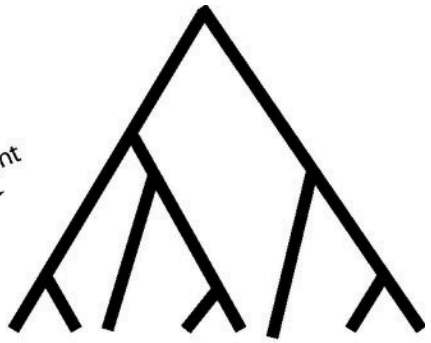
The future of genetic delimitation models will be those that bring speciation into the multispecies coalescent framework

Splitting events such as this are initiation of speciation through, e.g., population isolation

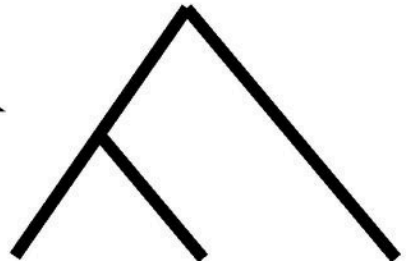
Color change indicates completion of speciation and development of true species from incipient species (i.e., lineage conversion)



Tree inferred under the multispecies coalescent

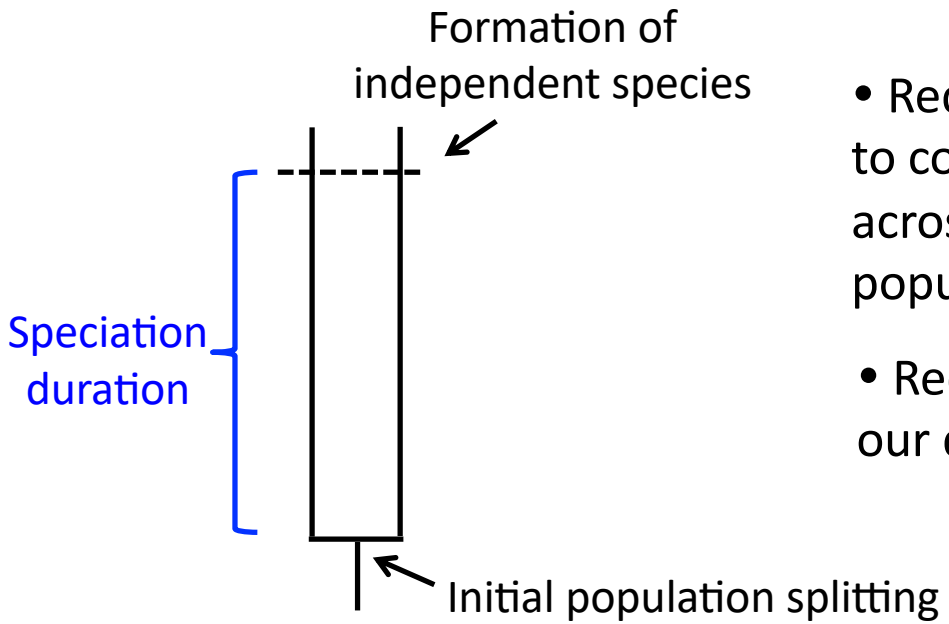


True speciation history



See e.g.: Sukumaran, Holder, Knowles (2021) *PLoS Comput Biol*

• We model the formation of new population lineages and their subsequent development into independent species modeled as separate processes



• Requires a shift in our sampling efforts: need to collect individuals from multiple populations across a species range to generate a combined population and species tree

• Requires incorporating well defined species in our delimitation analysis

Software: *DELINEATE*

https://github.com/jeetsukumaran/delineate

Evolutionary applications of model-based analyses:

(i) Inferring species boundaries (aka species delimitation)

(ii) Phylogenetic inference (and beyond the species tree)

(iii) Biogeographic study

(iv) Phylogeography

(v) Adaptive evolution

(vi) Speciation

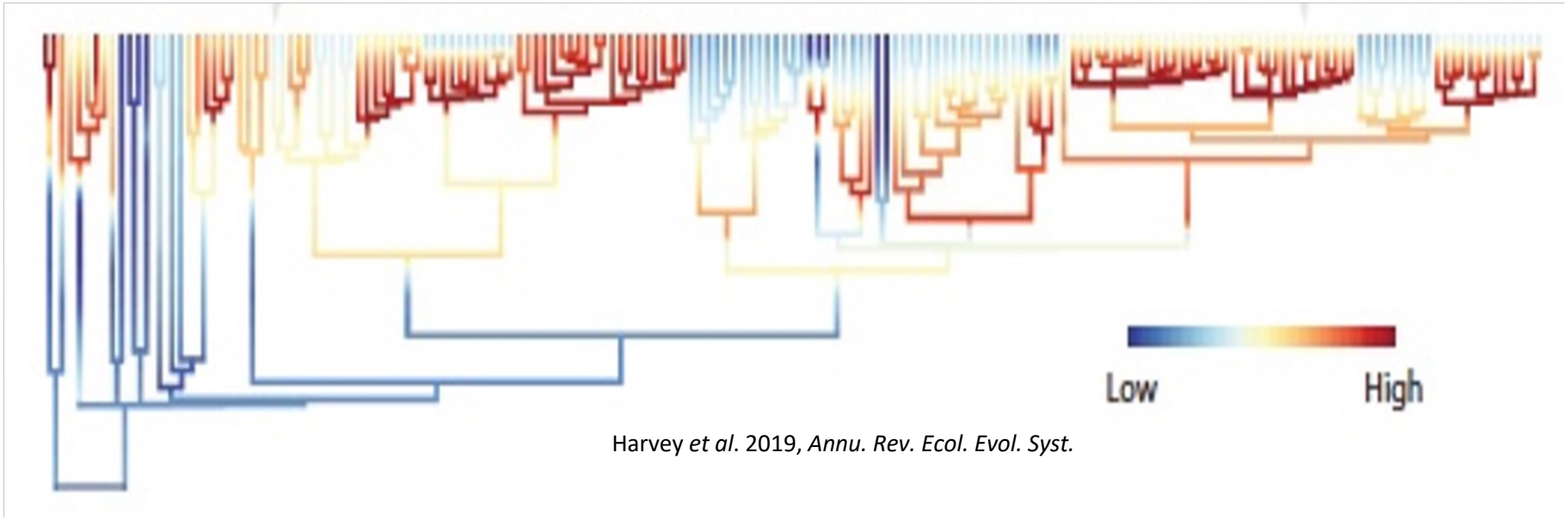
Controls on diversification

L Lacey Knowles

Describing diversity patterns versus understanding process

Macroevolutionary parameters of diversification for describing patterns of species diversity across space, time, and clades are NOT sufficient for understanding the processes generating those patterns.

Differences in diversification rate across phylogeny

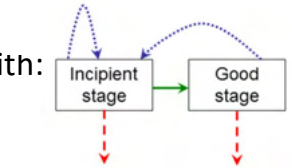


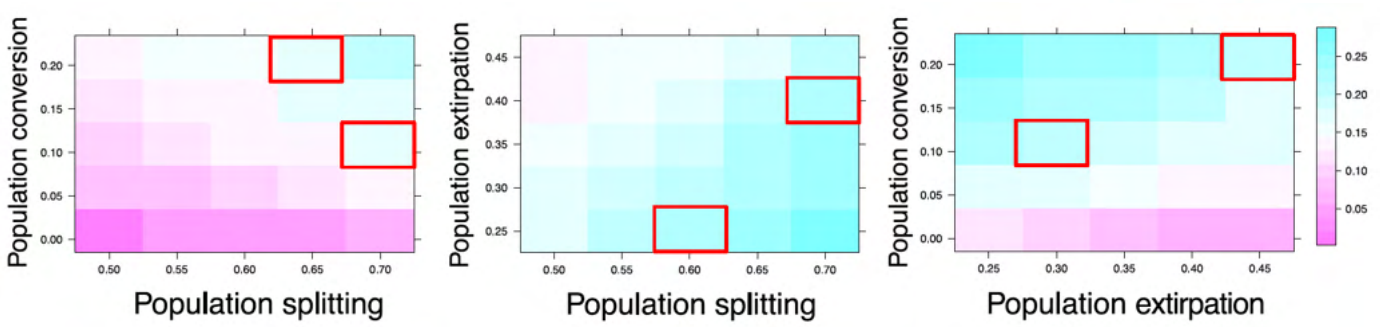
Harvey *et al*. 2019, *Annu. Rev. Ecol. Evol. Syst.*

- Different processes produce same rate of diversification
- Speciation is not an instantaneous process

Incomplete pictures of the controls on diversification from species lineages alone.

Simulated data sets with:

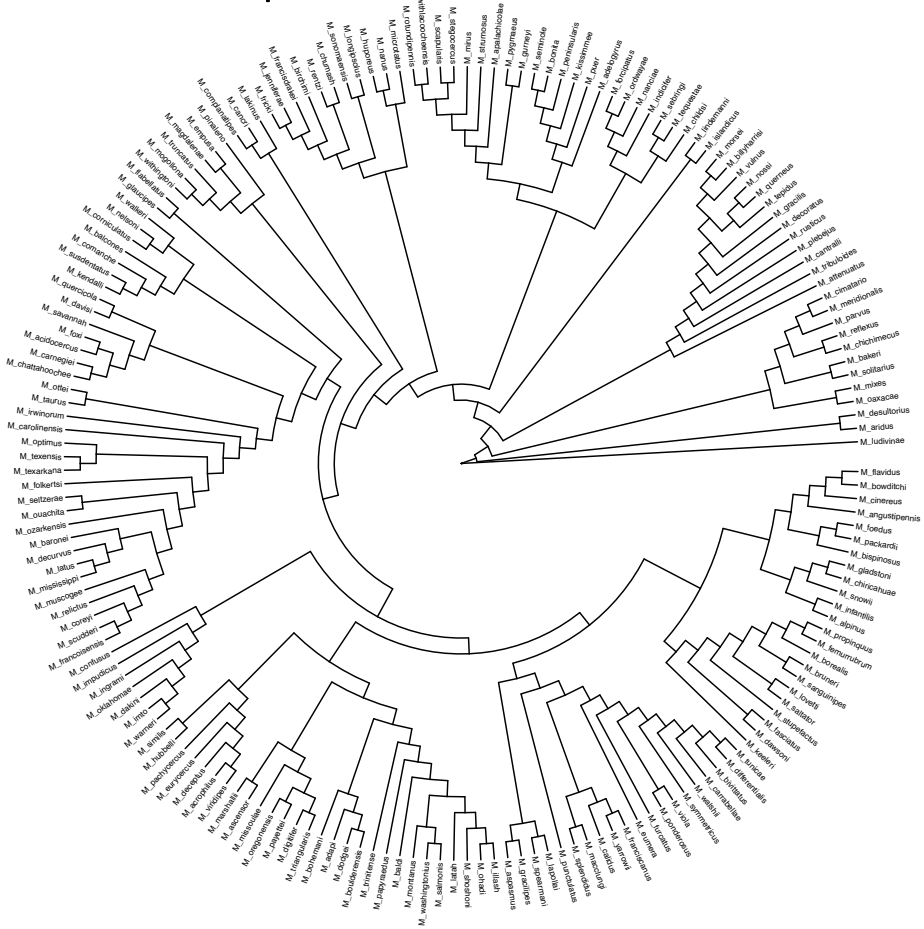




Same λ and μ estimated under B-D model

(from Li et al. 2018)

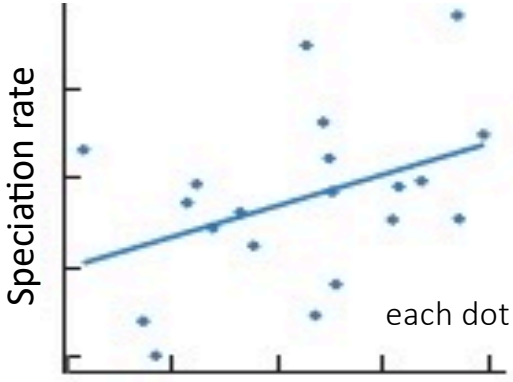
Species-level tests of the controls of diversification

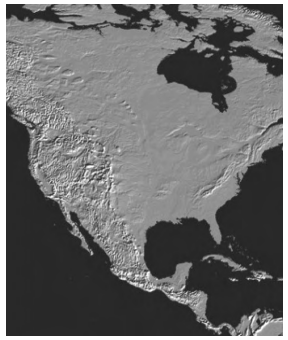


- 185 *Melanoplus* grasshopper species

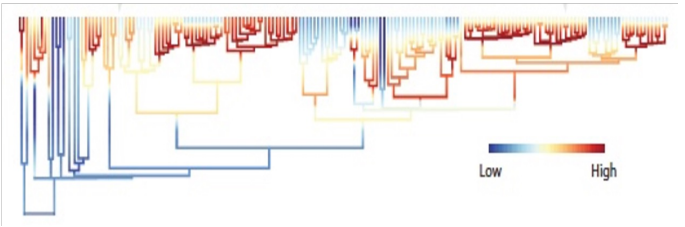
Cause of speciation rate differences across species lineages

Tests of associations between speciation rate and factors hypothesized to affect the formation of new species to explore the cause of differences in speciation rates.

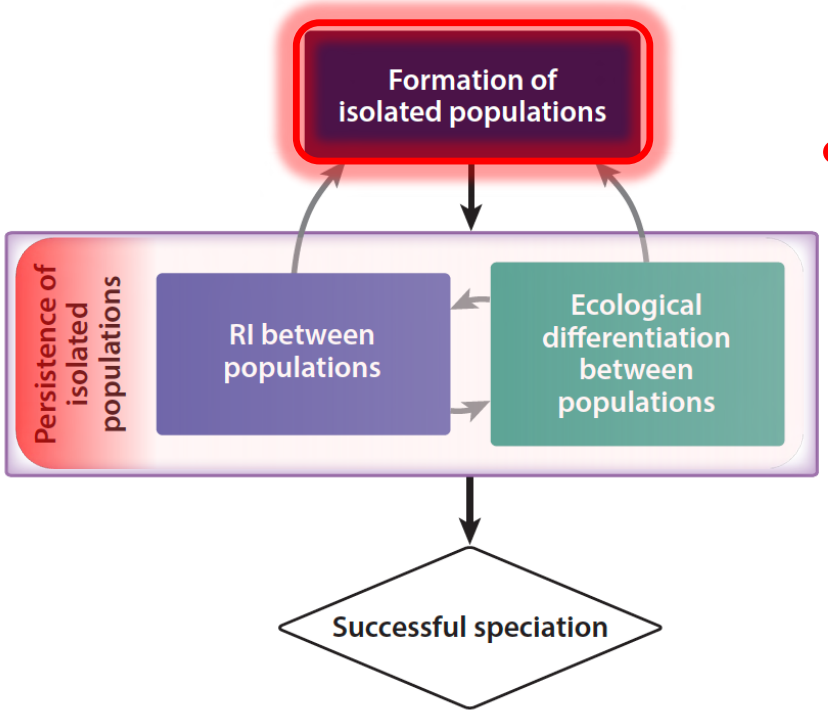




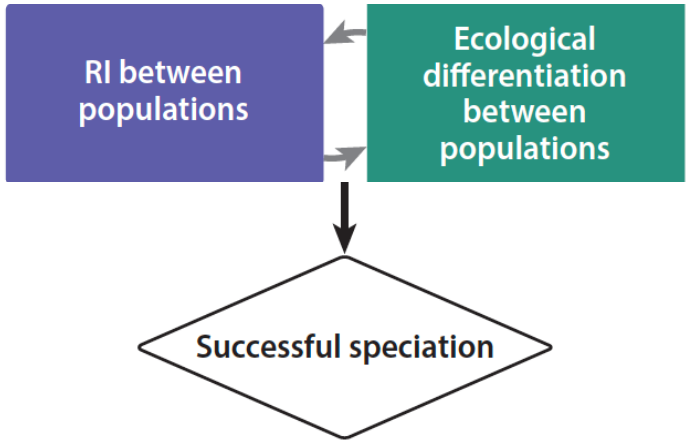
Differences in diversification rate across phylogeny



Controls on diversification

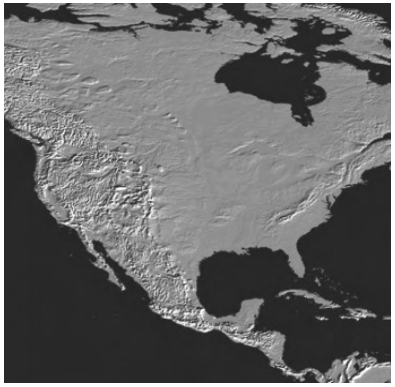


- Opportunities for speciation



- Fate of incipient divergences

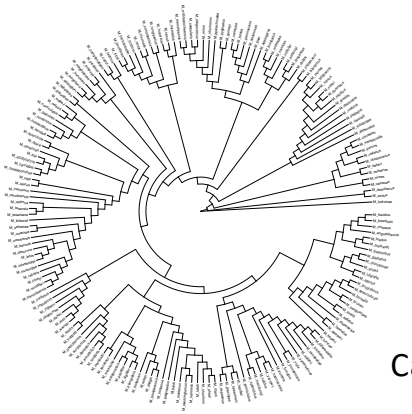
(modified from Harvey et al. 2019)



Do differences in species diversity across space, clades, and time, reflect differences in the opportunities for speciation

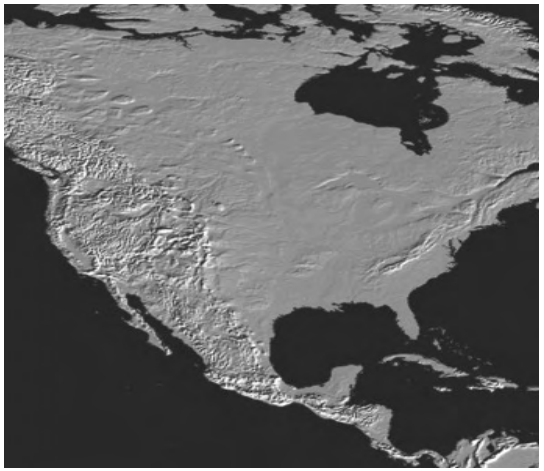
- higher rates of formation of isolated populations

- higher population persistence

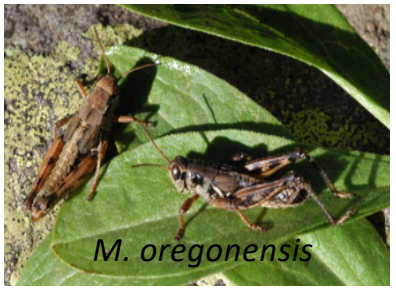


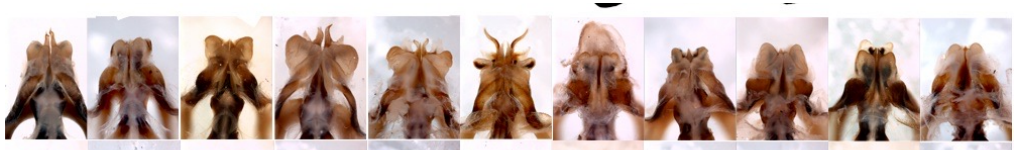
Cannot be tested from species lineage phylogenies

Hypotheses about how the opportunities for speciation may underly differences in diversity



- higher rates of formation of isolated populations
- higher population persistence





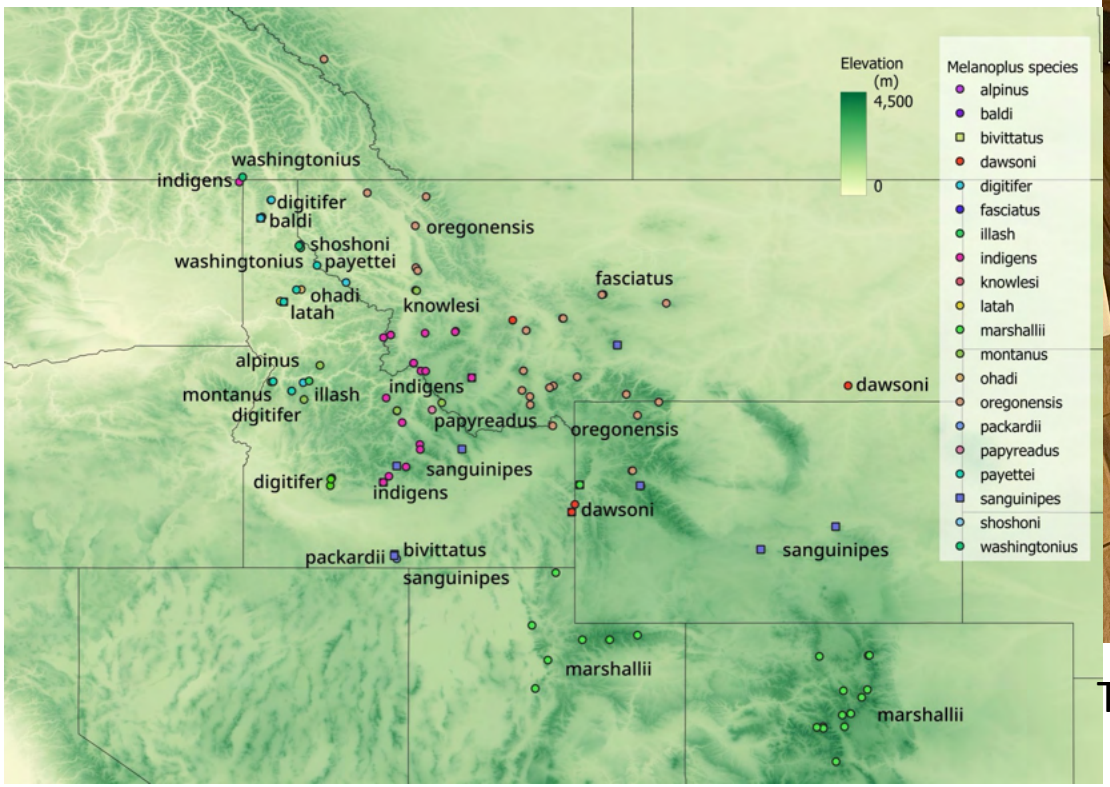
- higher rates of development of speciation isolating mechanisms

Secondary contact and recurrent gene flow are expected to hinder the completion of speciation, whereas small population sizes and habitat instability jeopardise the long-term persistence of newly formed lineages.

Hypotheses about species boundaries and diversification dynamics related to population-level processes

Big Data: Between and within species genetic structure; specifically, targeted capture of 15,000 loci developed from RADseq (RAPTURE)

- Have sequences for 15,000 individuals

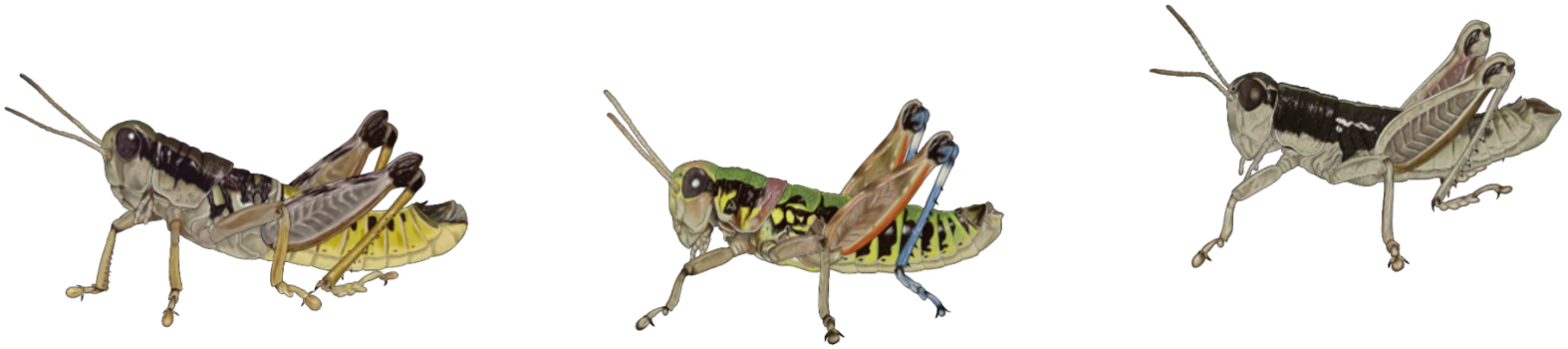


Tree for montane grasshoppers endemic to the northern Rocky Mountains

How latitudinal variation in species diversity reflects the balance between lineage formation and fusion—shaped by Pleistocene climatic oscillations—and determines the rates at which alpine microendemic species emerge and accumulate in temperate mountain regions

EXAMPLE: Melanoplinae: European Podismini

Melanoplinae is one of the most species-rich subfamilies of Acrididae (>1,250 species)

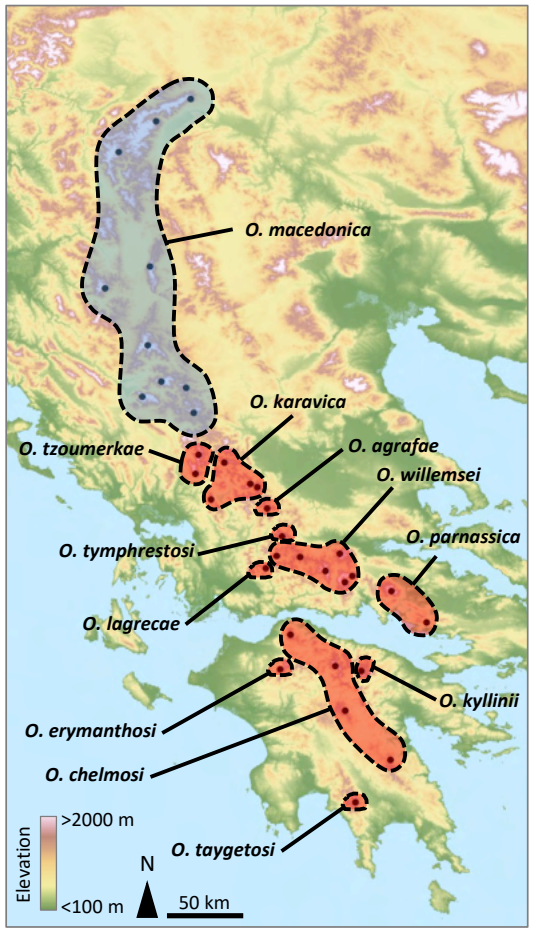


European Podismini

- Alpine-subalpine
- Most species are narrow-endemics
- A few species with large distributions

Joaquin Ortego et al. 2026 Latitudinal clines in gene flow and demographic stability reveal drivers of microendemism in a radiation of alpine grasshoppers. Mol. Ecol. 35, e70332

Speciation in alpine grasshoppers – Controls on Diversity



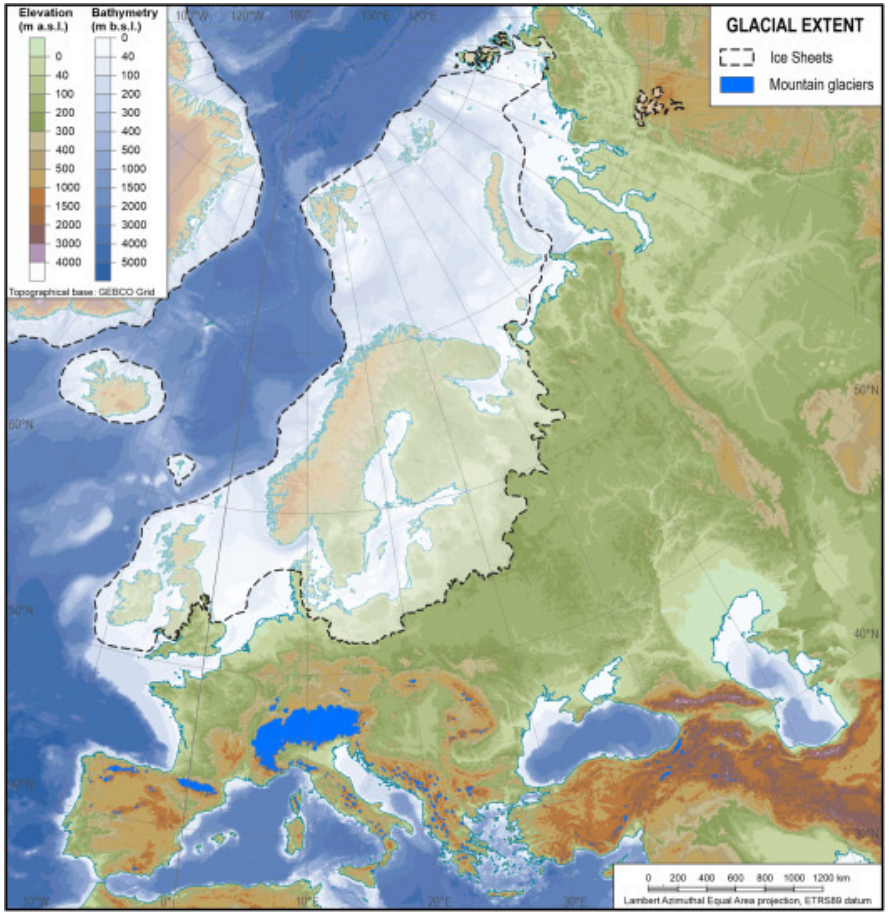
Higher latitudes

Lower latitudes

	Demographic	Microevolutionary	Macroecological
Higher latitudes	↑ Gene flow ↑ Demographic stability	**Inhibits** speciation	Few taxa, large distributions
Lower latitudes	↓ Gene flow ↓ Demographic stability	**Promotes** speciation	Many taxa, narrow distributions

Demographic and evolutionary dynamics: **Glaciations**

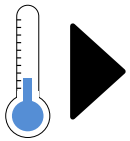
Contrasting impacts across latitudes



Hughes *et al*. 2022, *European Glacial Landscapes*

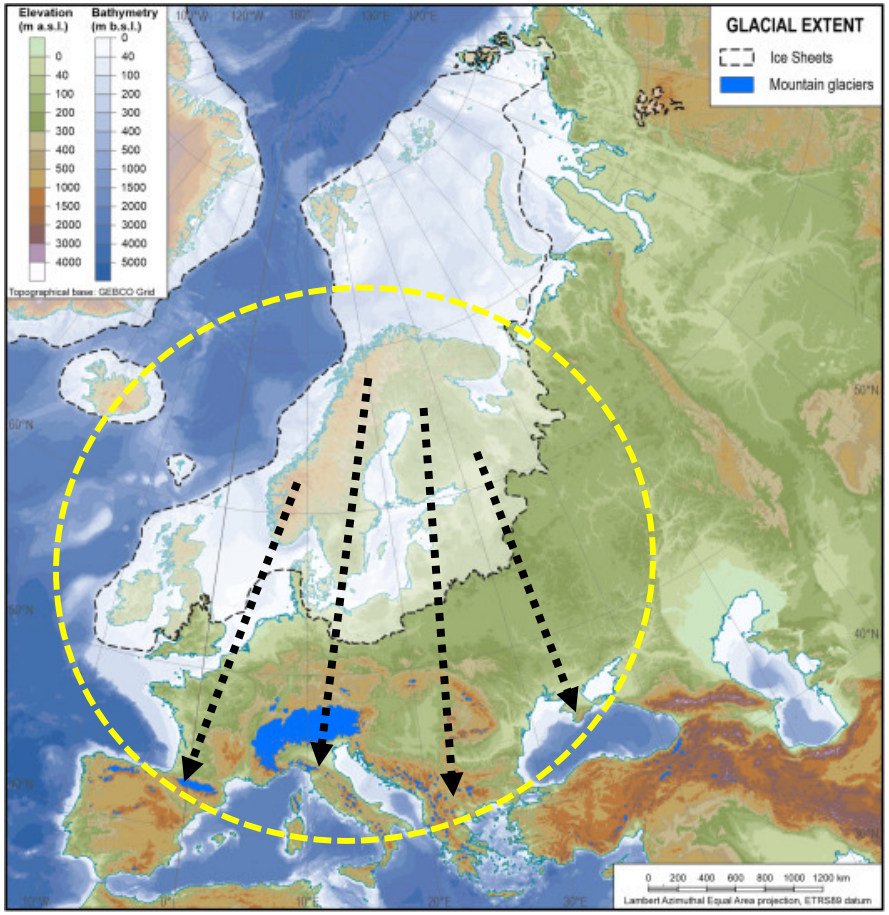
Demographic and evolutionary dynamics: **Glaciations**

Contrasting impacts across latitudes



Glacial periods

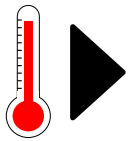
High latitudes



Hughes *et al*. 2022, *European Glacial Landscapes*

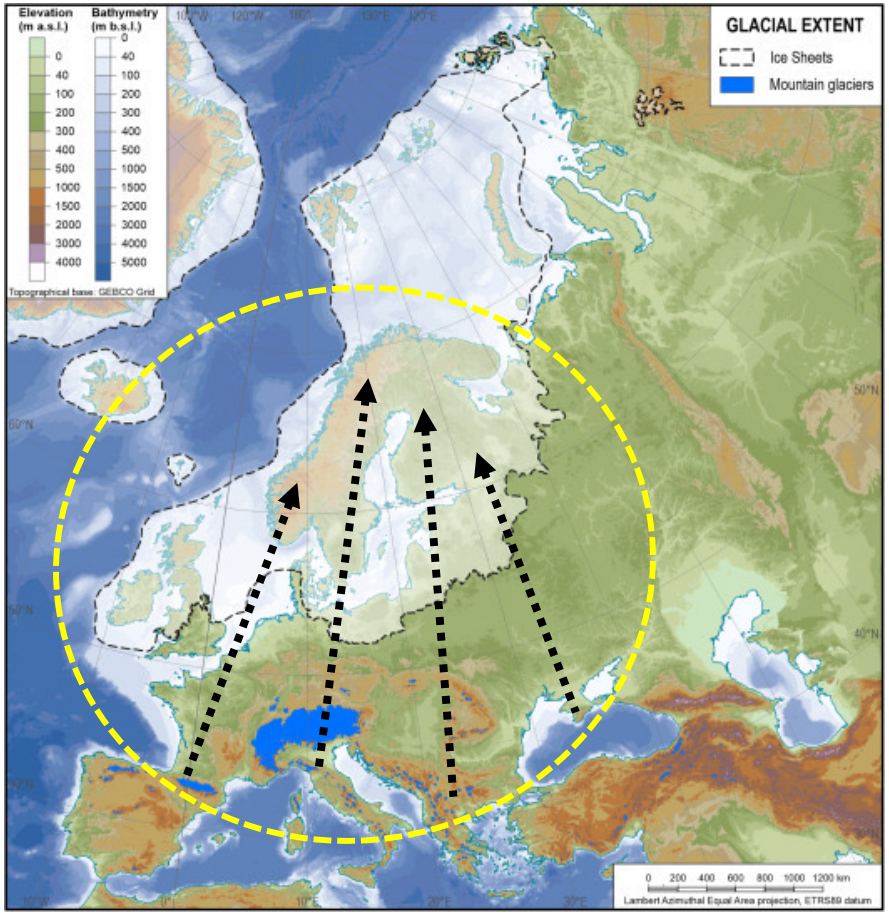
Demographic and evolutionary dynamics: **Glaciations**

Contrasting impacts across latitudes



Interglacial periods

High latitudes



Hughes *et al*. 2022, *European Glacial Landscapes*

Demographic and evolutionary dynamics: **Glaciations**

Contrasting impacts across latitudes

High latitudes

Genetic homogenization

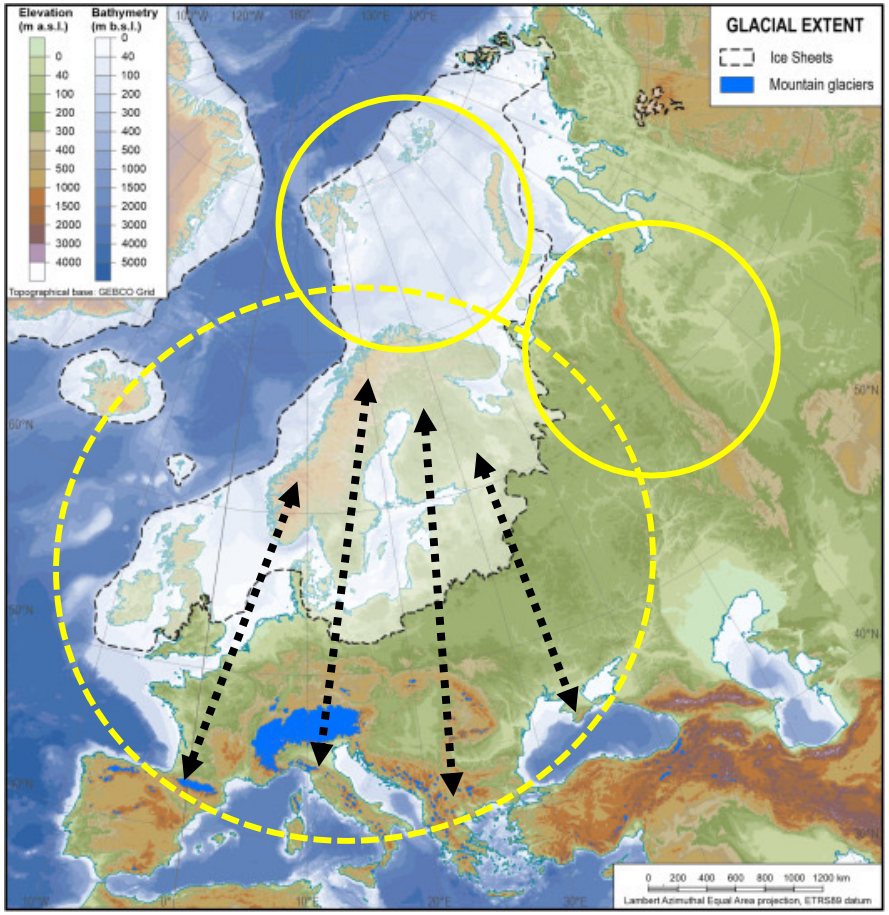
▼

Inhibit speciation

▼

Large distributional ranges

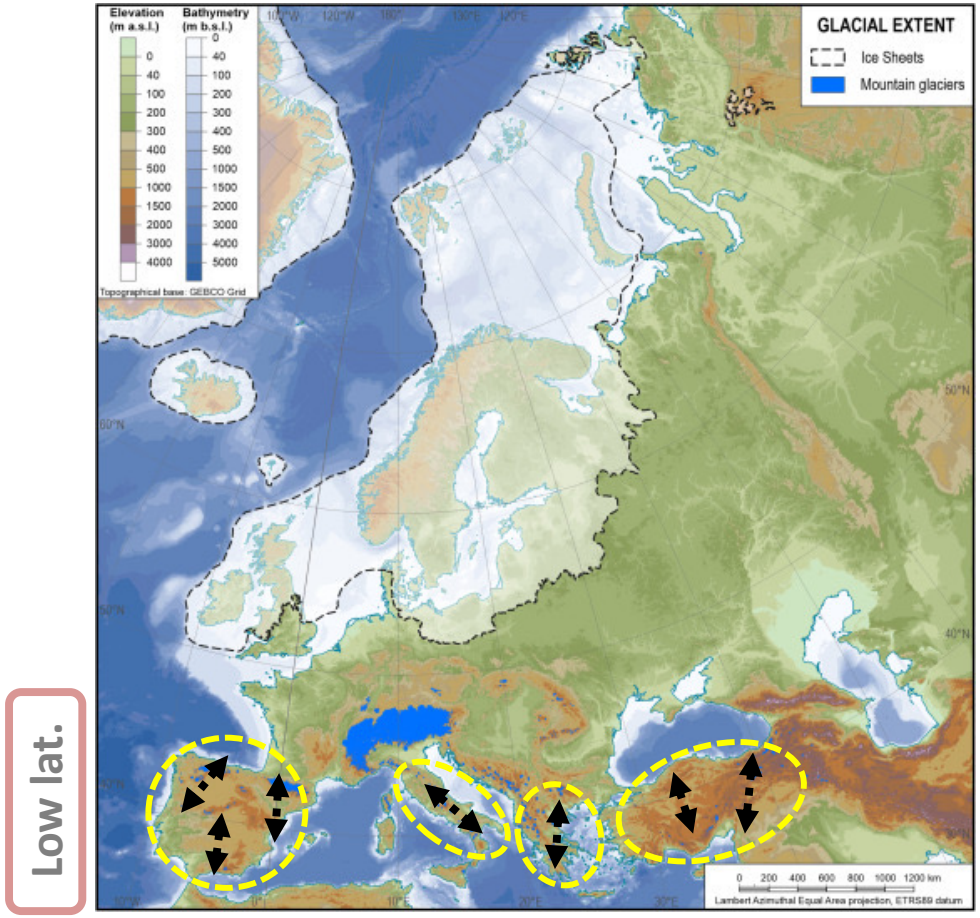
High latitudes



Jansson & Dynesius 2002, *Ann. Rev. Ecol. Syst.*

Demographic and evolutionary dynamics: **Glaciations**

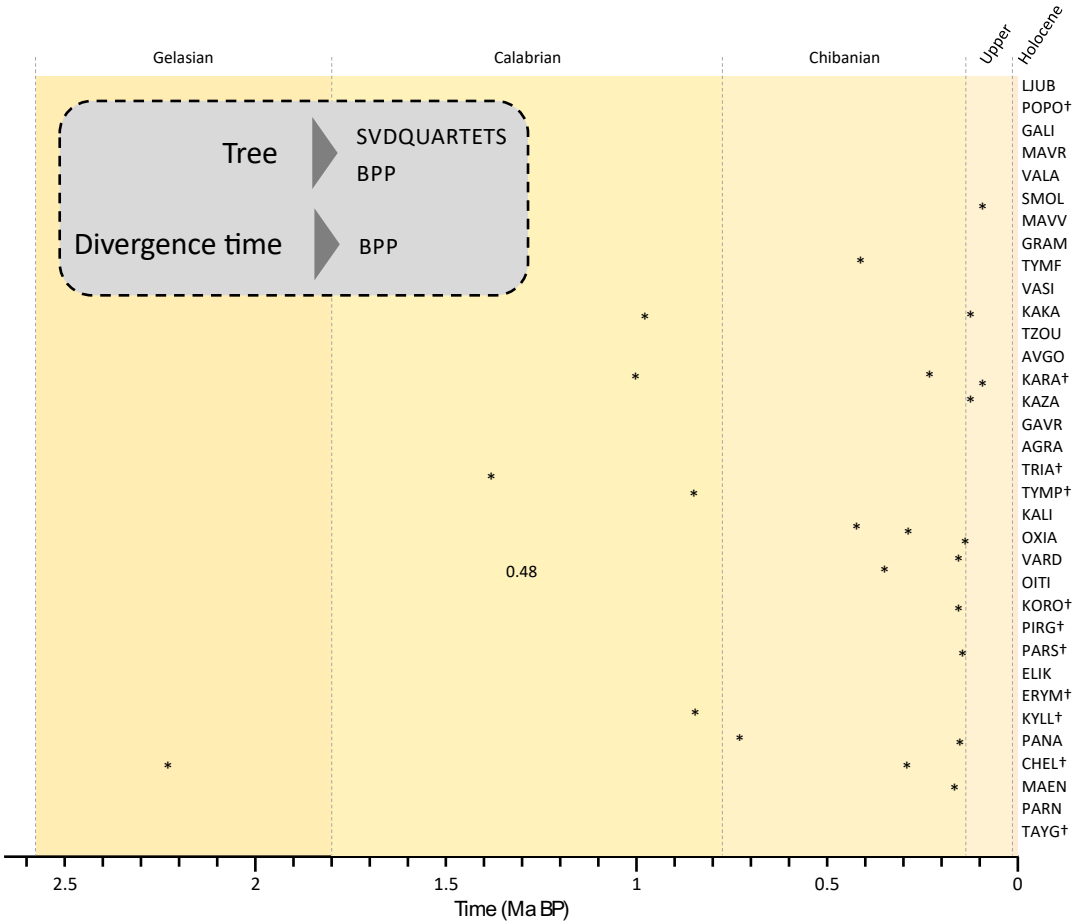
Contrasting impacts across latitudes



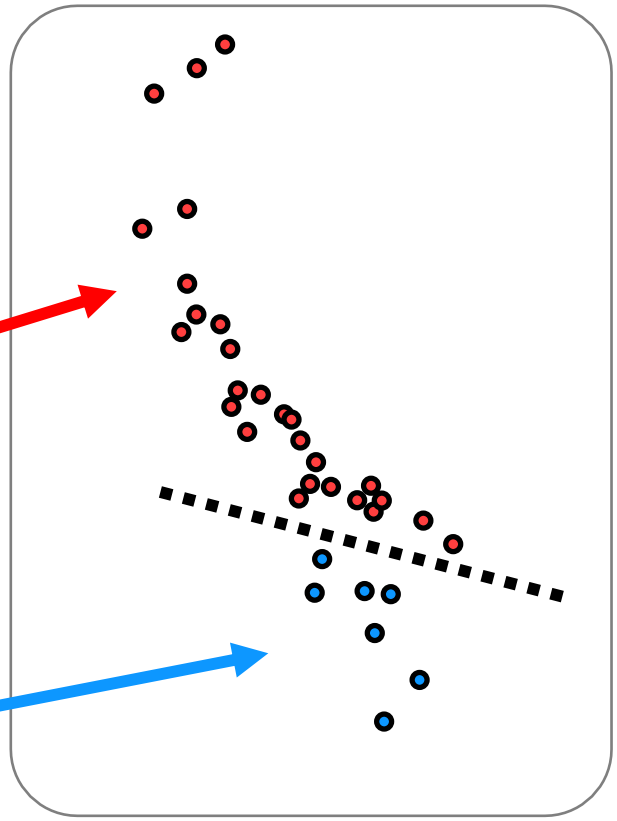
Low lat.

Jansson & Dynesius 2002, *Ann. Rev. Ecol. Syst.*

Speciation in alpine grasshoppers – Phylogeography



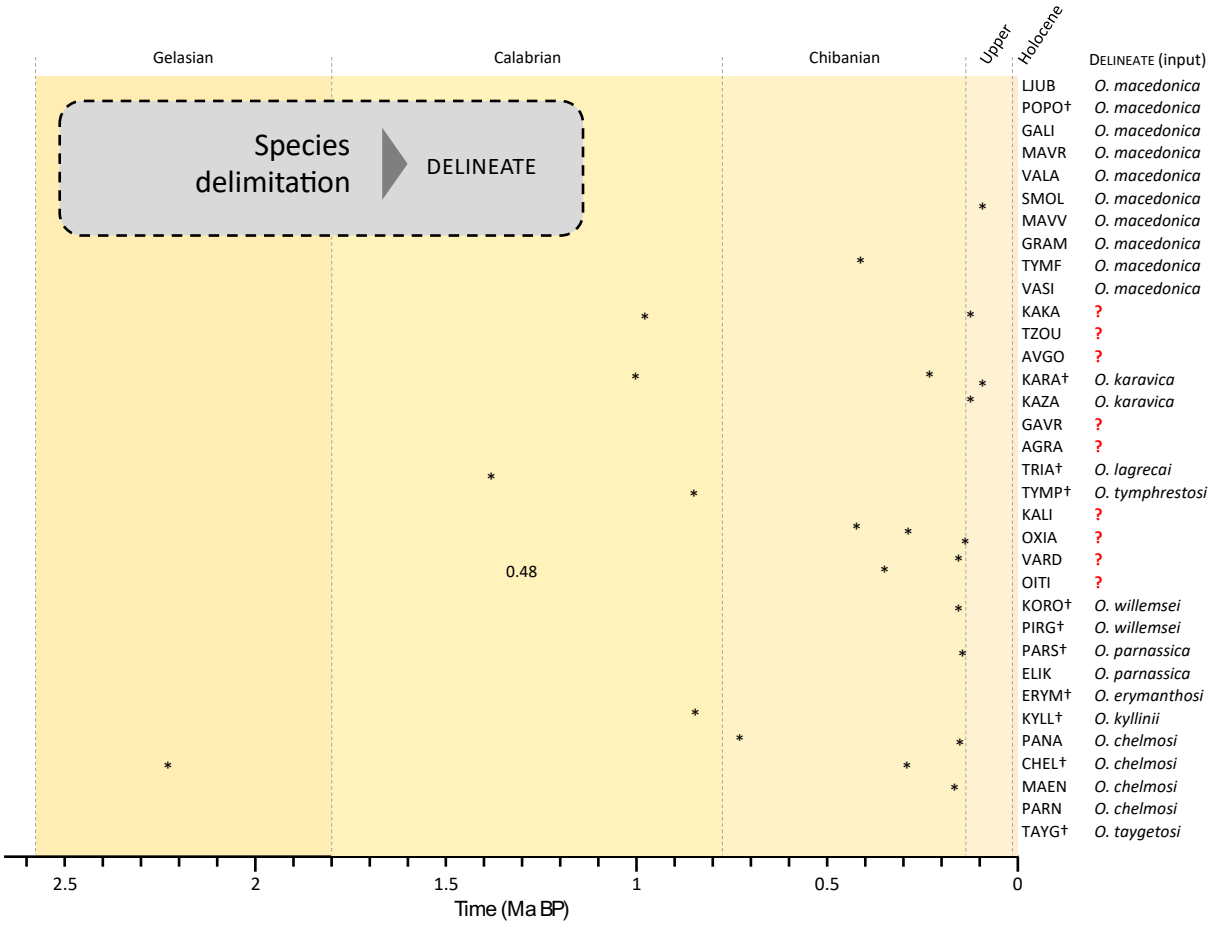
Biogeographic barrier: Corinthian Gulf

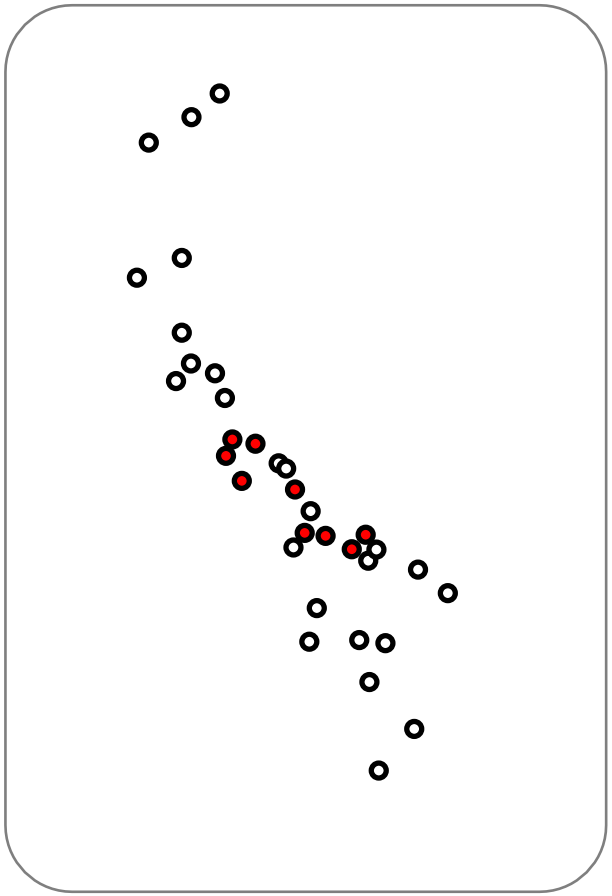


Kiourtsoglou *et al*. 2021, *J. Zool. Syst. Evol. Res.*

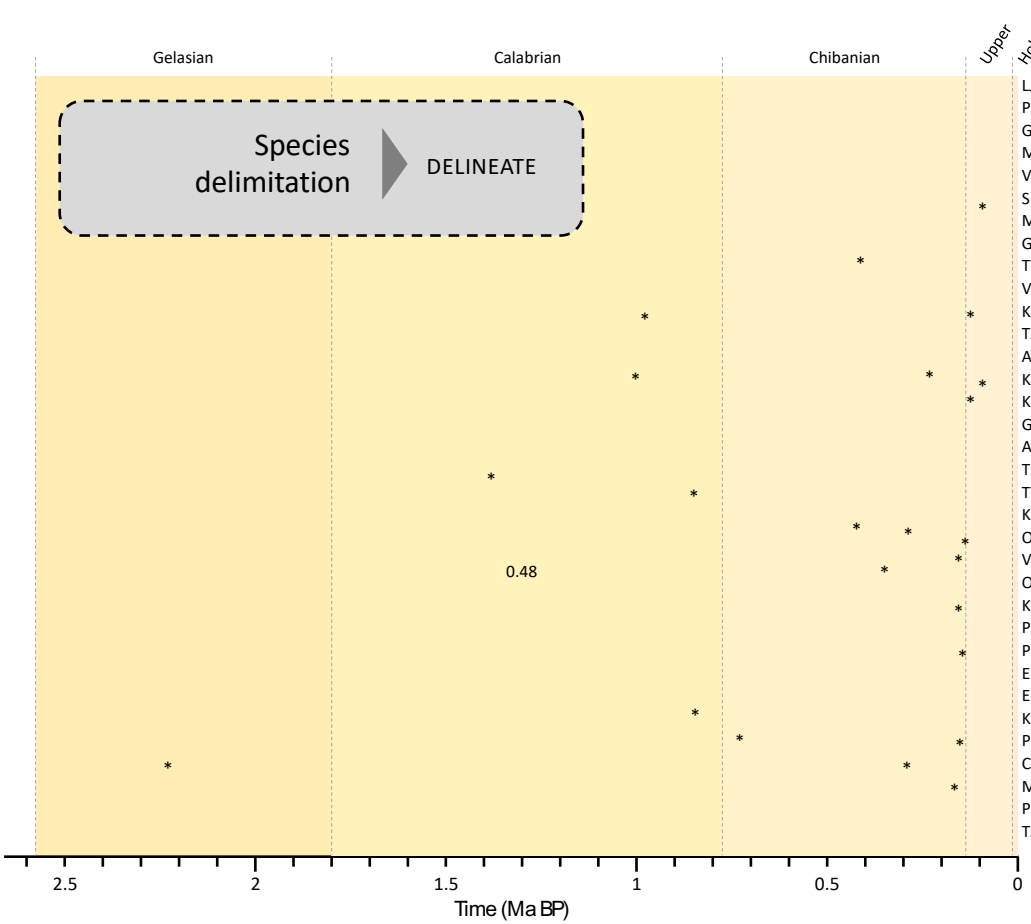
Taxonomically uncertain populations: new species or should they be assigned to existing described species?

Speciation in alpine grasshoppers – Species delimitation





Speciation in alpine grasshoppers – Species delimitation



	DELINEATE (input)		DELINEATE (output)
LJUB	*O. macedonica*	=	*O. macedonica*
POPO†	*O. macedonica*	=	*O. macedonica*
GALI	*O. macedonica*	=	*O. macedonica*
MAVR	*O. macedonica*	=	*O. macedonica*
VALA	*O. macedonica*	=	*O. macedonica*
SMOL	*O. macedonica*	=	*O. macedonica*
MAVV	*O. macedonica*	=	*O. macedonica*
GRAM	*O. macedonica*	=	*O. macedonica*
TYMF	*O. macedonica*	=	*O. macedonica*
VASI	*O. macedonica*	=	*O. macedonica*
KAKA	?		
TZOU	?		
AVGO	?	→	*O. karavica*
KARA†	*O. karavica*	=	*O. karavica*
KAZA	*O. karavica*	=	*O. karavica*
GAVR	?	→	*O. karavica*
AGRA	?		
TRIA†	*O. lagrecai*	=	*O. lagrecai*
TYMP†	*O. tymphrestosi*	=	*O. tymphrestosi*
KALI	?	→	*O. willemsei*
OXIA	?	→	*O. willemsei*
VARD	?	→	*O. willemsei*
OITI	?	→	*O. willemsei*
KORO†	*O. willemsei*	=	*O. willemsei*
PIRG†	*O. willemsei*	=	*O. willemsei*
PARS†	*O. parnassica*	=	*O. parnassica*
ELIK	*O. parnassica*	=	*O. parnassica*
ERYM†	*O. erymanthosi*	=	*O. erymanthosi*
KYLL†	*O. kyllinii*	=	*O. kyllinii*
PANA	*O. chelmosi*	=	*O. chelmosi*
CHEL†	*O. chelmosi*	=	*O. chelmosi*
MAEN	*O. chelmosi*	=	*O. chelmosi*
PARN	*O. chelmosi*	=	*O. chelmosi*
TAYG†	*O. taygetosi*	=	*O. taygetosi*

O. tzoumerkae sp. nov.

O. agrafae sp. nov.

Trillo & Ortego 2025, *ZooKeys*

Species delimitation analyses in delineate identified the recently described taxa *O. tzoumerkae* and *O. agrafae* as new species and assigned the rest of the taxonomically uncertain populations to the different delineated species.

Speciation in alpine grasshoppers – Species delimitation

Geometric morphometrics

GEOMORPH, STEREOMORPH

"Diagnostic traits"

♂

Phallus apex

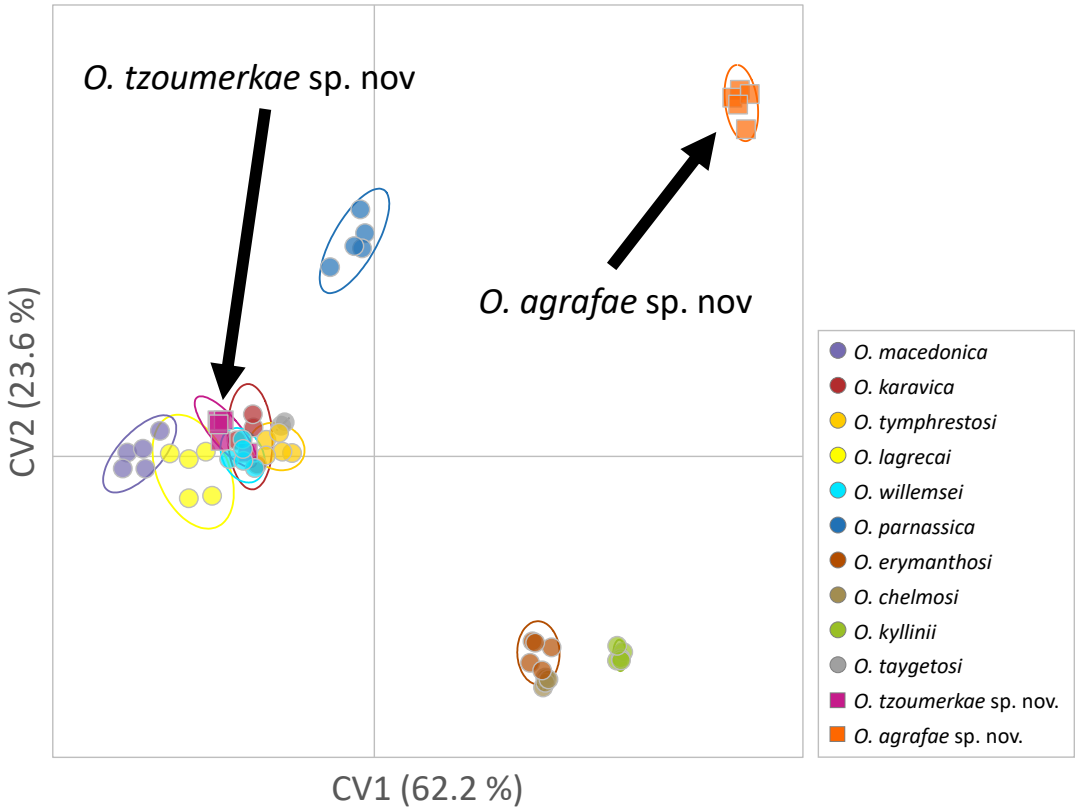
Furculae

Speciation in alpine grasshoppers – Species delimitation

Geometric morphometrics

GEOMORPH, STEREOMORPH

Phallus apex



Trillo & Ortego 2025, *ZooKeys*

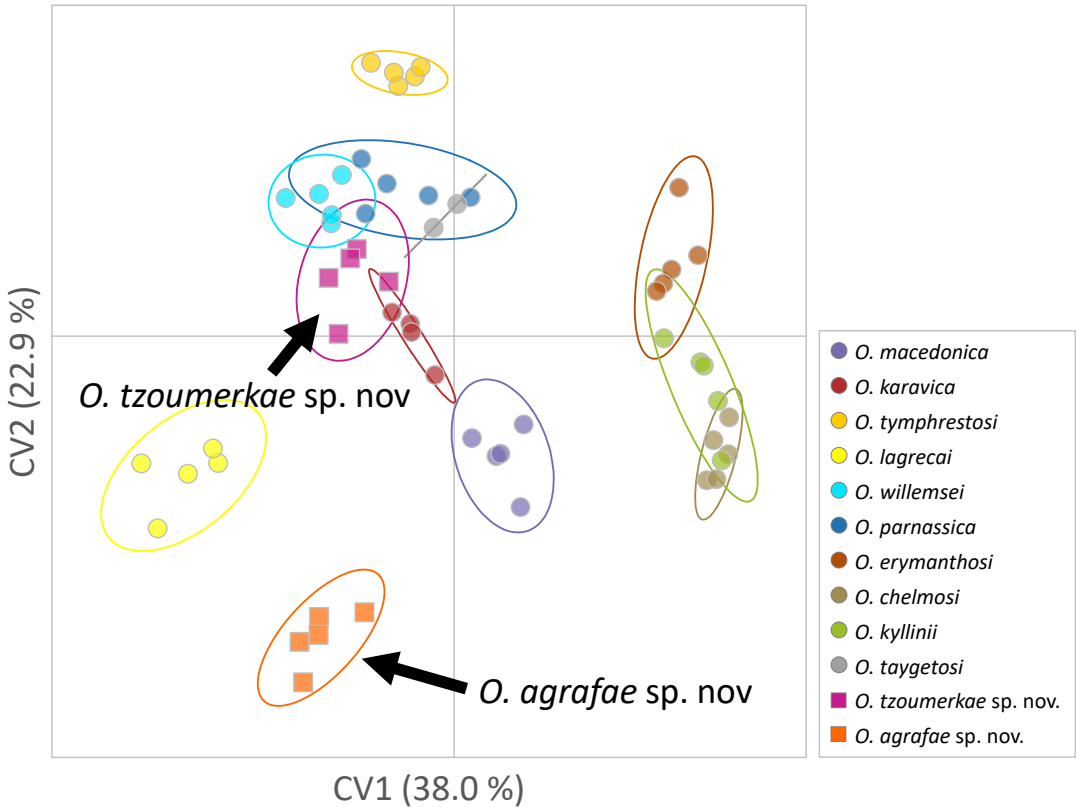
Speciation in alpine grasshoppers – Species delimitation

Geometric morphometrics

GEOMORPH, STEREOMORPH

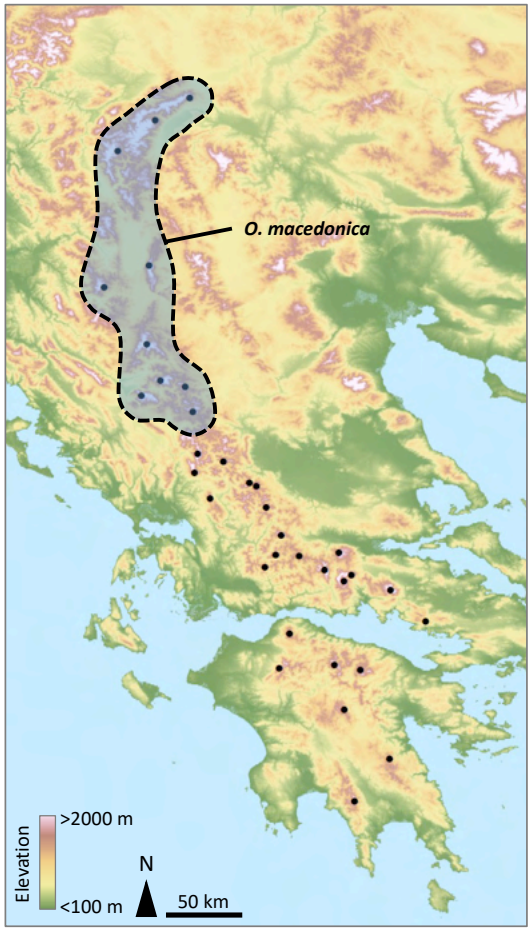
Furculae

O. tzoumerkae sp. nov: **Cryptic species**



Trillo & Ortego 2025, *ZooKeys*

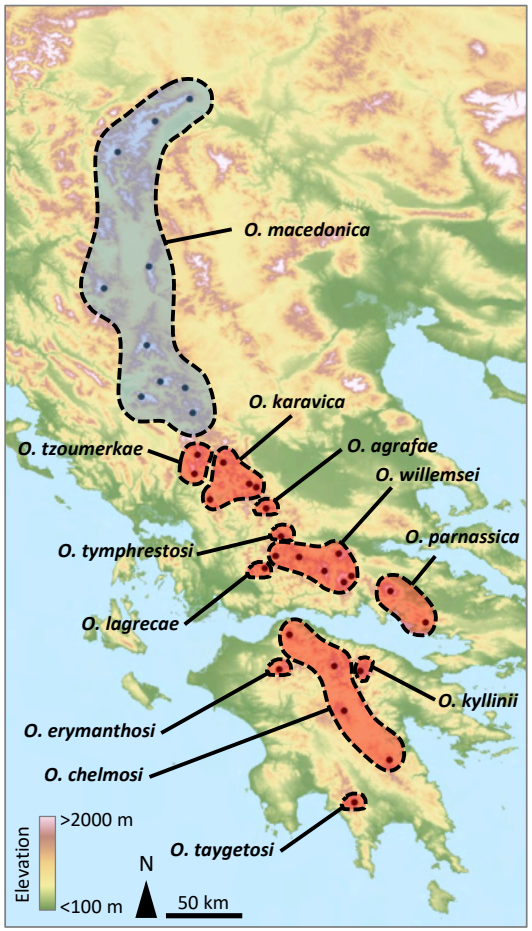
Speciation in alpine grasshoppers – Range size variation



Higher latitudes

1 species, large distribution

Speciation in alpine grasshoppers – Range size variation



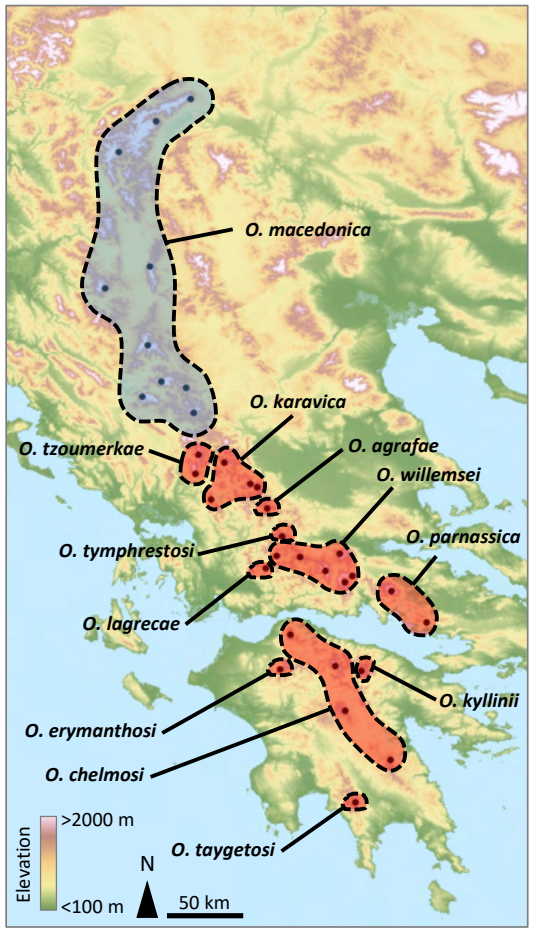
Higher latitudes

1 species, large distribution

Lower latitudes

11 species, narrow distributions

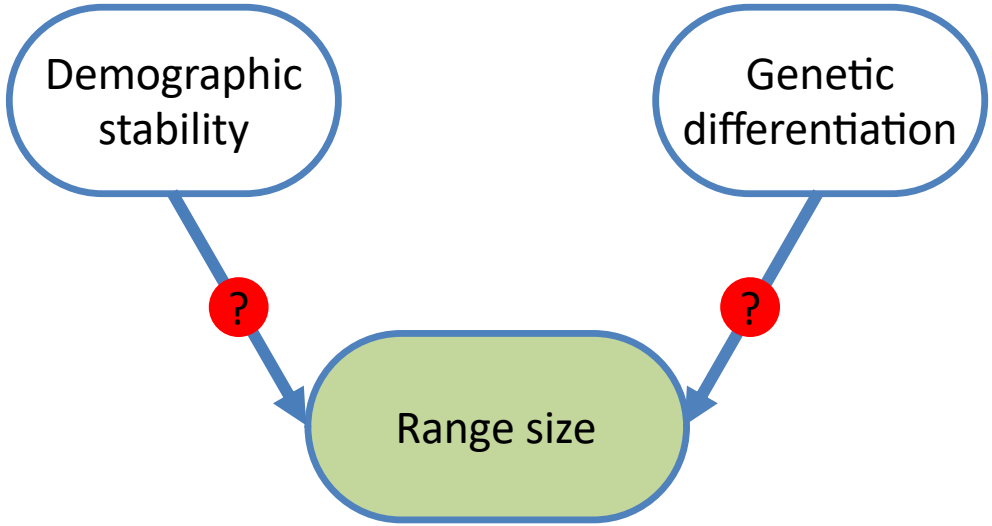
Speciation in alpine grasshoppers – Range size variation



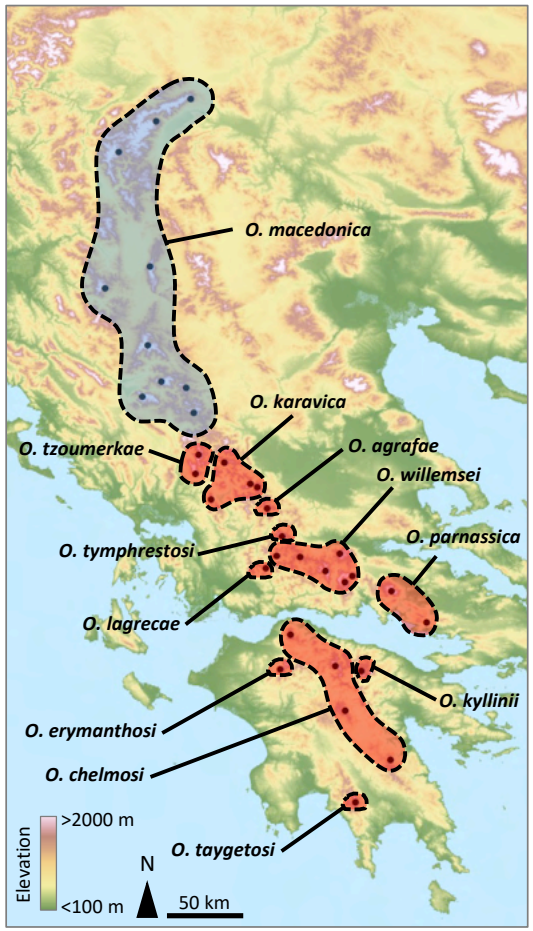
Higher latitudes

Lower latitudes

Is range size variation explained by microevolutionary and demographic processes?



Speciation in alpine grasshoppers – Controls of range size variation



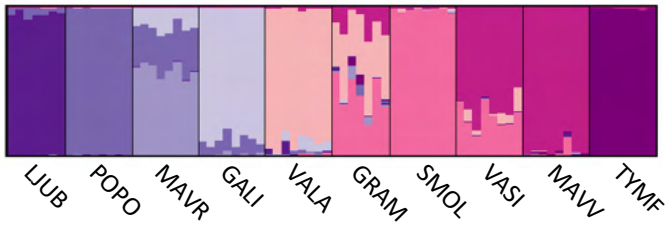
Higher latitudes

Lower latitudes

Genetic structure (STRUCTURE)

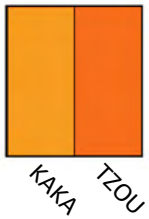
Genetic differentiation (F_{ST})

O. macedonica

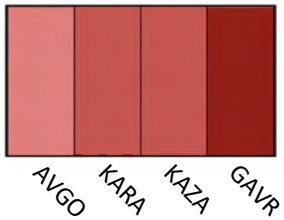


F_{ST} = **0.20**

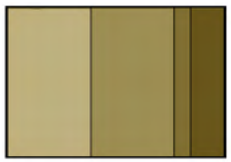
O. tzoumerkae



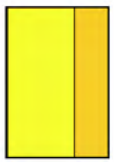
O. karavica



O. chelmosi

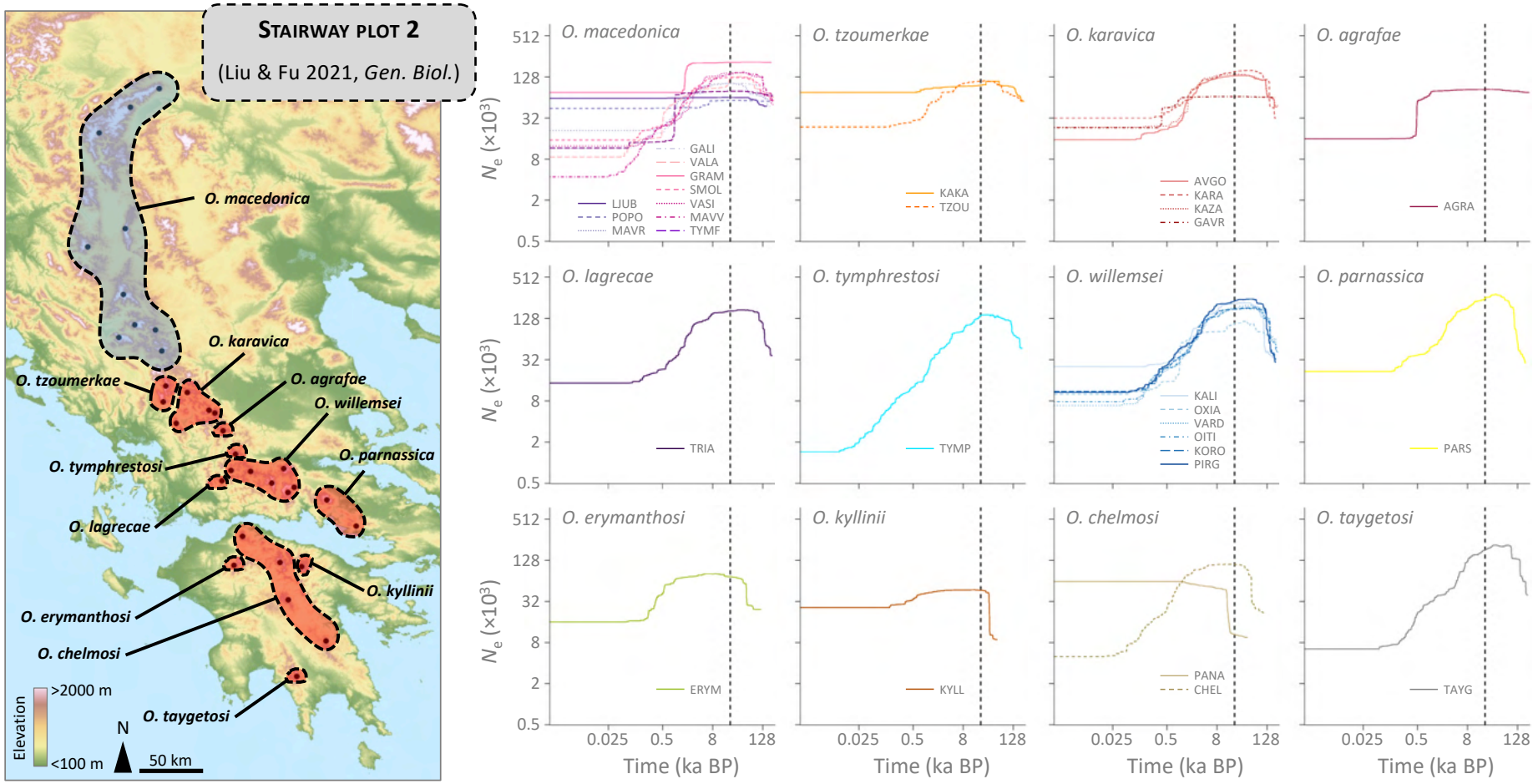


O. parnassica



F_{ST} = **0.43**

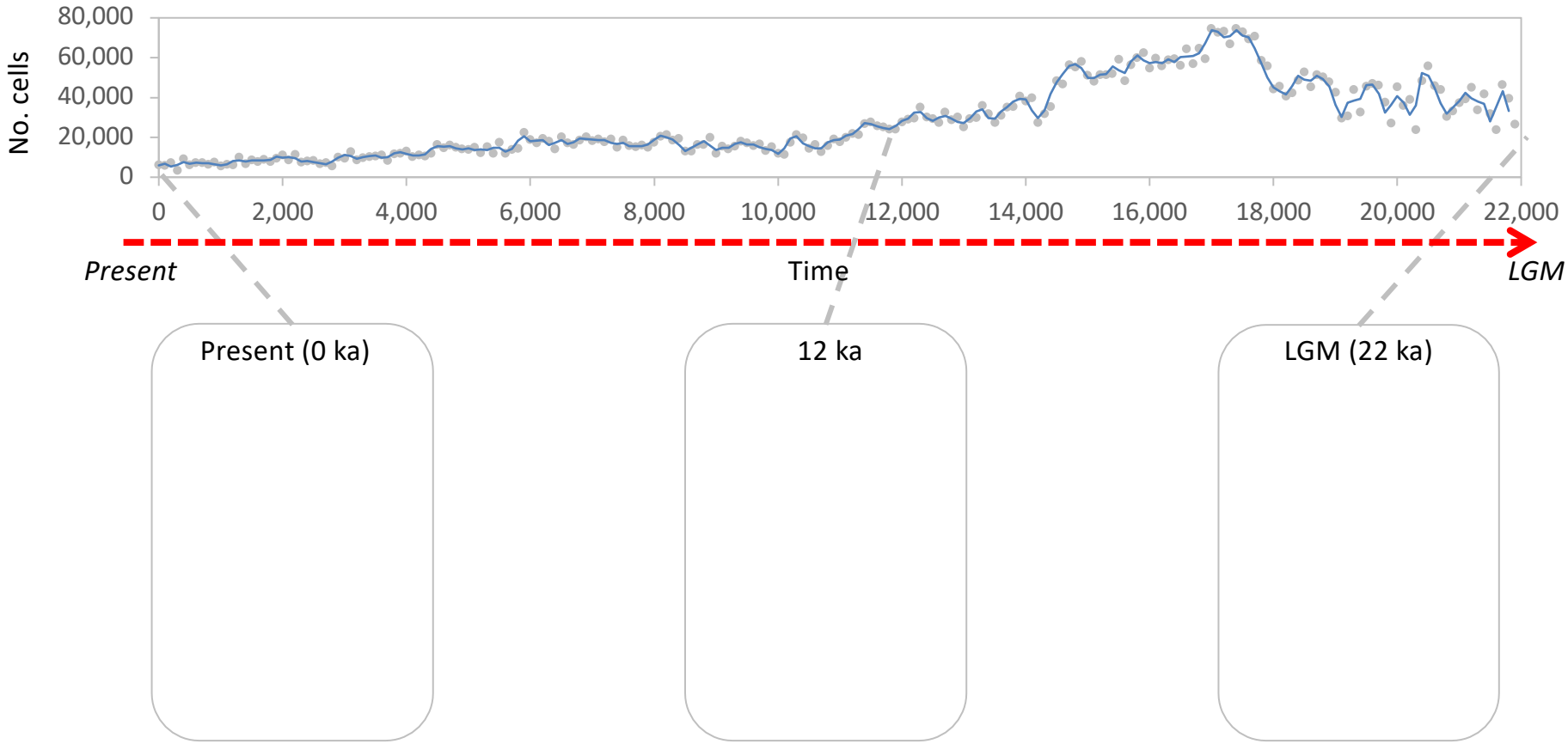
Speciation in alpine grasshoppers – Controls of range size variation



Demographic history of the studied populations for each taxon within the genus *Oropodisma* inferred using stairway plot. Only populations with $n \geq 7$ genotyped individuals were analysed. Panels show the median of effective population size (*N*e) through time, estimated assuming a mutation rate of 2.8 × 10−9 and one generation per year

Extent of climatically suitable habitats for the genus *Oropodisma* as inferred from projections of the environmental niche model (ENM) to bioclimatic conditions during the last 22,000 years

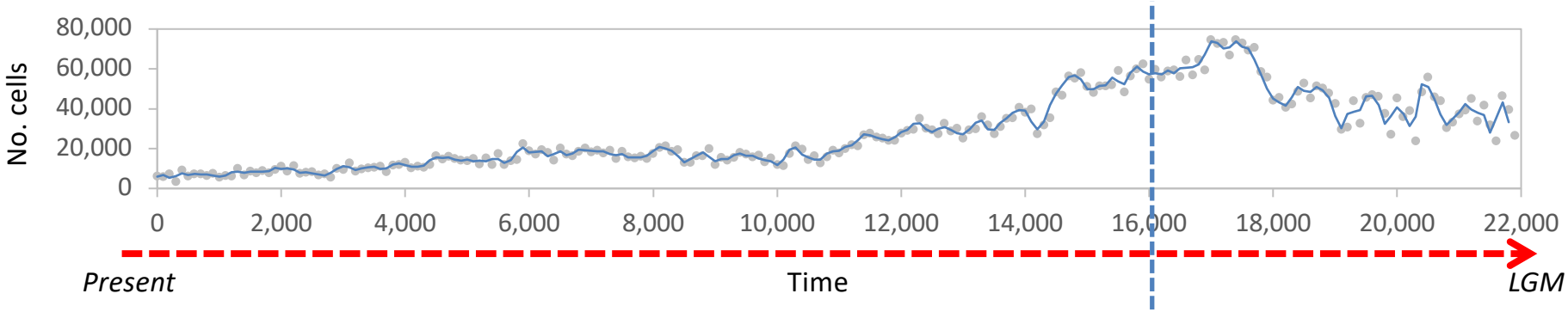
Speciation in alpine grasshoppers – Controls of range size variation

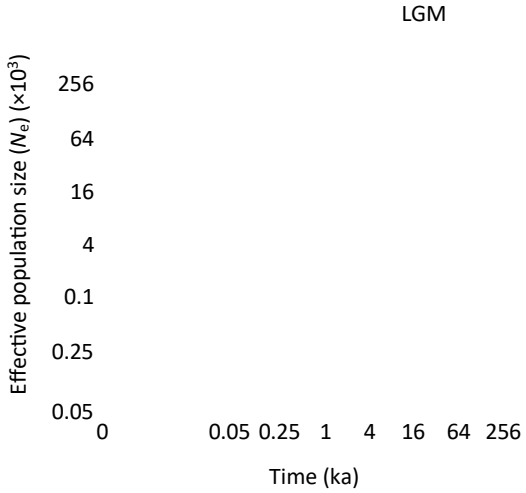


ENMEVAL: Muscarella *et al*. 2014, *Methods Ecol. Evol.*; **CHELSA-TraCE21k**: Karger *et al*. 2023, *Clim. Past*

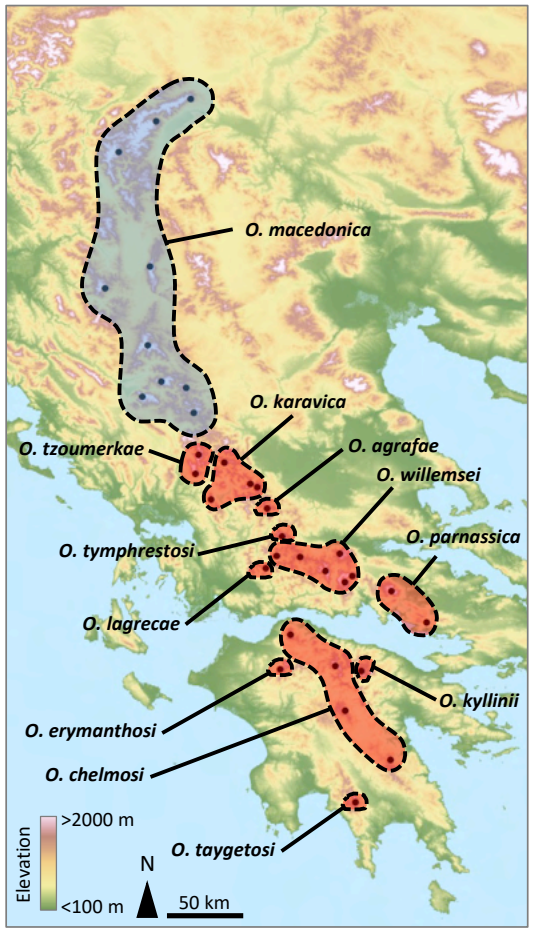
Correspondence between extent of climatically suitable habitats for the genus *Oropodisma* as inferred from projections of the ENMs and effective population size

Speciation in alpine grasshoppers – Controls of range size variation



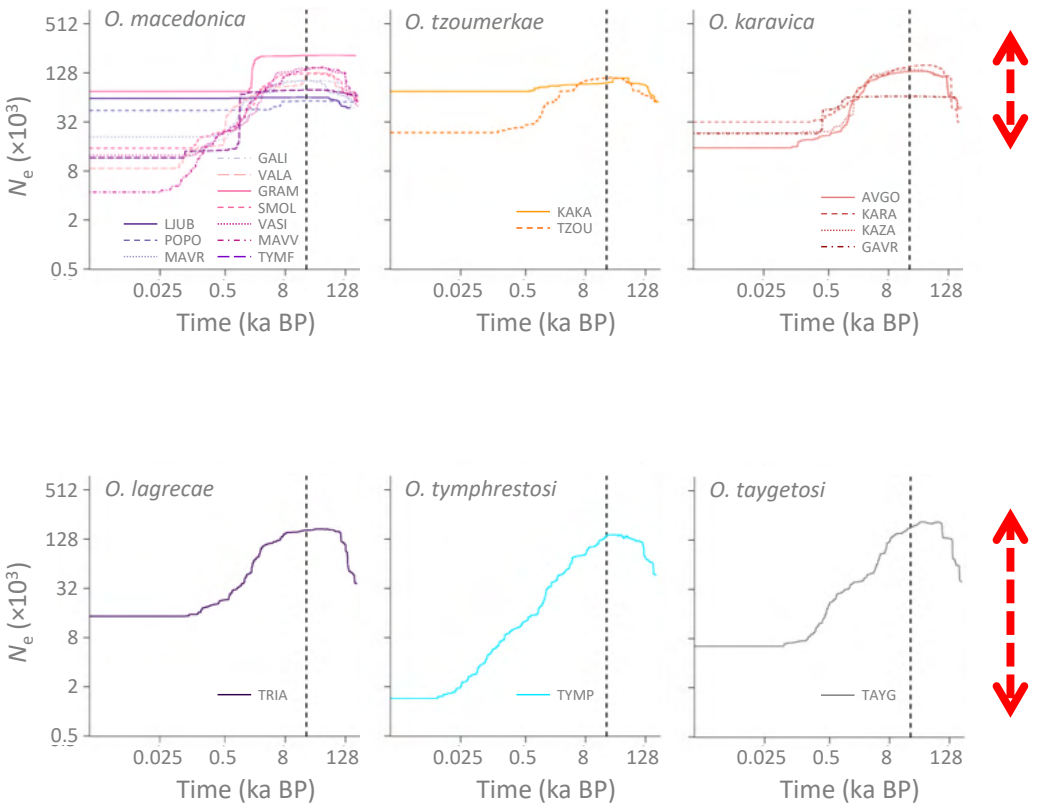


Speciation in alpine grasshoppers – Controls of range size variation



Higher latitudes

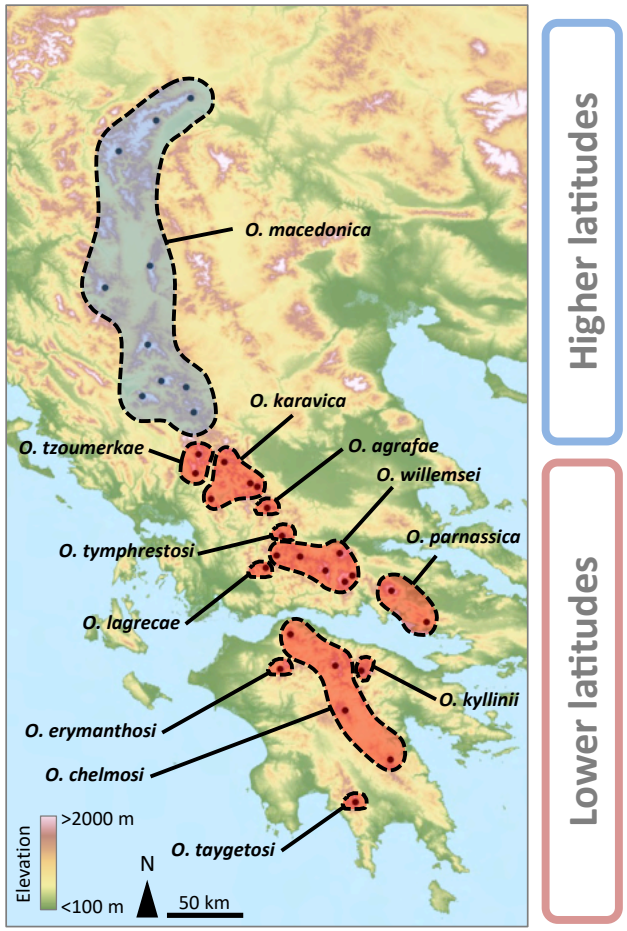
Lower latitudes



stairway plot analyses show that most populations of *Oropodisma* have experienced parallel demographic trajectories, undergoing severe declines of *N*e generally starting at the onset of the Holocene preceded, in most cases, by demographic expansions during the last glacial period

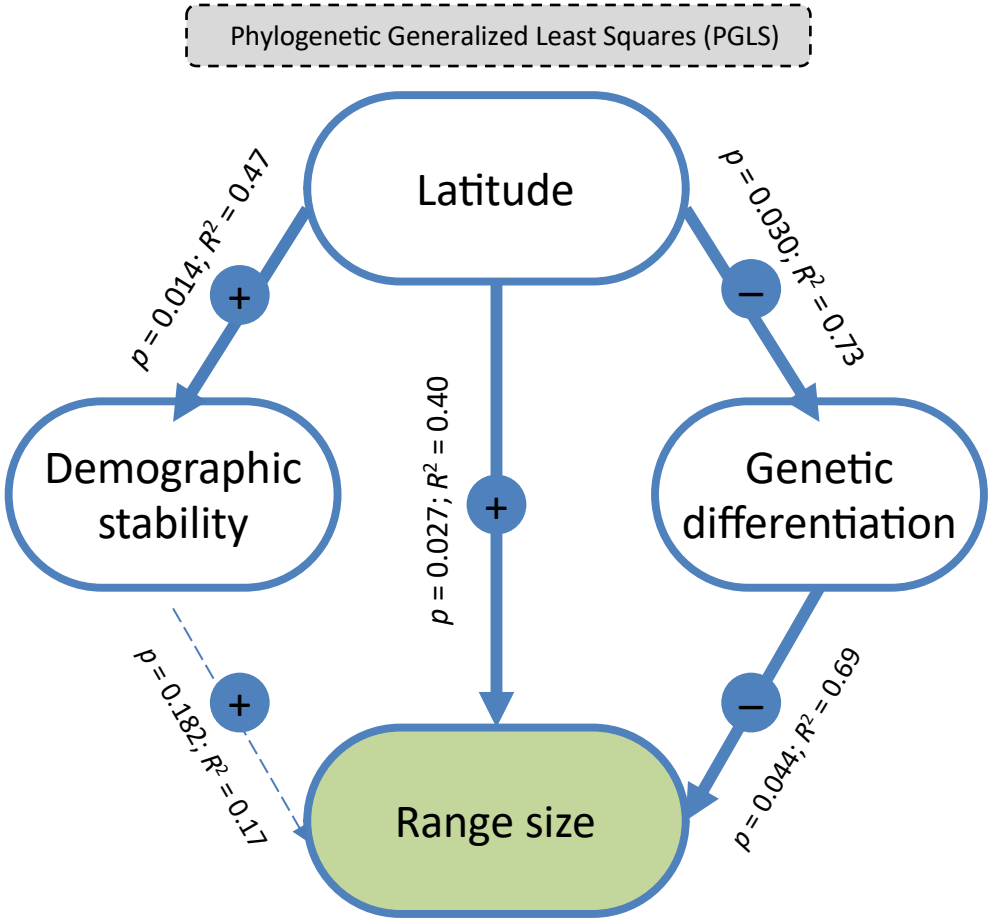
High rates of lineage formation triggered by geographical isolation (i.e., frequent splitting), the presence of climate refugia that prevent extinction (i.e., local persistence), and the rapid evolution of reproductive isolation that impedes incipiently diverging lineages from merging (i.e., short speciation times) are likely key processes driving the extraordinary levels of local microendemism that characterise alpine and montane biotas at mid to low latitudes

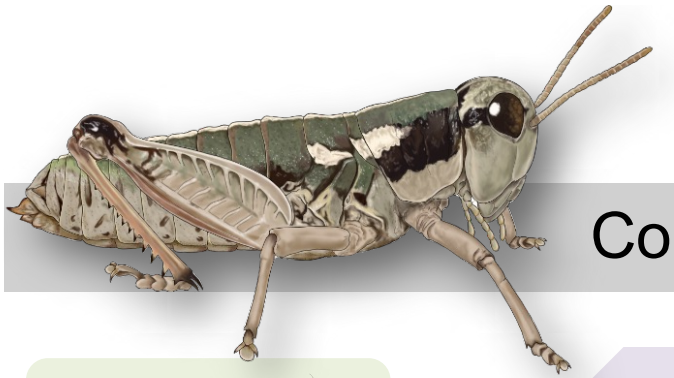
Speciation in alpine grasshoppers – Controls of range size variation



Higher latitudes

Lower latitudes





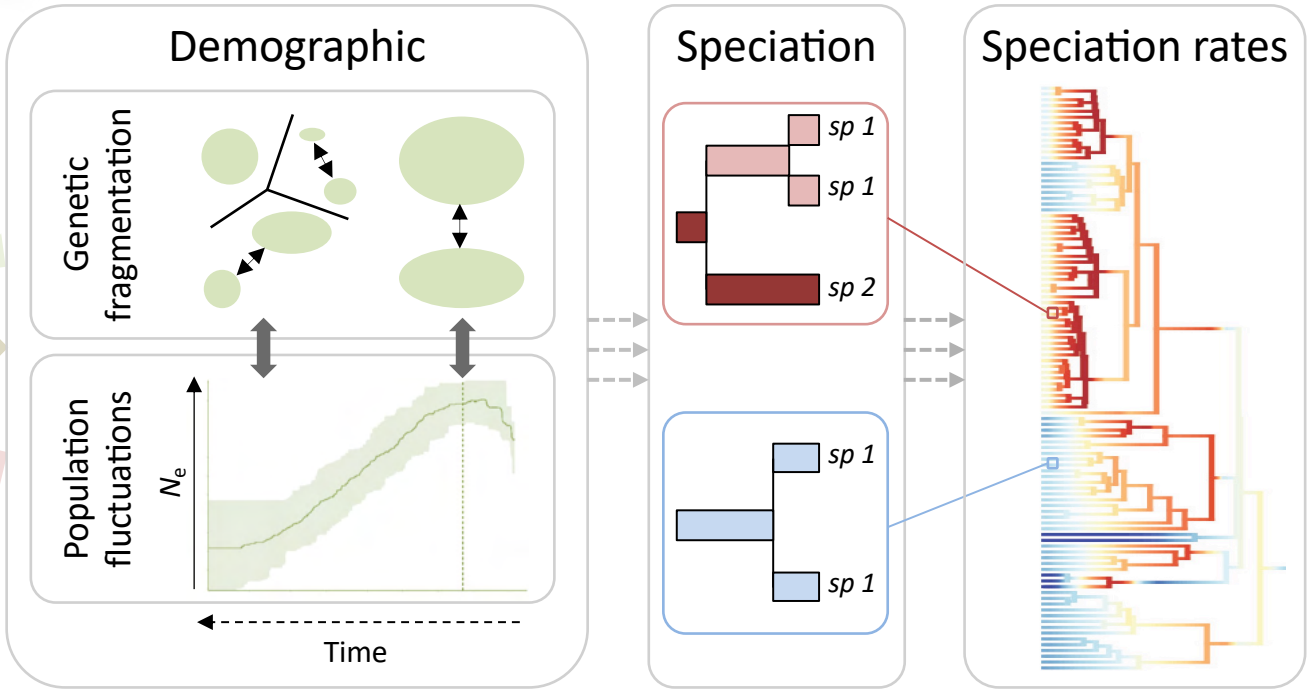
Controls on diversification

Biotic
- Dispersal capacity
- Niche breadth
- Interspecific interactions

- Patch size
- Range size
- Range shifts

Abiotic
- Topography
- Environment
- Climate change

Demographic | Evolutionary (micro-macro)



This suggests that greater genetic connectivity at higher relative latitudes, likely driven by regional range shifts and a more continuous availability of suitable habitats through time, has limited opportunities for microgeographic speciation and maintained genetic cohesiveness among populations across broader distributional ranges.

Great team of researchers!

Wonwoong Kim

Megan Sporre

Min-Xin Luo

UNIVERSITY OF MICHIGAN

Thank you!

Mark Holder
Univ. of Kansas

former Postdoctoral fellows:

Jeet Sukumaran
San Diego State Univ.

https://github.com/jeetsukumaran/delineate

Former UM
Ph. D. students:

Giorgia G. Auteri

Qixin He

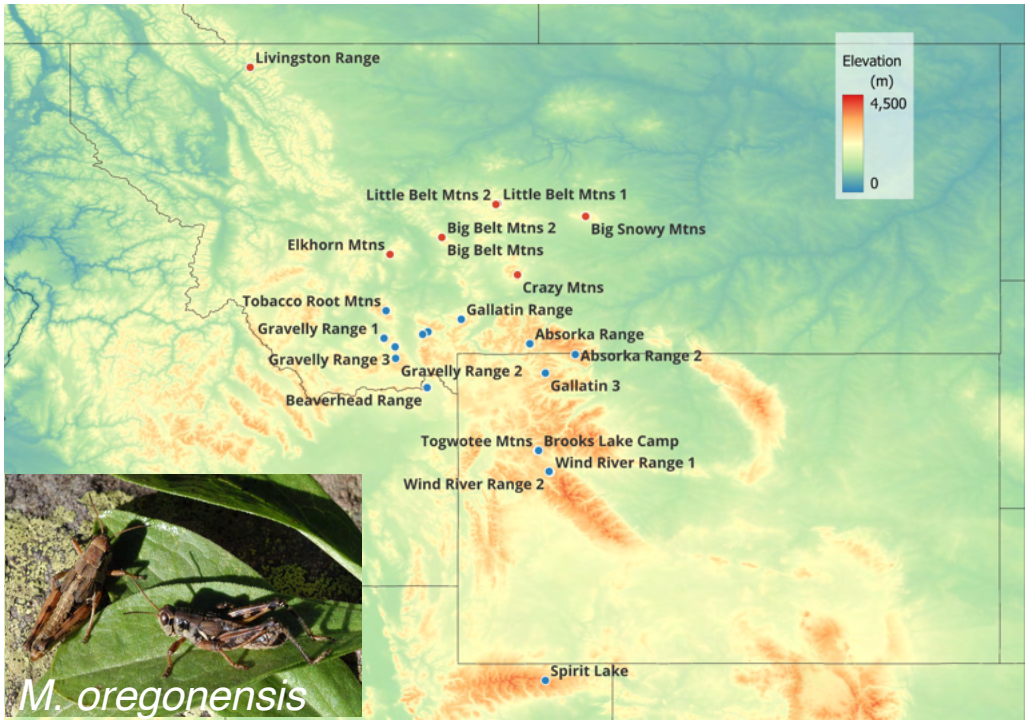
Rob Massatti

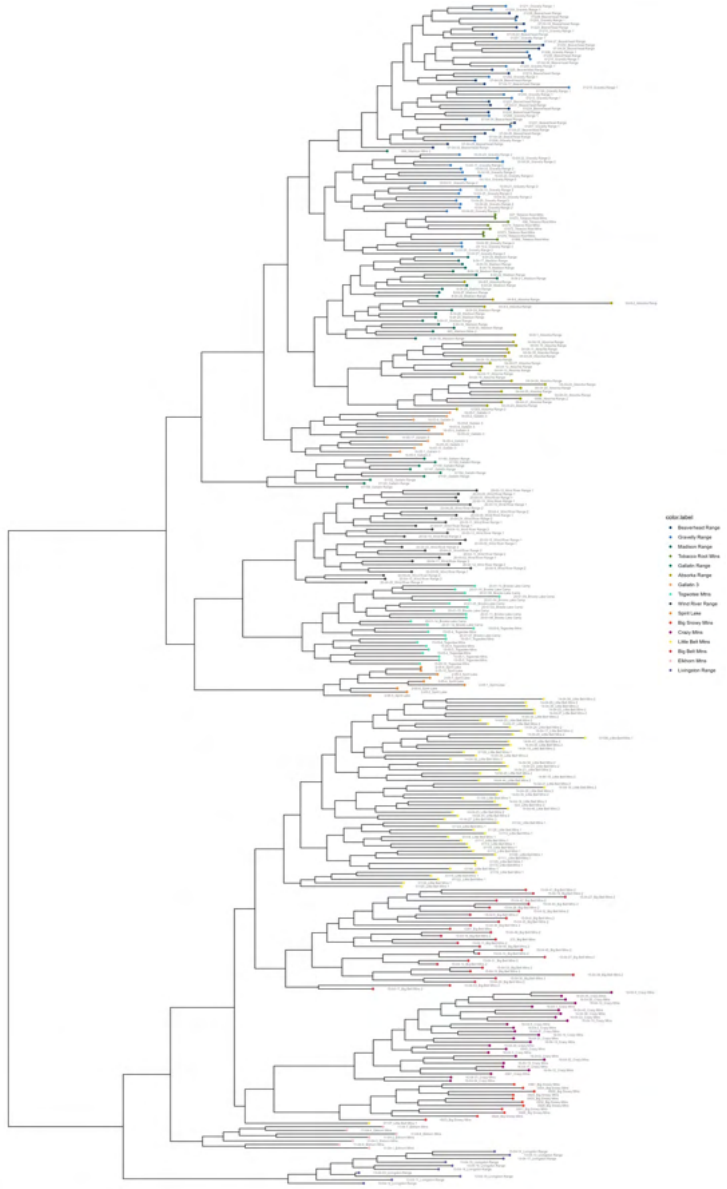
Anna Papadopoulou
University of Cyprus

support NSF & the UM

knowlesl@umich.edu

Test hypotheses about species boundaries and the effects of "opportunities for speciation" on species diversity patterns





- How does being restricted to isolated montane habitats across "sky islands" affect the opportunity for speciation and how do different properties of the sky islands and the history of climate change mediate the speciation process