

Islands and Integrals

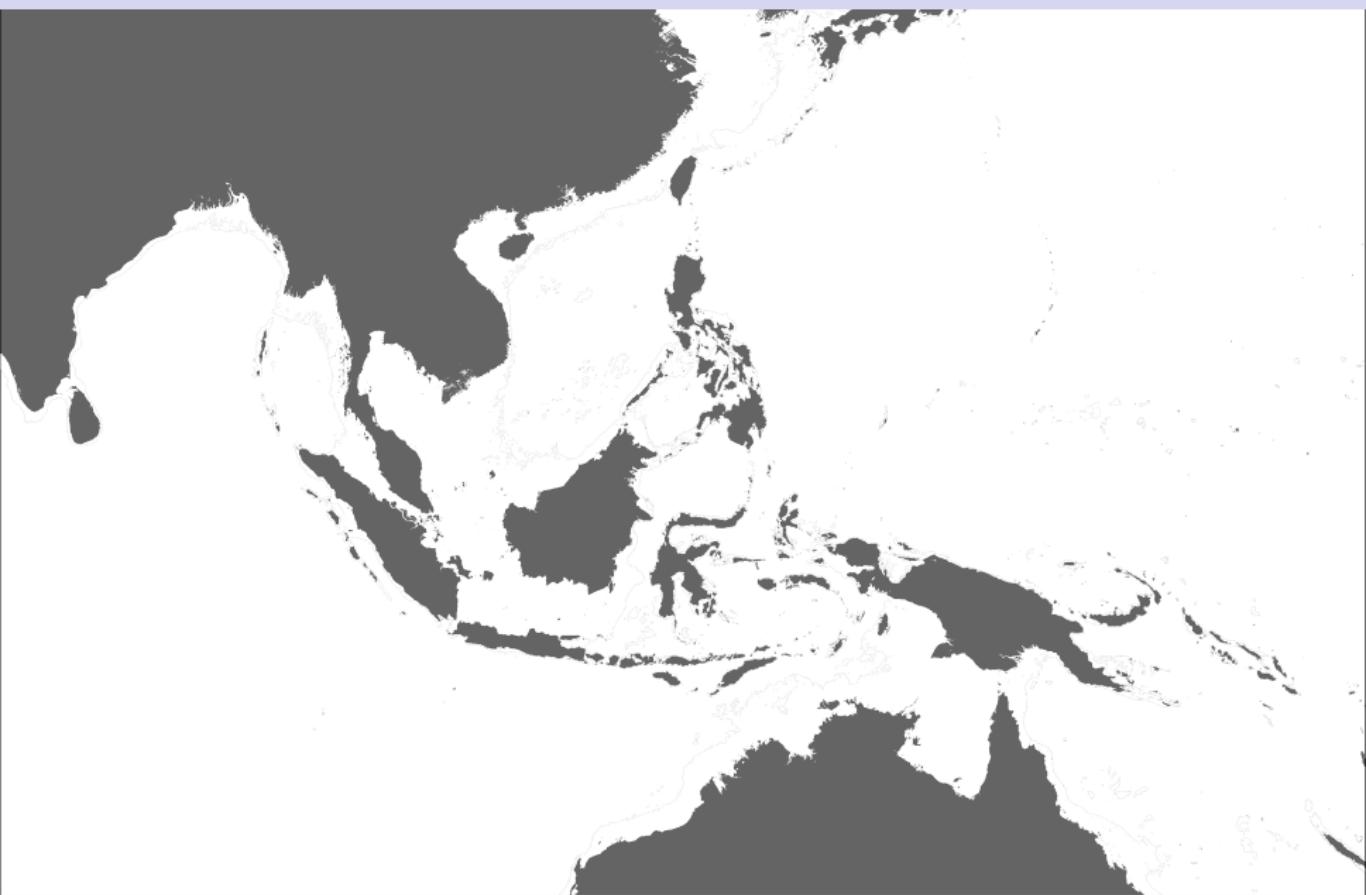
Processes of Diversification in an Island Archipelago and
Bayesian Methods of Comparative Phylogeographical Model
Choice

Jamie R. Oaks¹

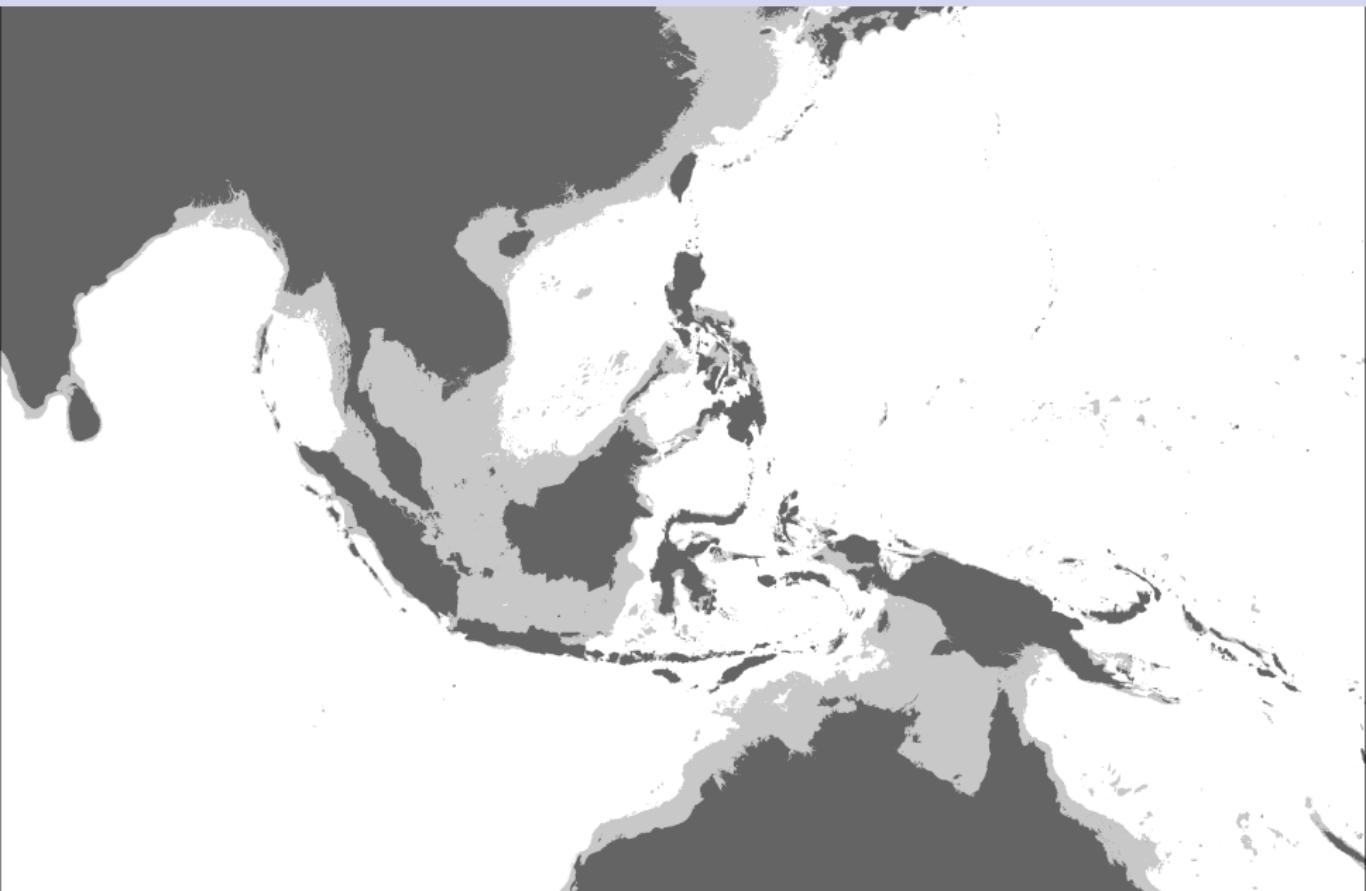
¹Department of Ecology and Evolutionary Biology, University of Kansas

October 16, 2013

Southeast Asia



Southeast Asia



Philippine Archipelago



Philippine Archipelago



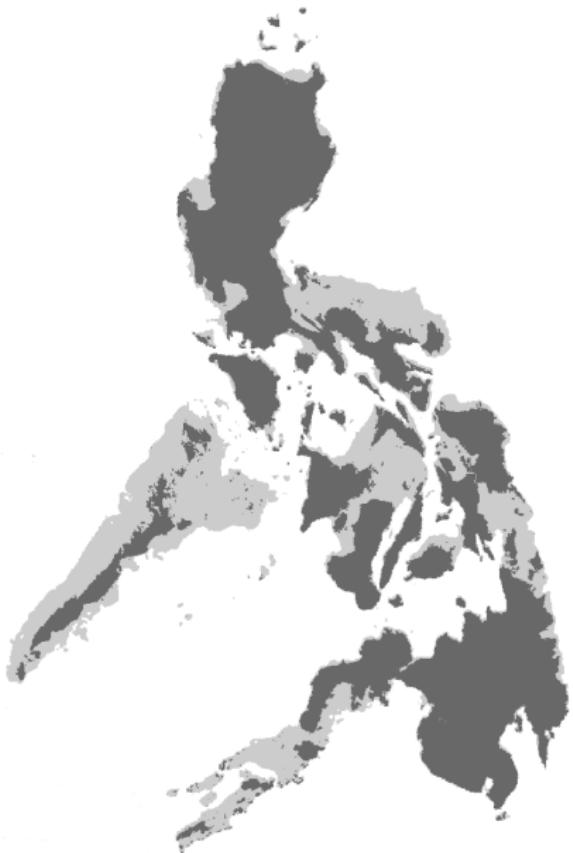
Climate-driven diversification model

- ▶ Repeated coalescence and fragmentation of island complexes



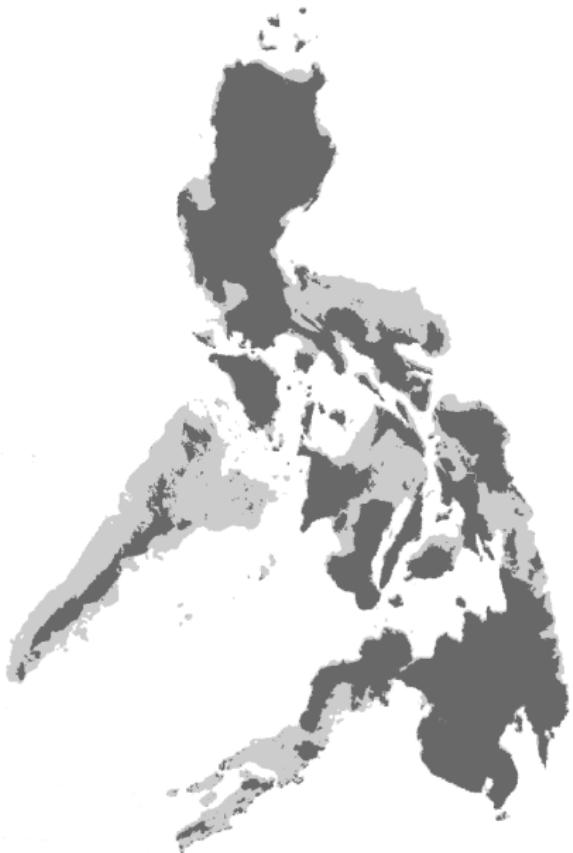
Climate-driven diversification model

- ▶ Repeated coalescence and fragmentation of island complexes
- ▶ Prominent paradigm for explaining Philippine biodiversity



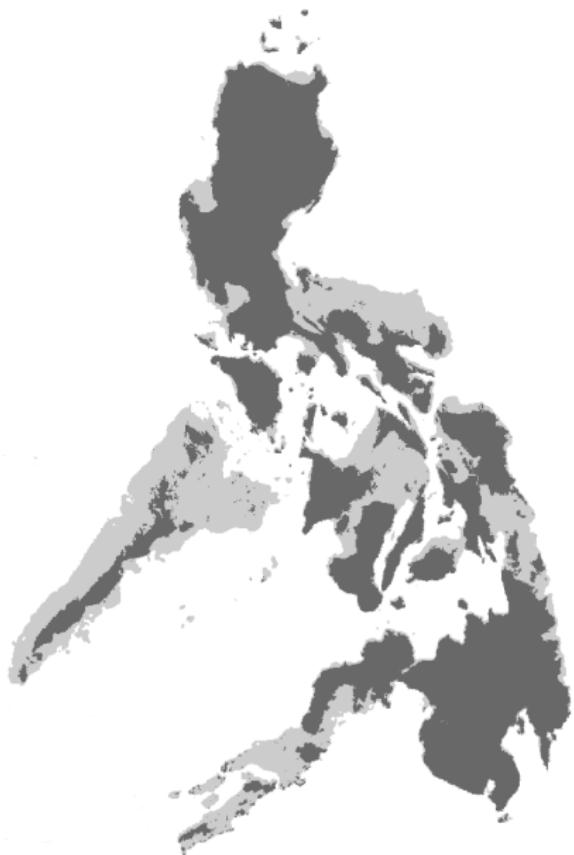
Climate-driven diversification model

- ▶ Repeated coalescence and fragmentation of island complexes
- ▶ Prominent paradigm for explaining Philippine biodiversity
- ▶ Proposed as model of diversification



Testing climate-driven diversification

**Did repeated fragmentation of
islands during inter-glacial
rises in sea level promote
diversification?**



Testing climate-driven diversification

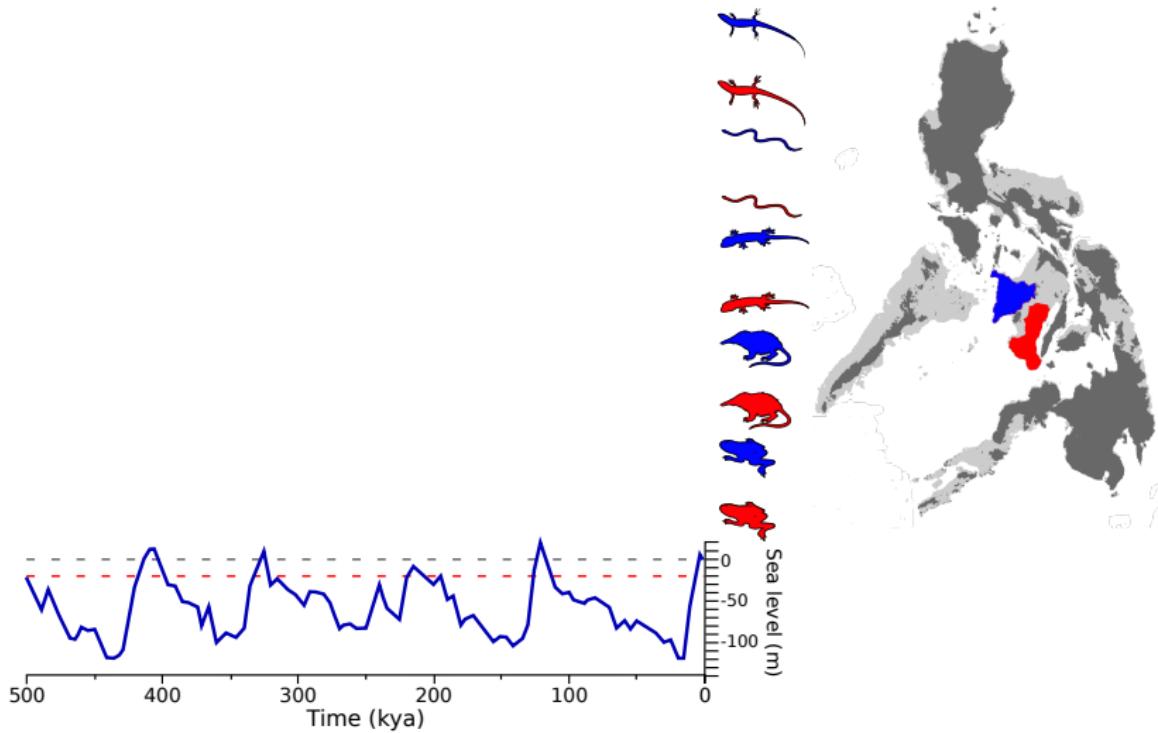
Did repeated fragmentation of islands during inter-glacial rises in sea level promote diversification?

Model has testable prediction:

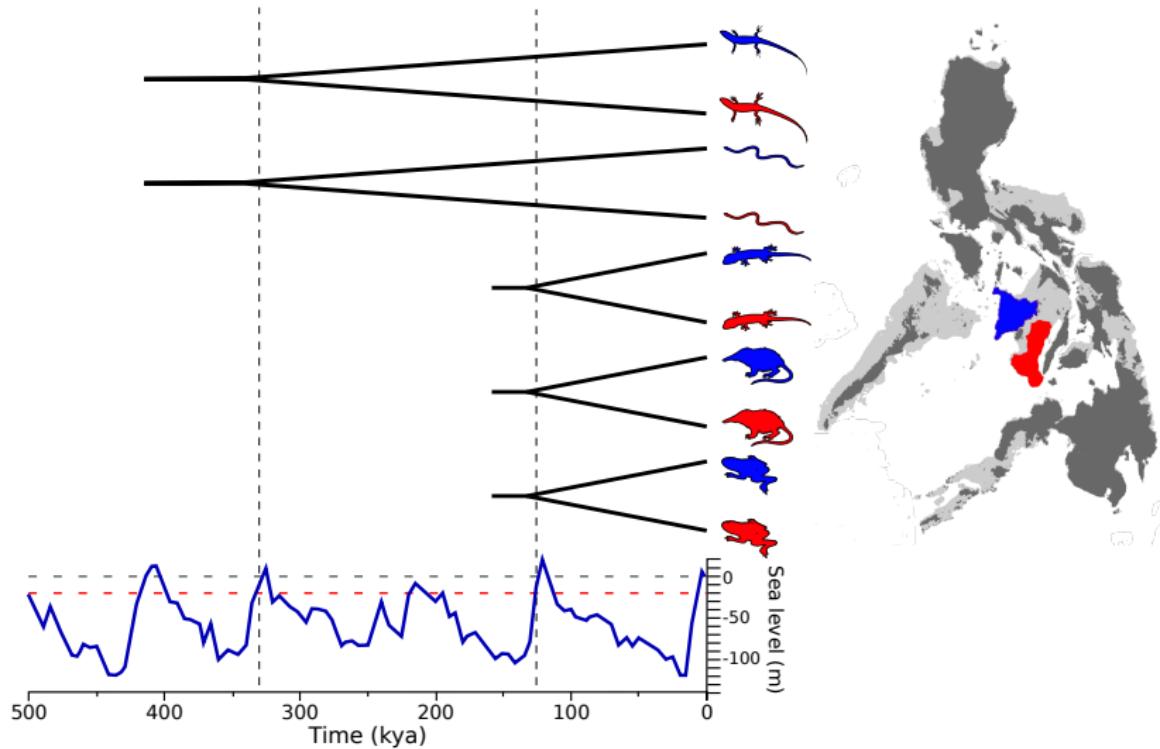
- ▶ Temporally clustered divergences among taxa co-distributed across fragmented islands



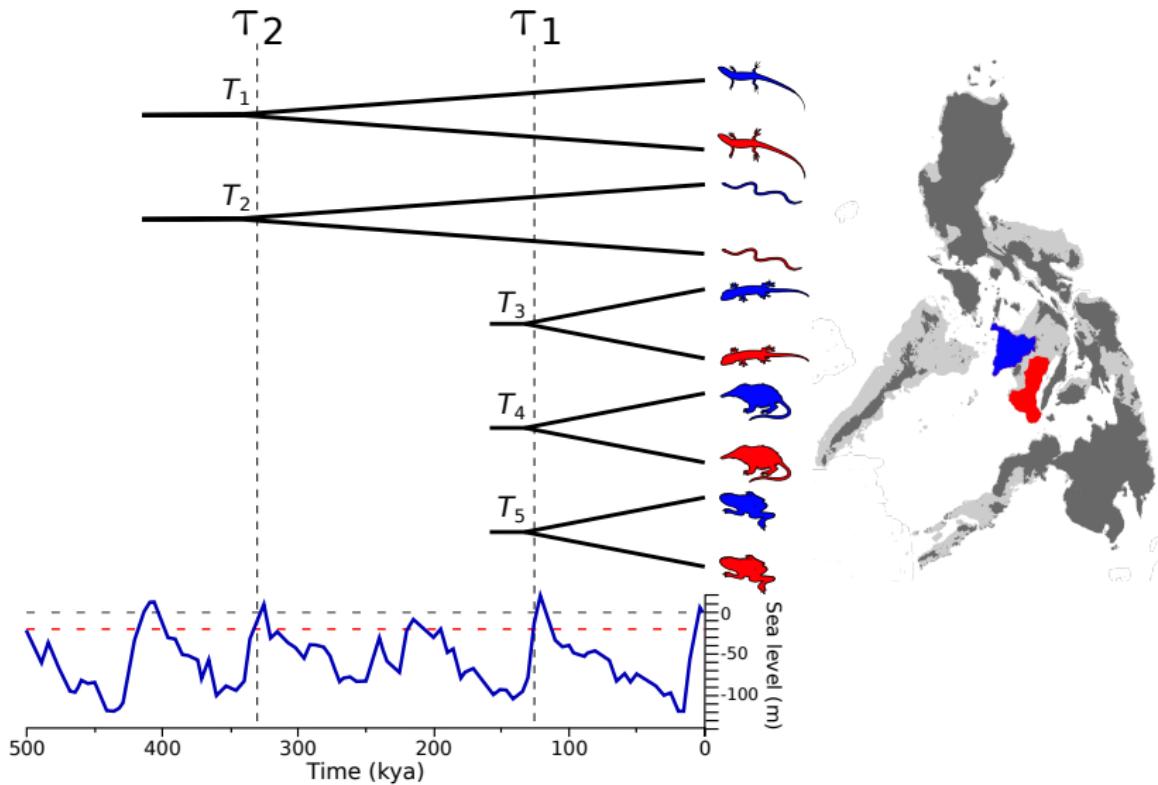
Climate-driven model: Prediction



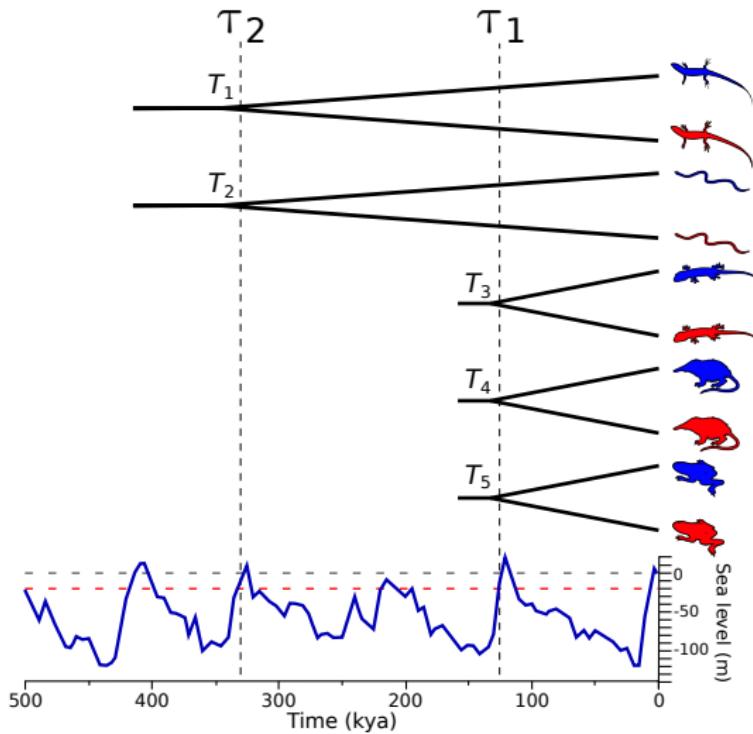
Climate-driven model: Prediction



Climate-driven model: Prediction



Divergence model choice

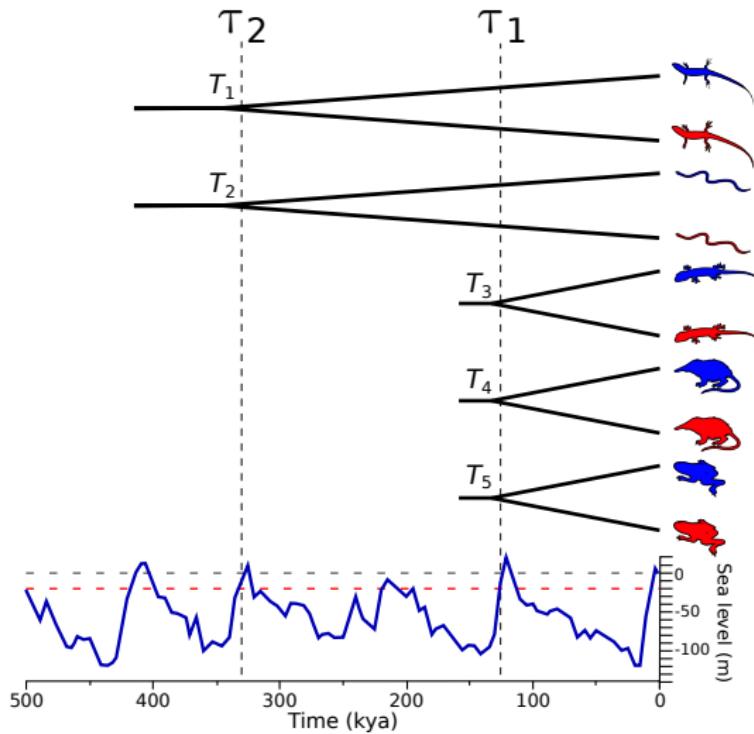


Divergence model choice

$$\mathbf{T} = (T_1, T_2, T_3, T_4, T_5)$$

$$\tau = \{\tau_1, \tau_2\}$$

$$|\tau| = 2$$

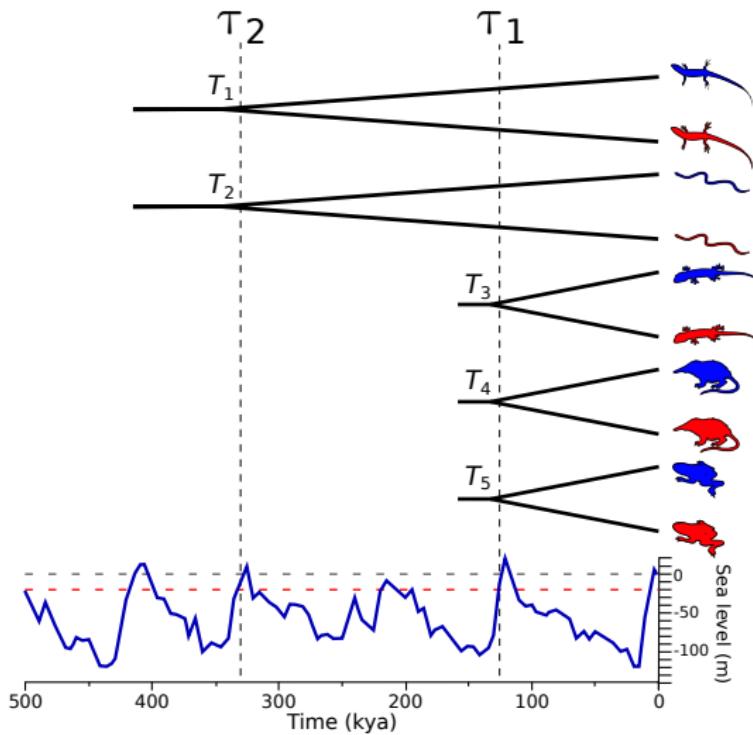


Divergence model choice

$$\mathbf{T} = (330, 330, 125, 125, 125)$$

$$\tau = \{125, 330\}$$

$$|\tau| = 2$$

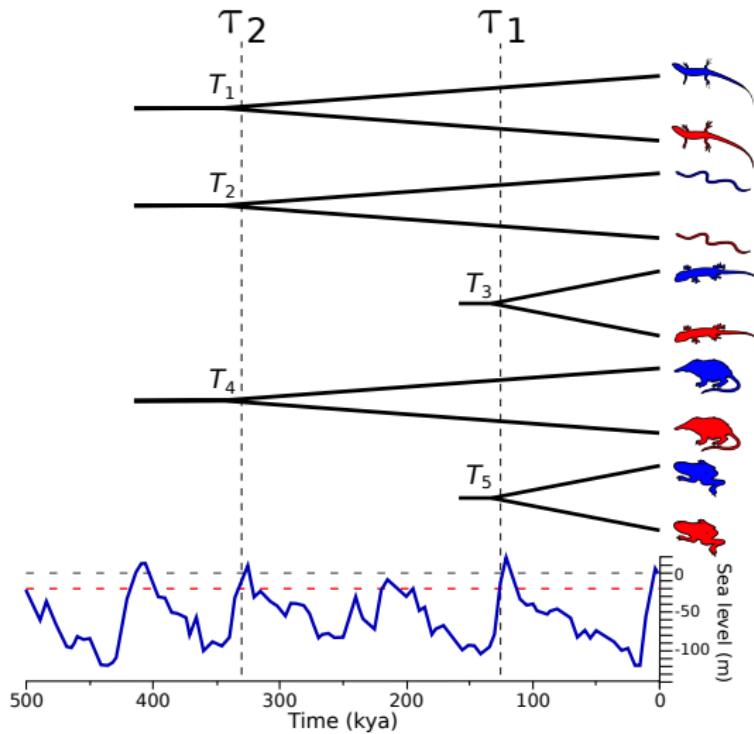


Divergence model choice

$$\mathbf{T} = (330, 330, 125, 330, 125)$$

$$\tau = \{125, 330\}$$

$$|\tau| = 2$$

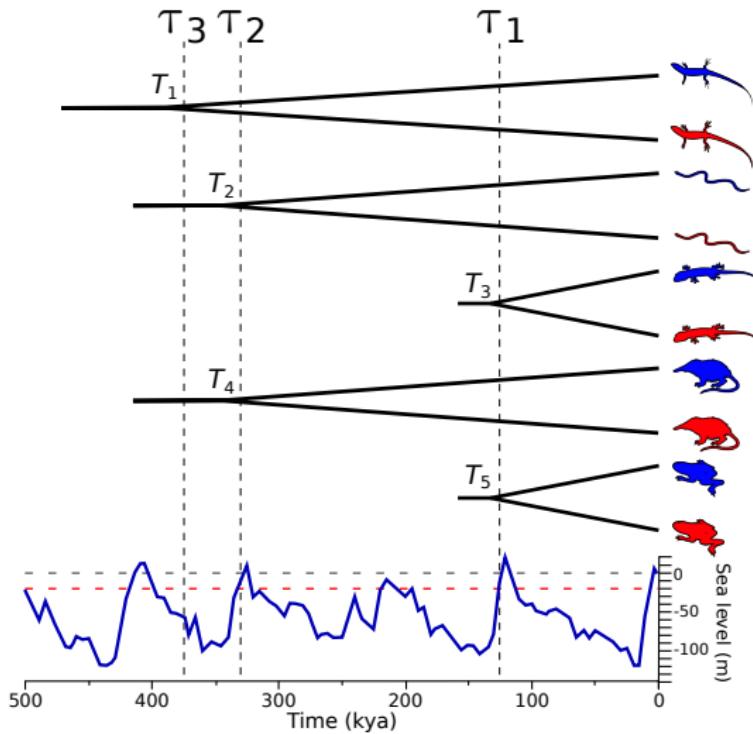


Divergence model choice

$$\mathbf{T} = (375, 330, 125, 330, 125)$$

$$\tau = \{125, 330, 375\}$$

$$|\tau| = 3$$

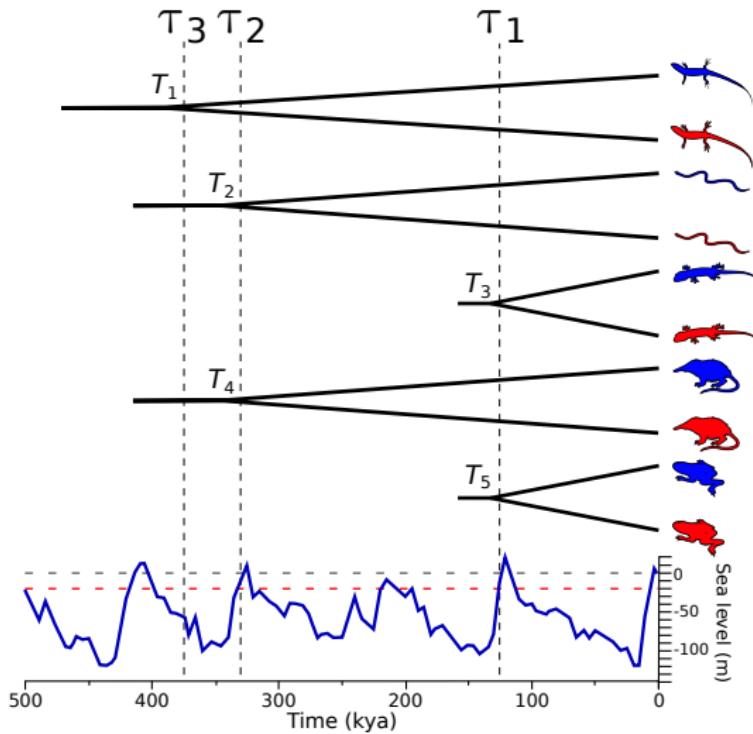


Divergence model choice

$$\mathbf{T} = (T_1, T_2, T_3, T_4, T_5)$$

$$\tau = \{\tau_1, \tau_2, \tau_3\}$$

$$|\tau| = 3$$

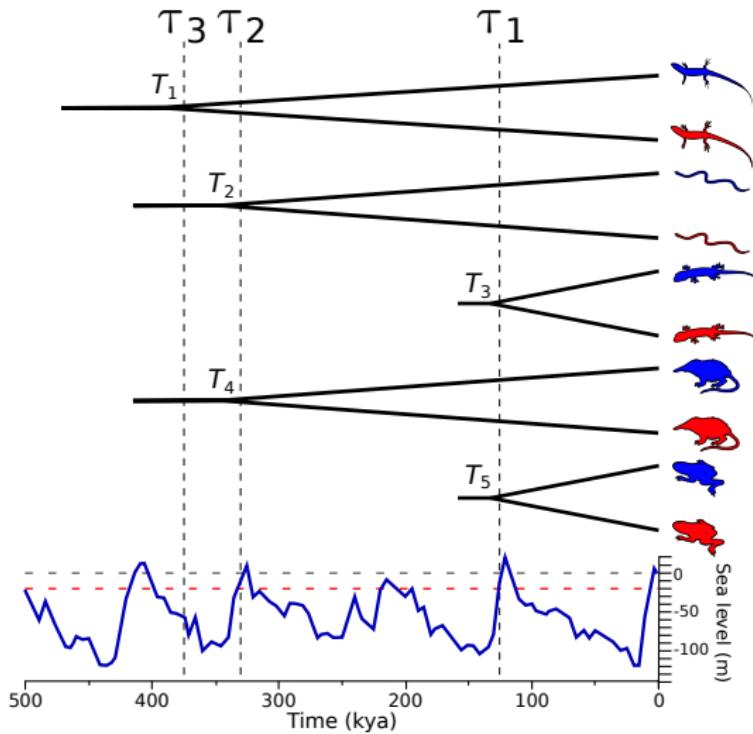


Divergence model choice

$$\mathbf{T} = (T_1, T_2, \dots, T_Y)$$

$$\tau = \{\tau_1, \dots, \tau_{|\tau|}\}$$

$$|\tau|$$



Divergence model choice

$$\mathbf{T} = (T_1, T_2, \dots, T_Y)$$

$$\boldsymbol{\tau} = \{\tau_1, \dots, \tau_{|\boldsymbol{\tau}|}\}$$

$$|\boldsymbol{\tau}|$$

- ▶ We want to infer \mathbf{T} given DNA sequence alignments \mathbf{X}

Divergence model choice

$$\mathbf{T} = (T_1, T_2, \dots, T_Y)$$

$$\boldsymbol{\tau} = \{\tau_1, \dots, \tau_{|\boldsymbol{\tau}|}\}$$

$$|\boldsymbol{\tau}|$$

- ▶ We want to infer \mathbf{T} given DNA sequence alignments \mathbf{X}



$$p(\mathbf{T} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{T})p(\mathbf{T})}{p(\mathbf{X})}$$

Divergence model choice

$$\mathbf{T} = (T_1, T_2, \dots, T_Y)$$

$$\boldsymbol{\tau} = \{\tau_1, \dots, \tau_{|\boldsymbol{\tau}|}\}$$

$$|\boldsymbol{\tau}|$$

- ▶ We want to infer \mathbf{T} given DNA sequence alignments \mathbf{X}
- ▶
$$p(\mathbf{T} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{T})p(\mathbf{T})}{p(\mathbf{X})}$$
- ▶ This approach implemented in msBayes

Divergence model choice

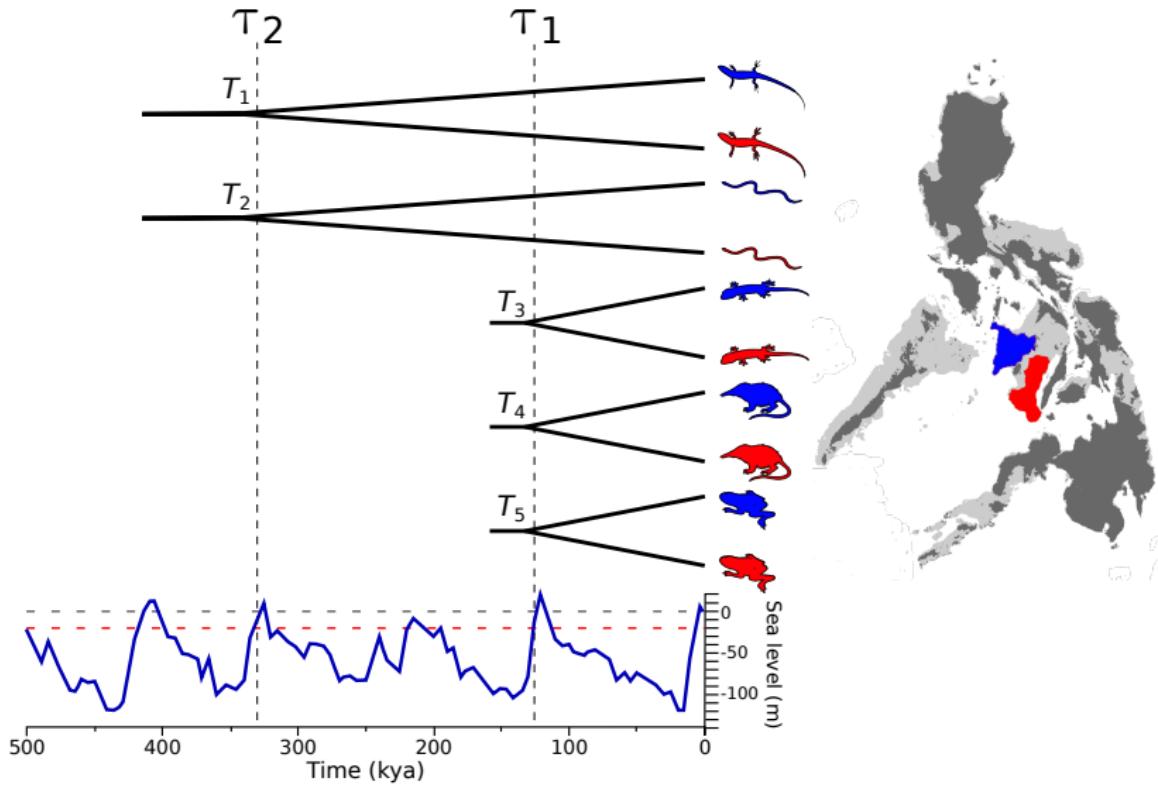
$$\mathbf{T} = (T_1, T_2, \dots, T_Y)$$

$$\boldsymbol{\tau} = \{\tau_1, \dots, \tau_{|\boldsymbol{\tau}|}\}$$

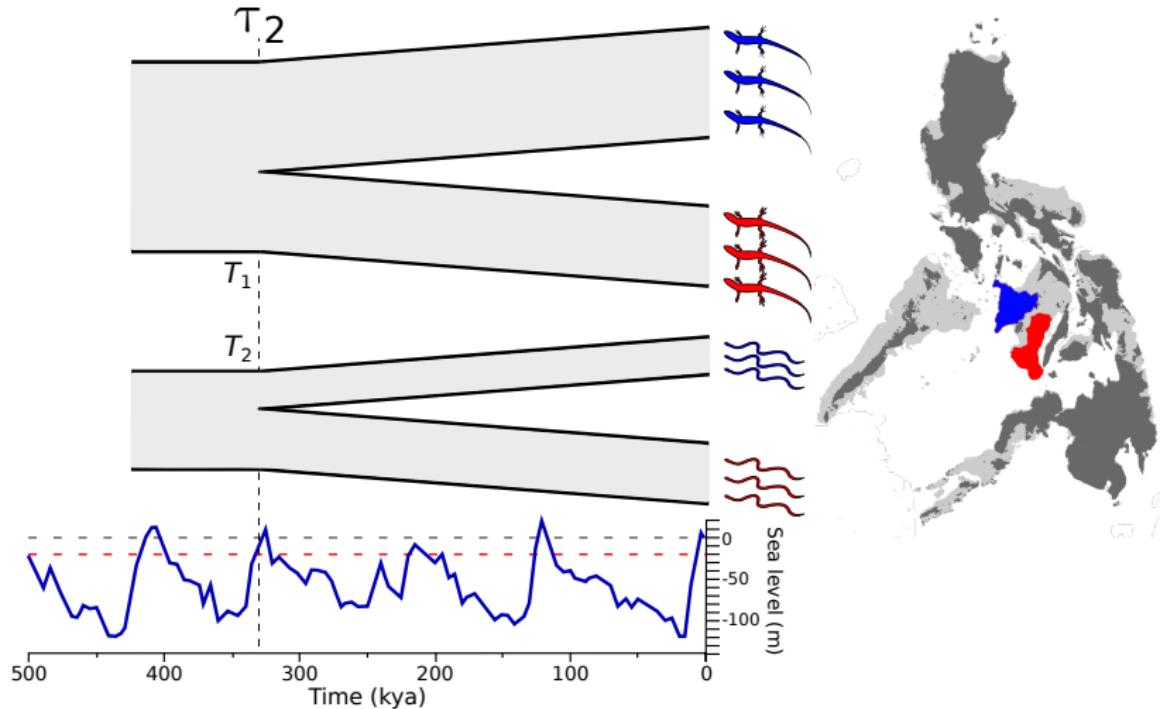
$$|\boldsymbol{\tau}|$$

- ▶ We want to infer \mathbf{T} given DNA sequence alignments \mathbf{X}
- ▶
$$p(\mathbf{T} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{T})p(\mathbf{T})}{p(\mathbf{X})}$$
- ▶ This approach implemented in msBayes
 - ▶ Not that simple

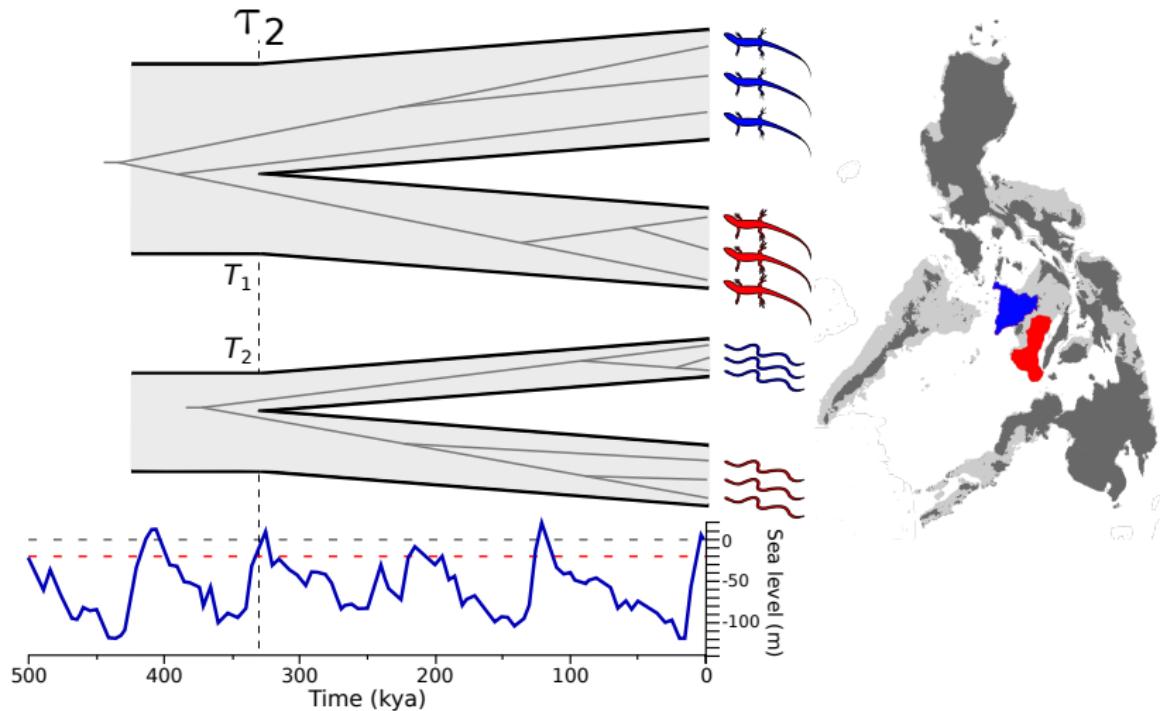
The msBayes model



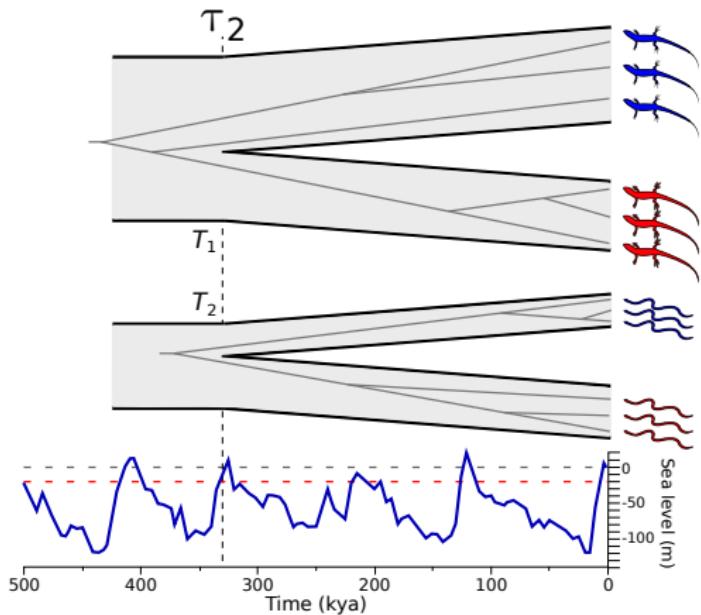
The msBayes model



The msBayes model

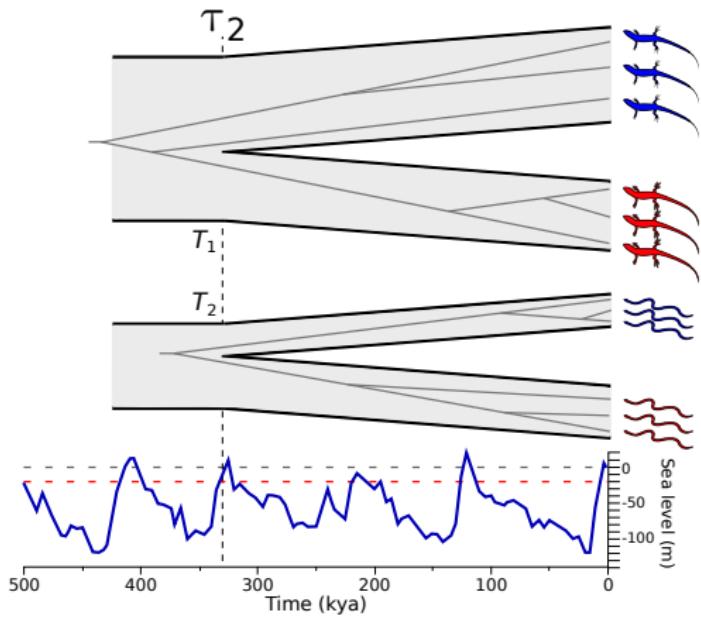


The msBayes model



The msBayes model

- X Sequence alignments
- G Gene trees
- T Divergence times
- Θ Demographic parameters



The msBayes model

- X** Sequence alignments
- G** Gene trees
- T** Divergence times
- Θ** Demographic parameters

The msBayes model

Full Model:

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(\mathbf{X})}$$

- X** Sequence alignments
- G** Gene trees
- T** Divergence times
- Θ** Demographic parameters

The msBayes model

Full Model:

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(\mathbf{X})}$$

$$p(\mathbf{G}, \mathbf{T}, \theta_A, \theta_{D1}, \theta_{D2}, \tau_B, \zeta_{D1}, \zeta_{D2}, m, \alpha, v | \mathbf{X}, \phi, \rho, \nu)$$

$$= \frac{1}{p(\mathbf{X})} p(\mathbf{T}) f(\alpha) \left[\prod_{i=1}^Y p(\theta_{A,i}) p(\theta_{D1,i}, \theta_{D2,i}) p(\tau_{B,i}) p(\zeta_{D1,i}) f(\zeta_{D2,i}) p(m_i) \right. \\ \left. \prod_{j=1}^{k_j} p(X_{i,j} | G_{i,j}, \phi_{i,j}) p(G_{i,j} | T_i, \theta_{A,i}, \theta_{D1,i}, \theta_{D2,i}, \rho_{i,j}, \nu_{i,j}, v_j, \tau_{B,i}, \zeta_{D1,i}, \zeta_{D2,i}, m_i) \right] \left[\prod_{j=1}^K f(v_j | \alpha) \right]$$

The msBayes model

Full Model:

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(\mathbf{X})}$$

Approximate Bayesian computation (ABC)

$$\mathbf{X} \rightarrow \mathbf{S}^* \rightarrow B_\epsilon(\mathbf{S}^*)$$

The msBayes model

Full Model:

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(\mathbf{X})}$$

Approximate Bayesian computation (ABC)

$$\mathbf{X} \rightarrow \mathbf{S}^* \rightarrow B_\epsilon(\mathbf{S}^*)$$

Approximate Model:

$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | B_\epsilon(\mathbf{S}^*)) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(B_\epsilon(\mathbf{S}^*))}$$

The msBayes model

Full Model:

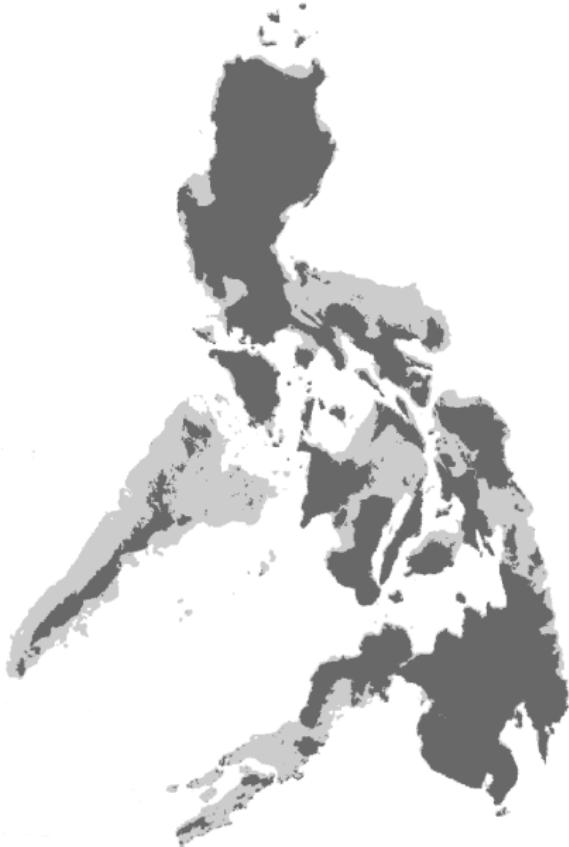
$$p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta} | \mathbf{X}) = \frac{p(\mathbf{X} | \mathbf{G}, \mathbf{T}, \boldsymbol{\Theta}) p(\mathbf{G}, \mathbf{T}, \boldsymbol{\Theta})}{p(\mathbf{X})}$$

- \mathbf{T} Vector of divergence times across pairs of populations
- $|\tau|$ Number of divergence parameters
- D_T The variance of \mathbf{T}

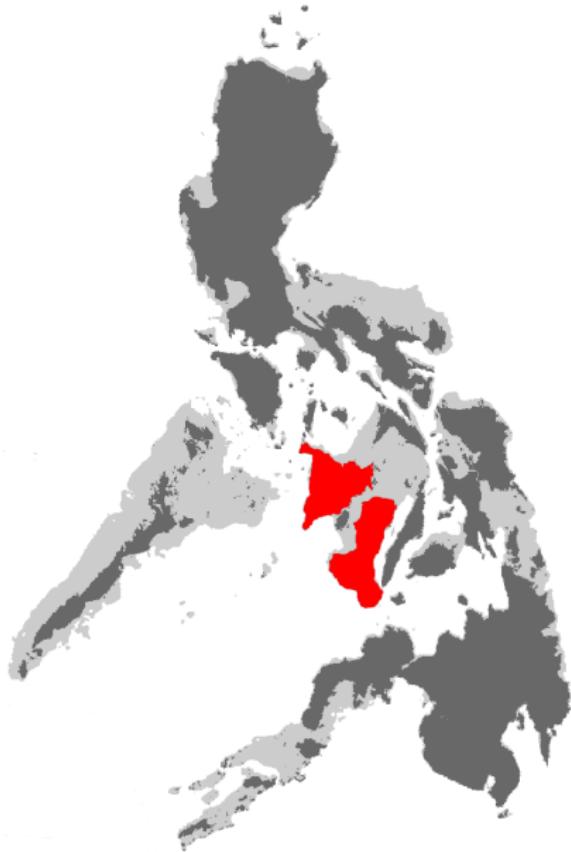
Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



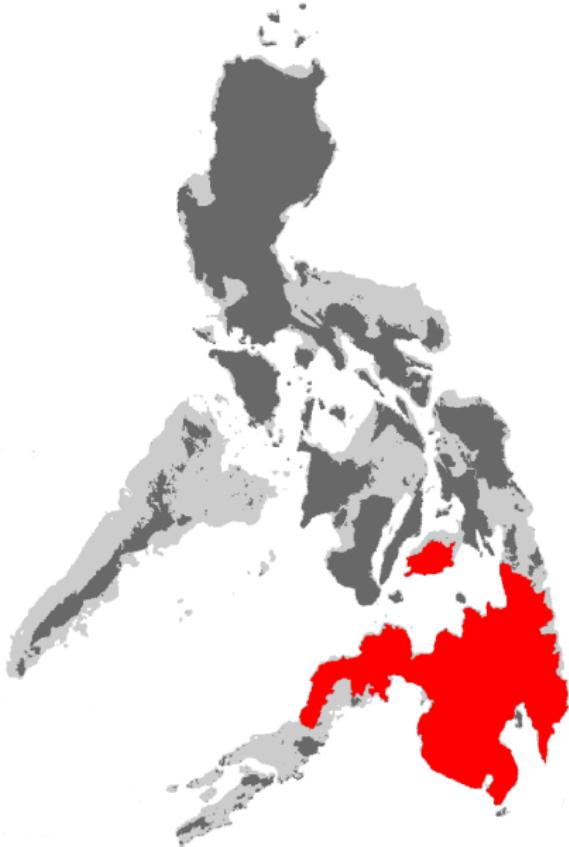
Species	n₁	n₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



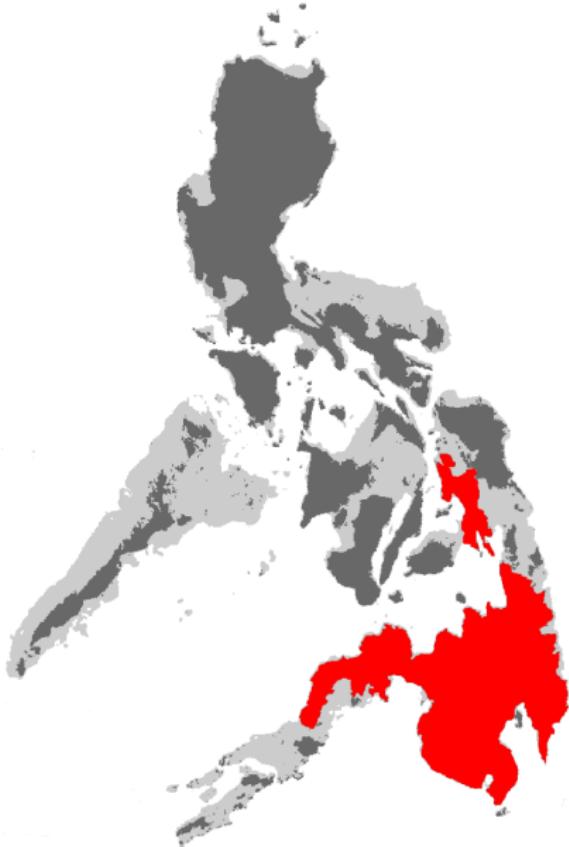
Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



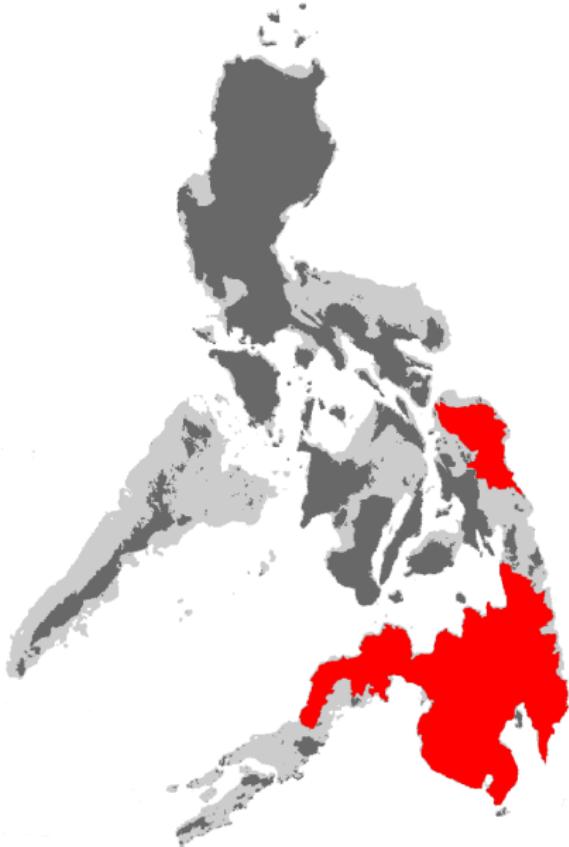
Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



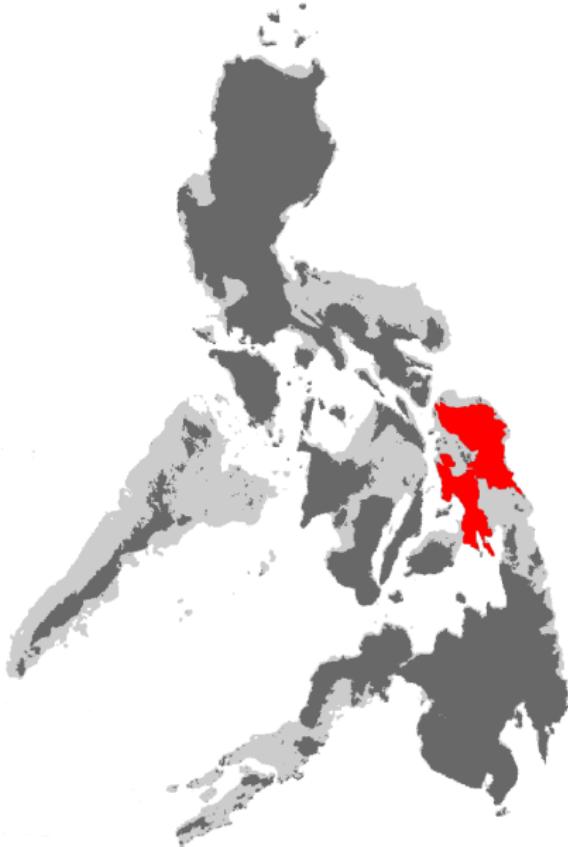
Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3

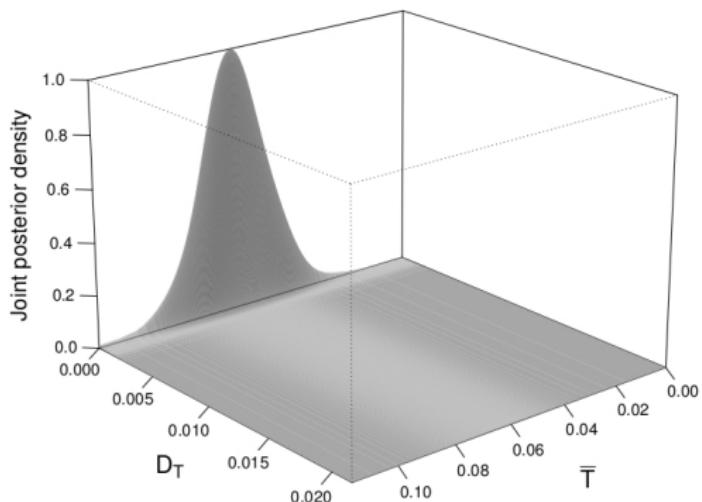


Empirical results

Strong support for simultaneous divergence of all 22 taxon pairs

pp > 0.96

~100,000–250,000 years ago



Simulation-based power analyses

What is “simultaneous”?

Simulation-based power analyses

What is “simultaneous”?

- ▶ Simulate datasets in which all 22 divergence times are random

Simulation-based power analyses

What is “simultaneous”?

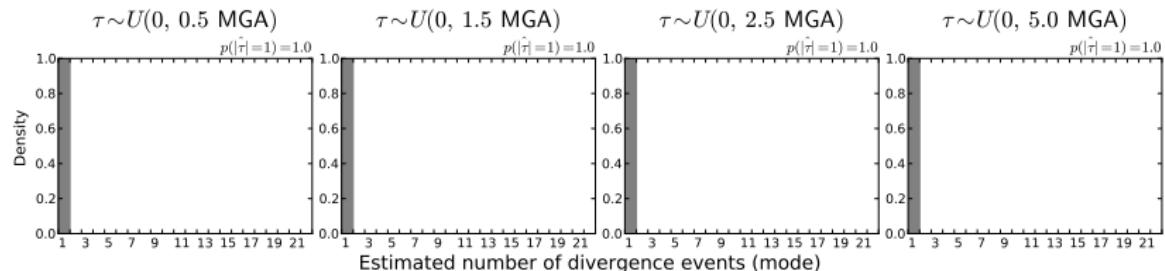
- ▶ Simulate datasets in which all 22 divergence times are random
 - ▶ $\tau \sim U(0, 0.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 1.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 2.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 5.0 \text{ MGA})$
- ▶ $MGA = \text{Millions of Generations Ago}$

Simulation-based power analyses

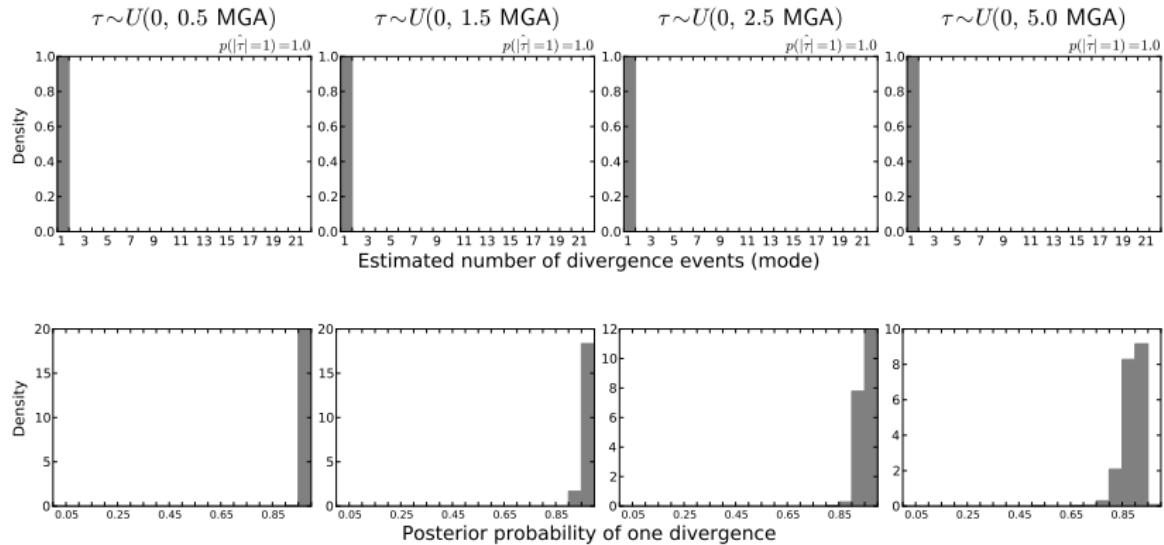
What is “simultaneous”?

- ▶ Simulate datasets in which all 22 divergence times are random
 - ▶ $\tau \sim U(0, 0.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 1.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 2.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 5.0 \text{ MGA})$
- ▶ $MGA = \text{Millions of Generations Ago}$
- ▶ Simulate 1000 datasets for each τ distribution
- ▶ Analyze all 4000 datasets as we did the empirical data

Simulation-based power analyses: Results



Simulation-based power analyses: Results



Simulation-based power analyses: Results

Strong support for highly clustered divergences when divergence times are random over 5 million generations

Our empirical results are likely spurious

Why the bias?

Potential causes of the bias:

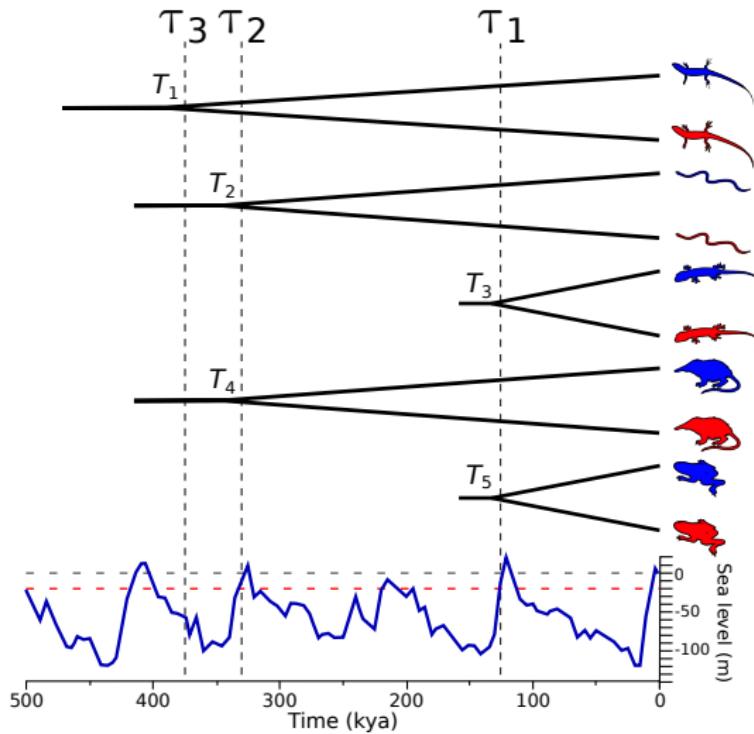
1. The prior on divergence models
2. Broad uniform priors on many of the model's parameters, including divergence times

Causes of bias: Prior on divergence models

$$\mathbf{T} = (375, 330, 125, 330, 125)$$

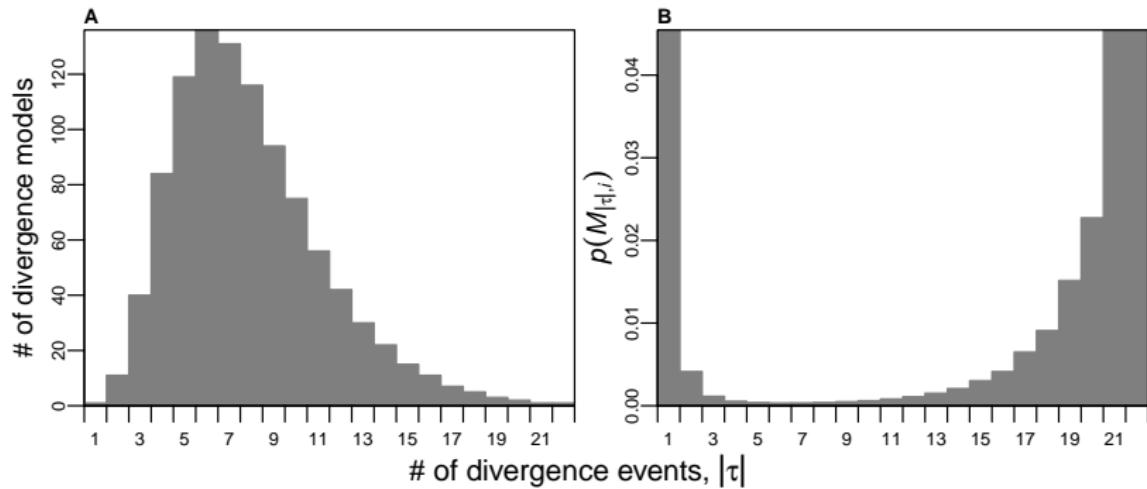
$$\tau = \{125, 330, 375\}$$

$$|\tau| = 3$$



Causes of bias: Prior on divergence models

- ▶ msBayes uses a discrete uniform prior on the *number* of divergence events, $|\tau|$



Causes of bias: Broad priors

- ▶ msBayes uses uniform priors on most model parameters, including divergence times

Causes of bias: Broad priors

- ▶ msBayes uses uniform priors on most model parameters, including divergence times
- ▶ This requires the use of broad priors

Causes of bias: Broad priors

- ▶ msBayes uses uniform priors on most model parameters, including divergence times
- ▶ This requires the use of broad priors
- ▶ Models with more divergence-time parameters have much greater parameter space, much of it with low likelihood

Causes of bias: Broad priors

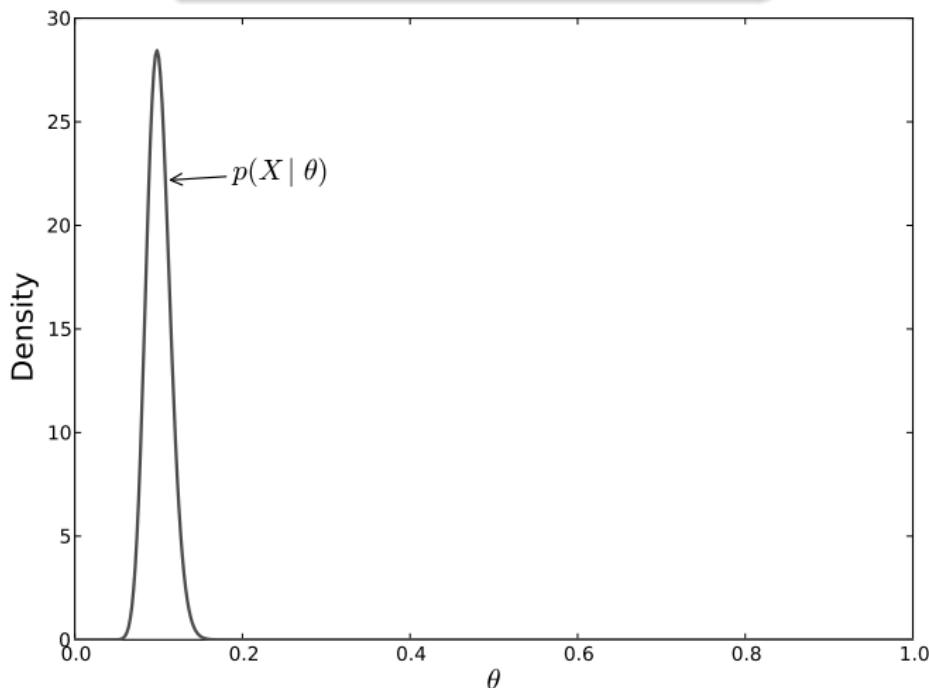
- ▶ msBayes uses uniform priors on most model parameters, including divergence times
- ▶ This requires the use of broad priors
- ▶ Models with more divergence-time parameters have much greater parameter space, much of it with low likelihood
- ▶ This vast space can cause problems with Bayesian model choice
 - ▶ Reduced marginal likelihoods

Causes of bias: Marginal likelihoods

$$p(X) = \int_{\theta} p(X | \theta)p(\theta)d\theta$$

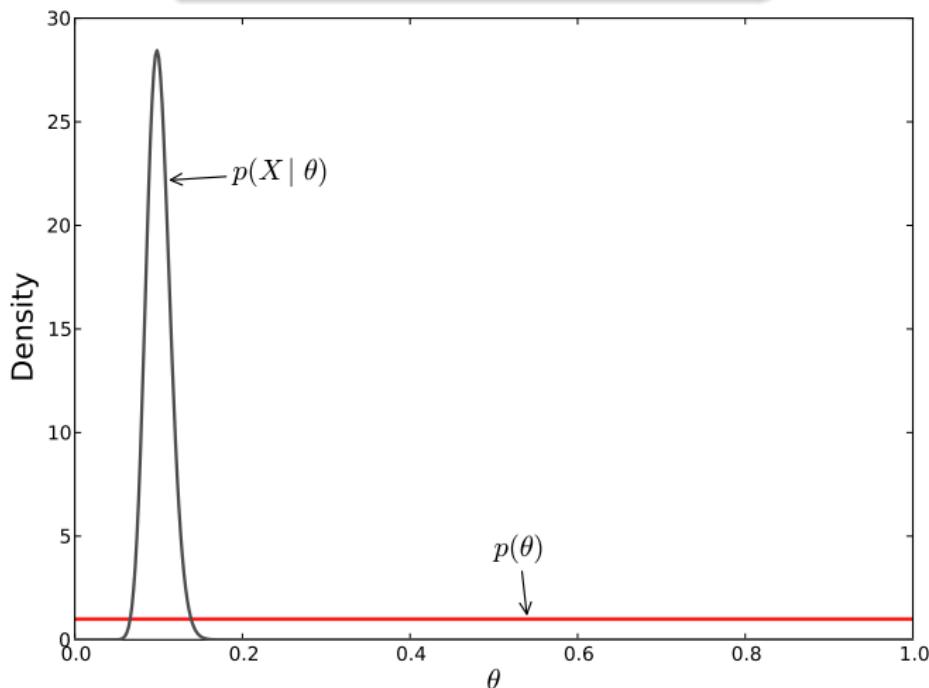
Causes of bias: Marginal likelihoods

$$p(X) = \int_{\theta} p(X | \theta)p(\theta)d\theta$$

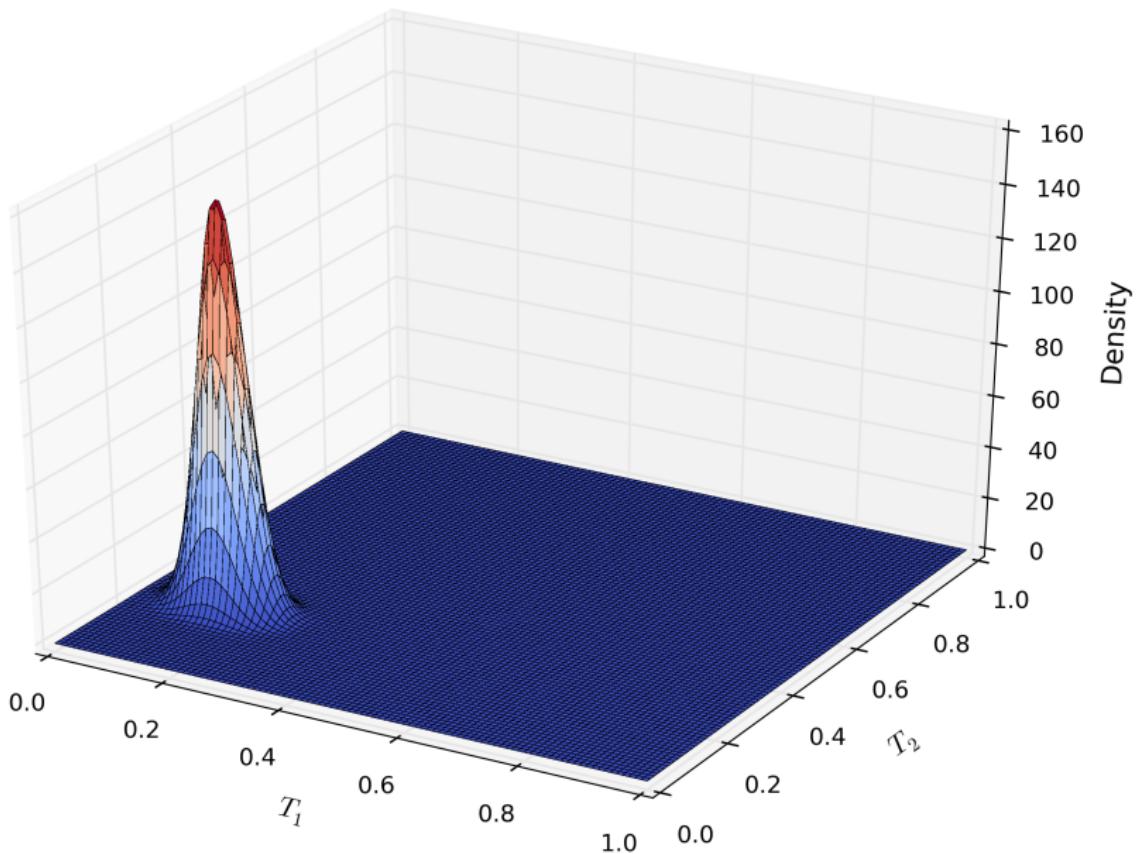


Causes of bias: Marginal likelihoods

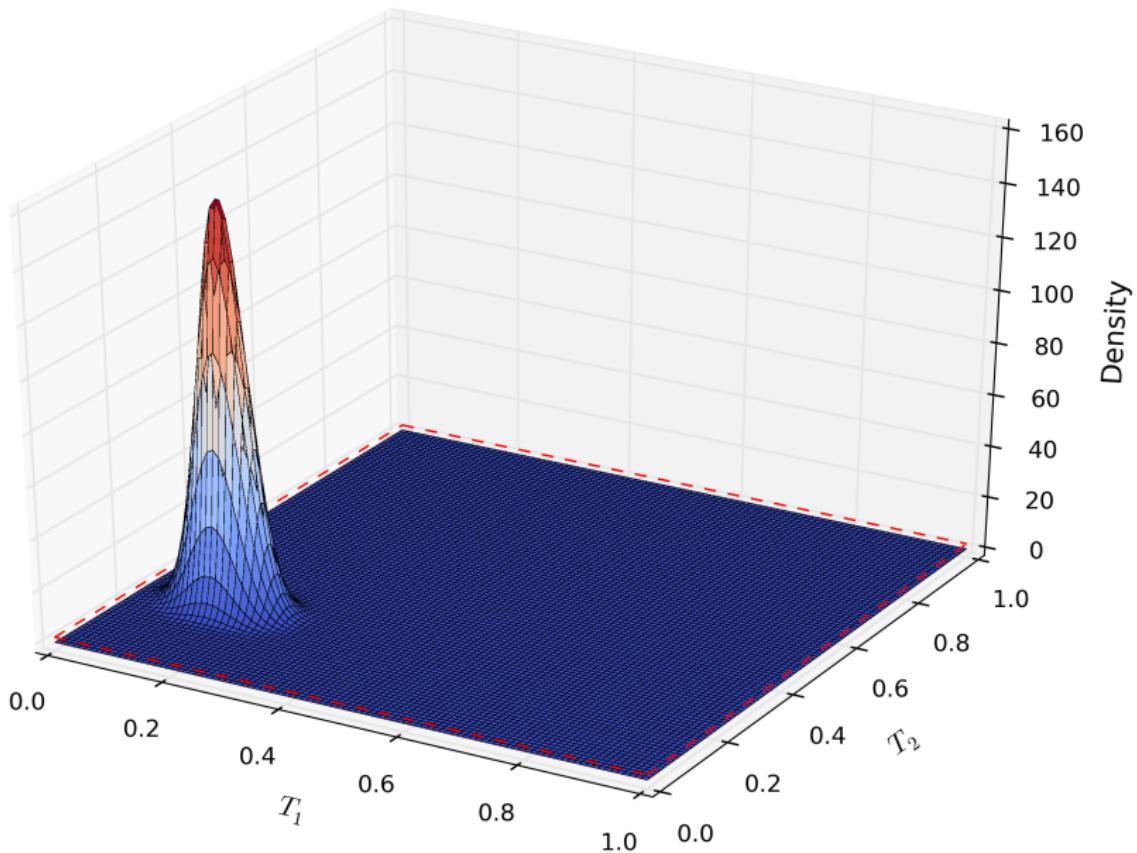
$$p(X) = \int_{\theta} p(X | \theta)p(\theta)d\theta$$



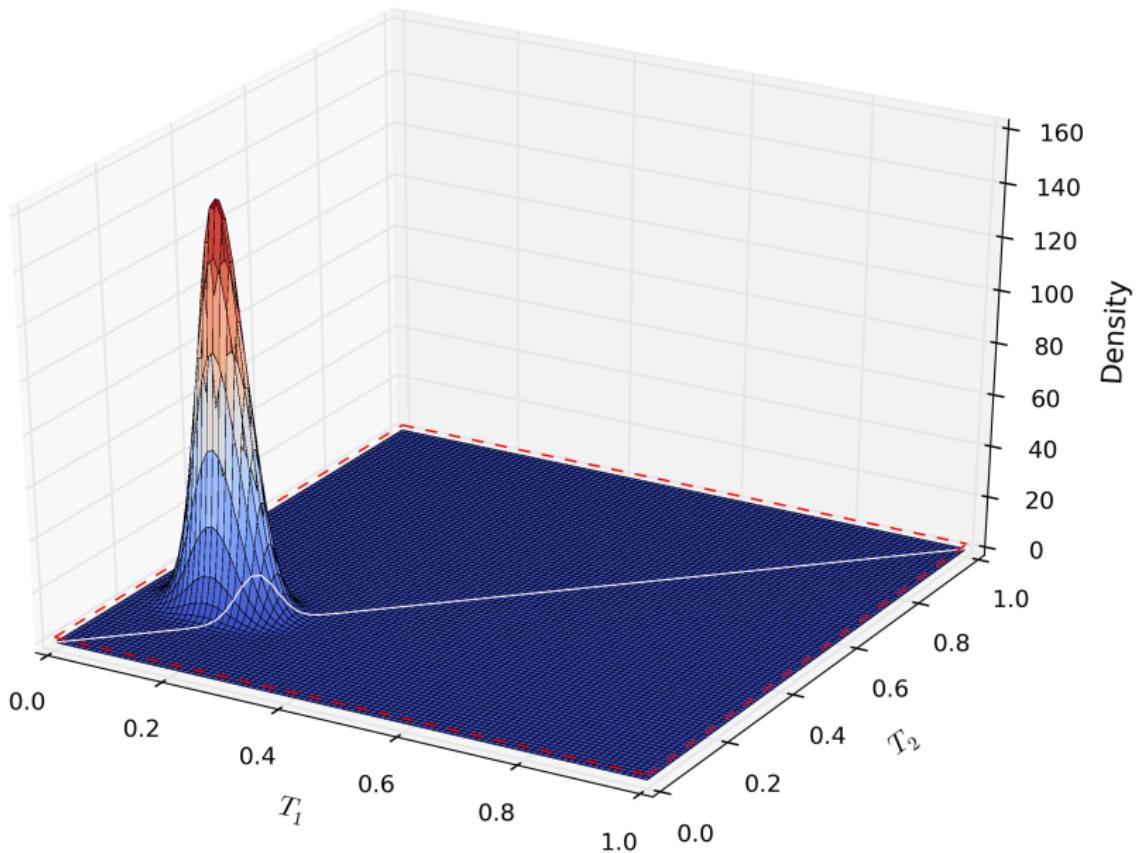
Causes of bias: Marginal likelihoods



Causes of bias: Marginal likelihoods



Causes of bias: Marginal likelihoods



Causes of bias: Marginal likelihoods

$$p(\theta | X) = \frac{p(X | \theta)p(\theta)}{p(X)}$$

$$p(X) = \int_{\theta} p(X | \theta)p(\theta)d\theta$$

Causes of bias: Marginal likelihoods

$$p(\theta_1 | X, M_1) = \frac{p(X | \theta_1, M_1)p(\theta_1 | M_1)}{p(X | M_1)}$$

$$p(X | M_1) = \int_{\theta_1} p(X | \theta_1, M_1)p(\theta_1 | M_1)d\theta_1$$

Causes of bias: Marginal likelihoods

$$p(\theta_1 | X, M_1) = \frac{p(X | \theta_1, M_1)p(\theta_1 | M_1)}{p(X | M_1)}$$

$$p(X | M_1) = \int_{\theta_1} p(X | \theta_1, M_1)p(\theta | M_1)d\theta_1$$

$$p(M_1 | X) = \frac{p(X | M_1)p(M_1)}{p(X | M_1)p(M_1) + p(X | M_2)p(M_2)}$$

Causes of bias: Marginal likelihoods

Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

Causes of bias: Marginal likelihoods

Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

Testing prior sensitivity:

Causes of bias: Marginal likelihoods

Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

Testing prior sensitivity:

1. Analyze empirical data under several different prior settings
 - ▶ Results are very sensitive

Causes of bias: Marginal likelihoods

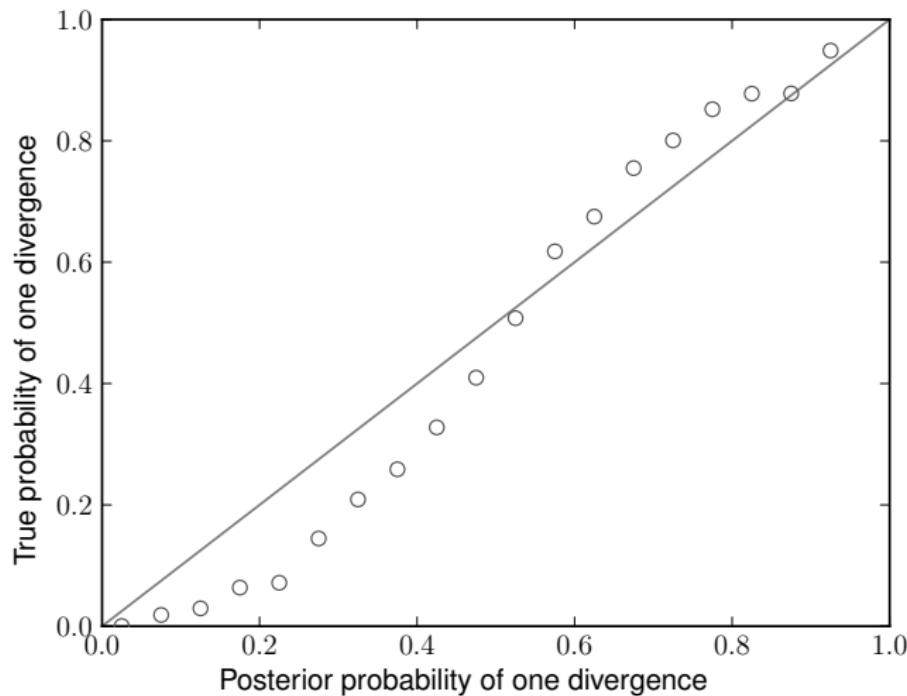
Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

Testing prior sensitivity:

1. Analyze empirical data under several different prior settings
 - ▶ Results are very sensitive
2. Use simulations to assess behavior when priors are correct

Simulation results: Performance when priors are correct



msBayes performs well when all assumptions are met

Causes of bias: Marginal likelihoods

Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

Testing prior sensitivity:

1. Analyze empirical data under several different prior settings
 - ▶ Results are very sensitive
2. Use simulations to assess behavior when priors are correct

Causes of bias: Marginal likelihoods

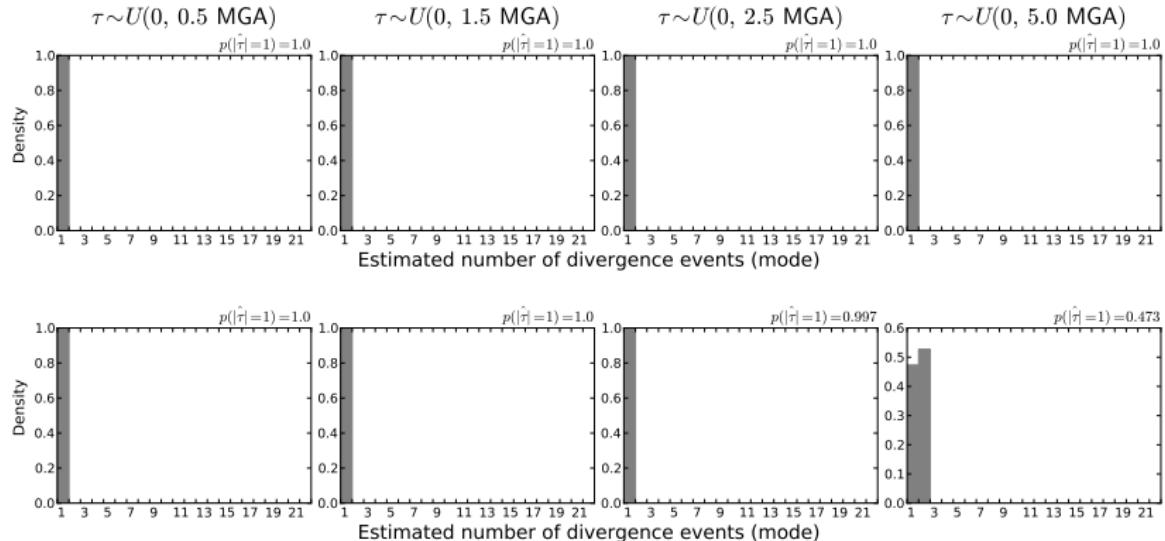
Predictions:

- ▶ Posterior estimates should be sensitive to priors
- ▶ As prior converges to distribution underlying the data, the bias should disappear

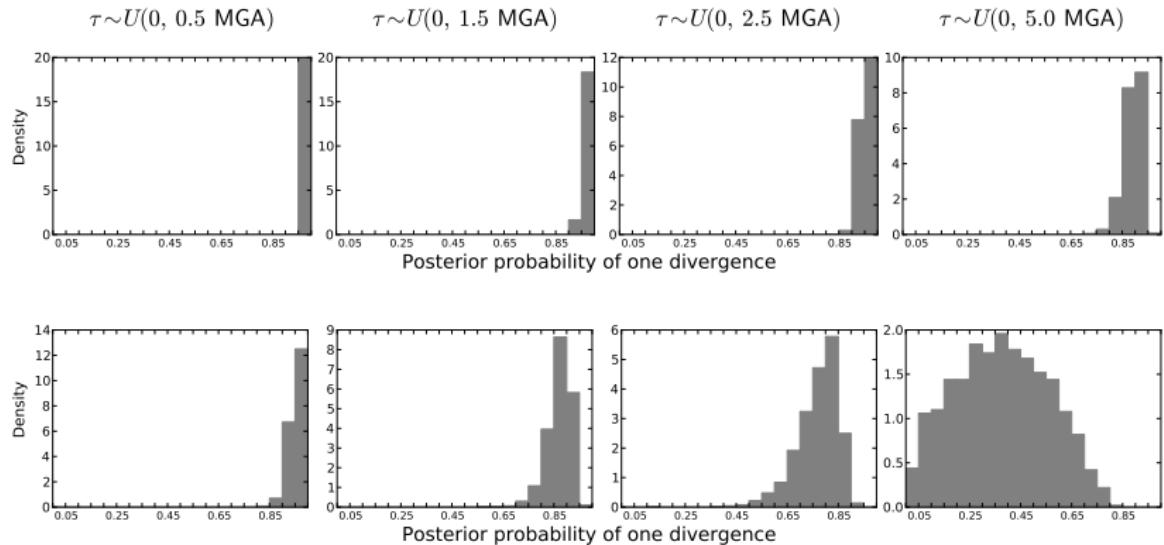
Testing prior sensitivity:

1. Analyze empirical data under several different prior settings
 - ▶ Results are very sensitive
2. Use simulations to assess behavior when priors are correct
3. Use simulations to assess behavior under “ideal” real-world priors

Simulation results: Power with informed priors



Simulation results: Power with informed priors



Causes of bias: Simulation results

Broad uniform priors are reducing marginal likelihoods of models with more divergence events

Even when uniform priors are informed by the data the bias remains

Causes of bias: Simulation results

Broad uniform priors are reducing marginal likelihoods of models with more divergence events

Even when uniform priors are informed by the data the bias remains

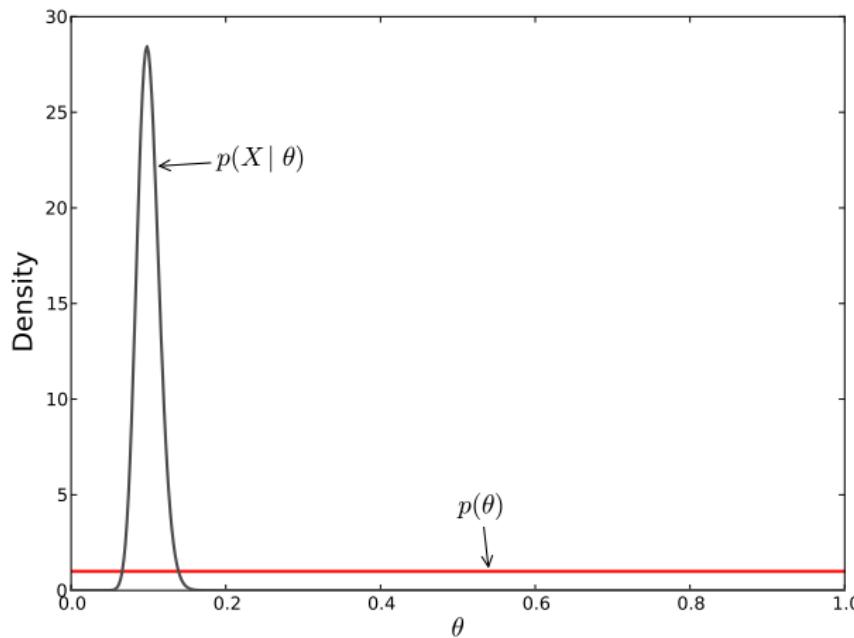
Potential solution:

More flexible priors

Mitigating the bias

Potential solution:

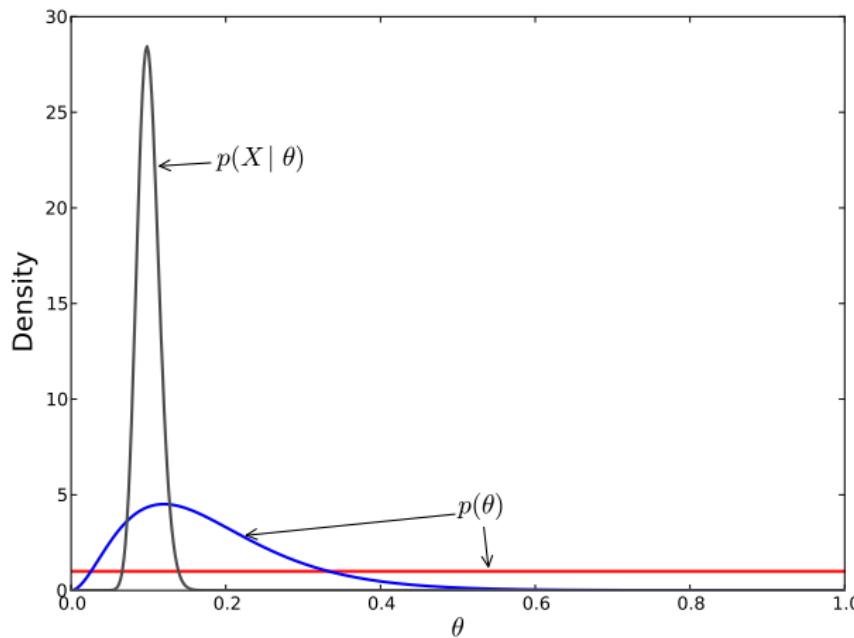
More flexible priors



Mitigating the bias

Potential solution:

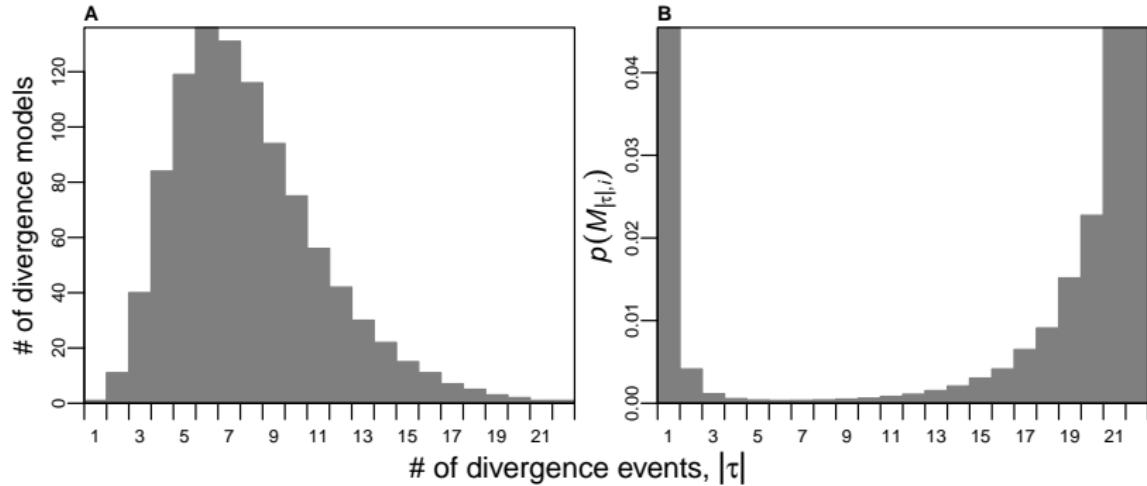
More flexible priors



Mitigating the bias

Potential solution:

More flexible priors



Mitigating the bias

Potential solution:

More flexible priors

Potential solution:

Alternative prior over divergence models (e.g., uniform or Dirichlet process)

New method: dpp-msbayes

- ▶ Reparameterized the model implemented in msBayes

New method: dpp-msbayes

- ▶ Reparameterized the model implemented in msBayes
- ▶ Replaced uniform priors on continuous parameters with gamma and beta distributions

New method: dpp-msbayes

- ▶ Reparameterized the model implemented in msBayes
- ▶ Replaced uniform priors on continuous parameters with gamma and beta distributions
- ▶ Dirichlet process prior (DPP) over all possible discrete divergence models

New method: dpp-msbayes

- ▶ Reparameterized the model implemented in msBayes
- ▶ Replaced uniform priors on continuous parameters with gamma and beta distributions
- ▶ Dirichlet process prior (DPP) over all possible discrete divergence models
- ▶ Uniform prior over divergence models

dpp-msbayes: Simulation-based assessment

Simulate 50,000 datasets under four models

- $M_{msBayes}$ ▶ U-shaped prior on divergence models
▶ Uniform priors on continuous parameters
- $M_{Ushaped}$ ▶ U-shaped prior on divergence models
▶ Gamma priors on continuous parameters
- $M_{Uniform}$ ▶ Uniform prior on divergence models
▶ Gamma priors on continuous parameters
- M_{DPP} ▶ DPP prior on divergence models
▶ Gamma priors on continuous parameters

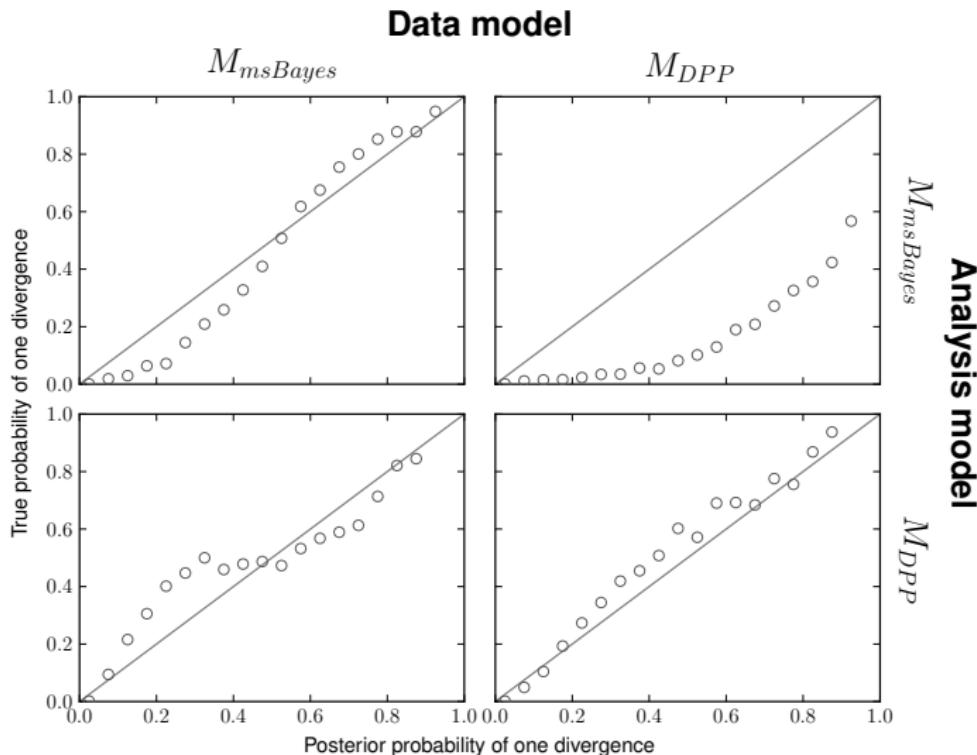
Analyze all datasets under each of the models

dpp-msbayes: Simulation-based assessment

Assess power

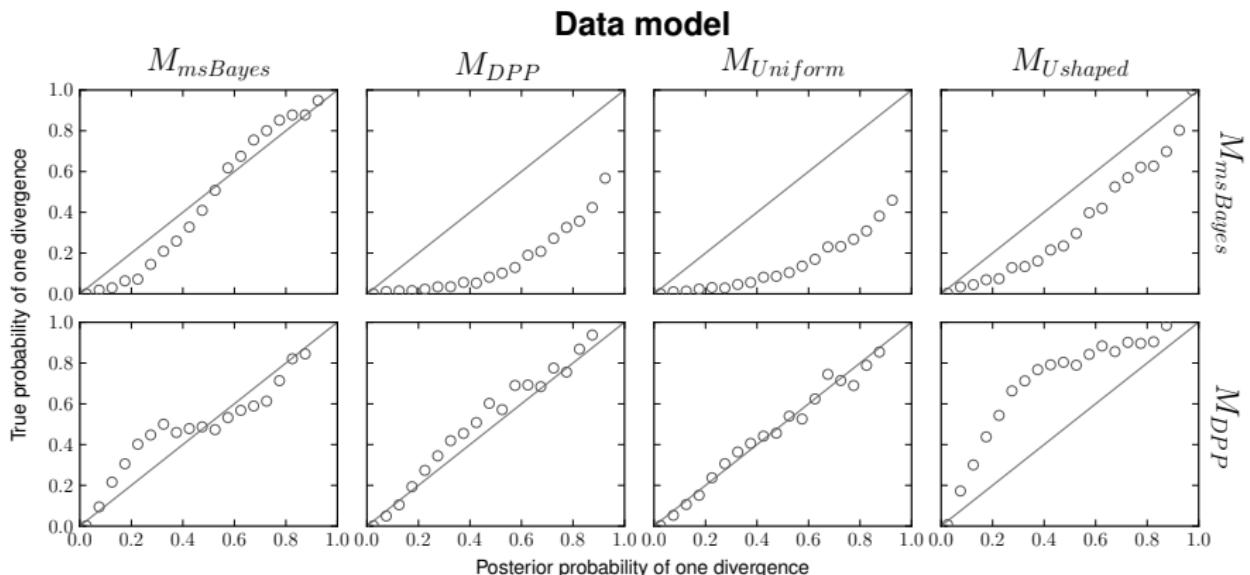
- ▶ Simulate datasets in which all 22 divergence times are random
 - ▶ $\tau \sim U(0, 0.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 1.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 2.5 \text{ MGA})$
 - ▶ $\tau \sim U(0, 5.0 \text{ MGA})$
- ▶ $\text{MGA} = \text{Millions of Generations Ago}$
- ▶ Simulate 1000 datasets for each τ distribution
- ▶ Analyze all 4000 datasets as we did the empirical data

dpp-msbayes: Simulation results

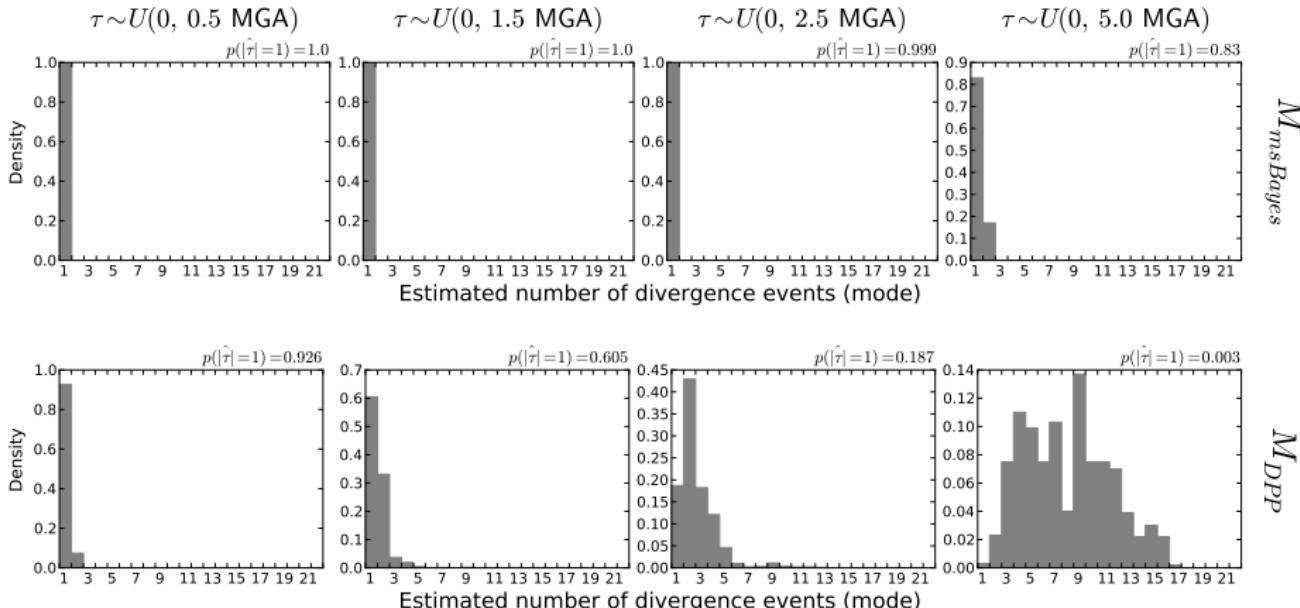


dpp-msbayes: Simulation results

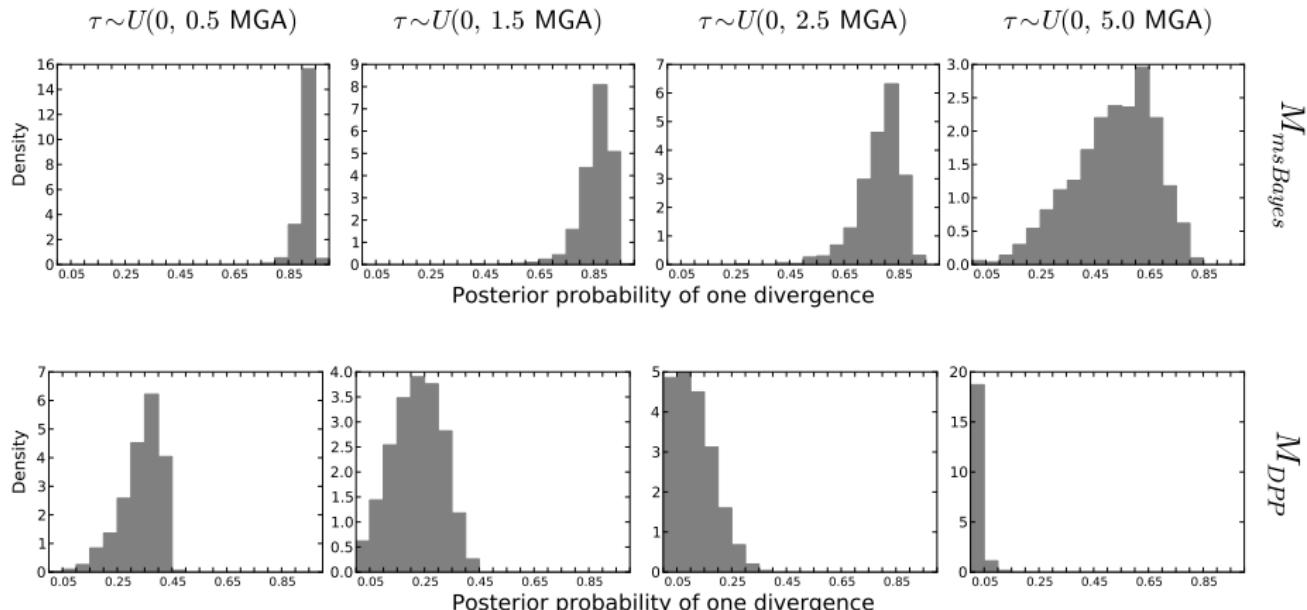
Analysis model



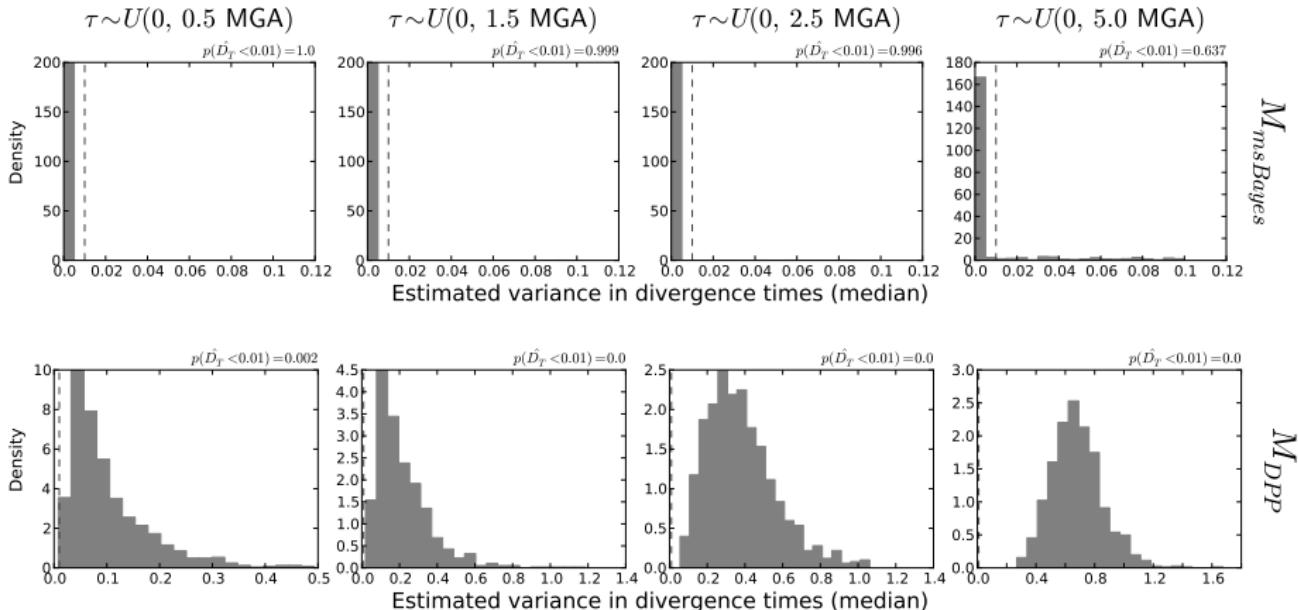
dpp-msbayes: Simulation results



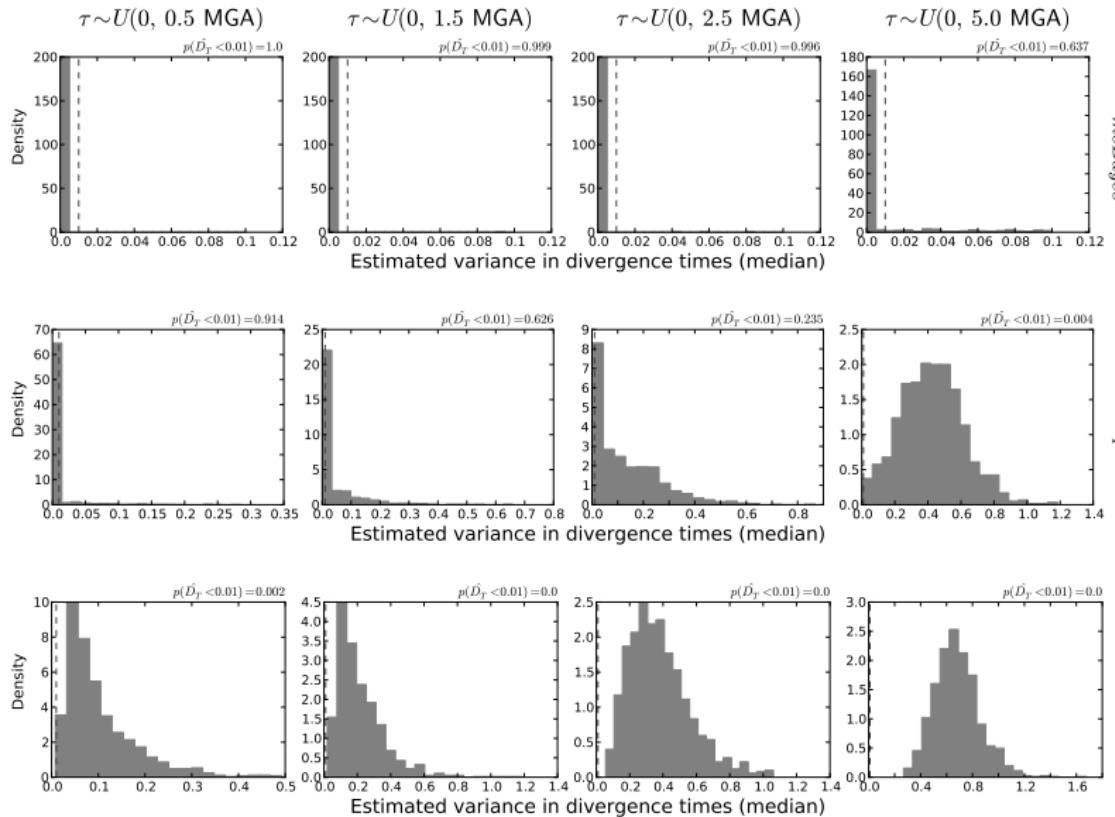
dpp-msbayes: Simulation results



dpp-msbayes: Simulation results



dpp-msbayes: Simulation results

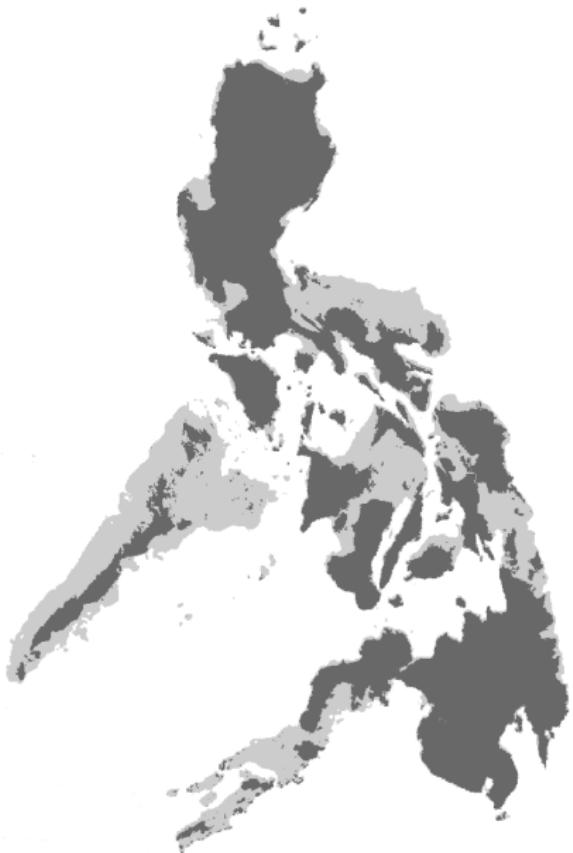


dpp-msbayes: Simulation results

- ▶ Results confirm the bias of msBayes was caused by
 1. Broad uniform priors
 2. U-shaped prior on divergence models
- ▶ The new model shows improved model-choice accuracy, power, and robustness

Testing climate-driven diversification

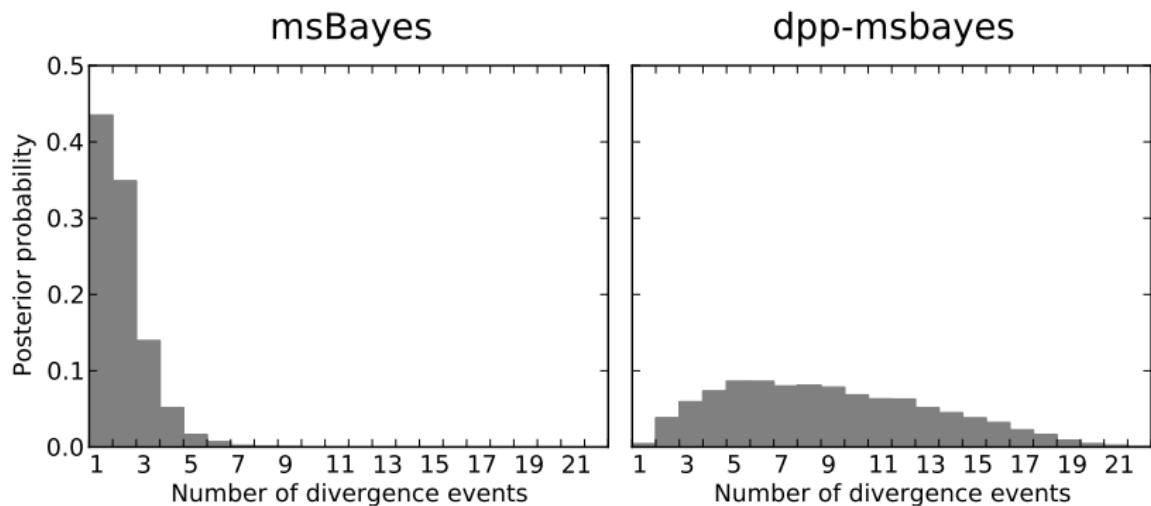
**Did repeated fragmentation of
islands during inter-glacial
rises in sea level promote
diversification?**



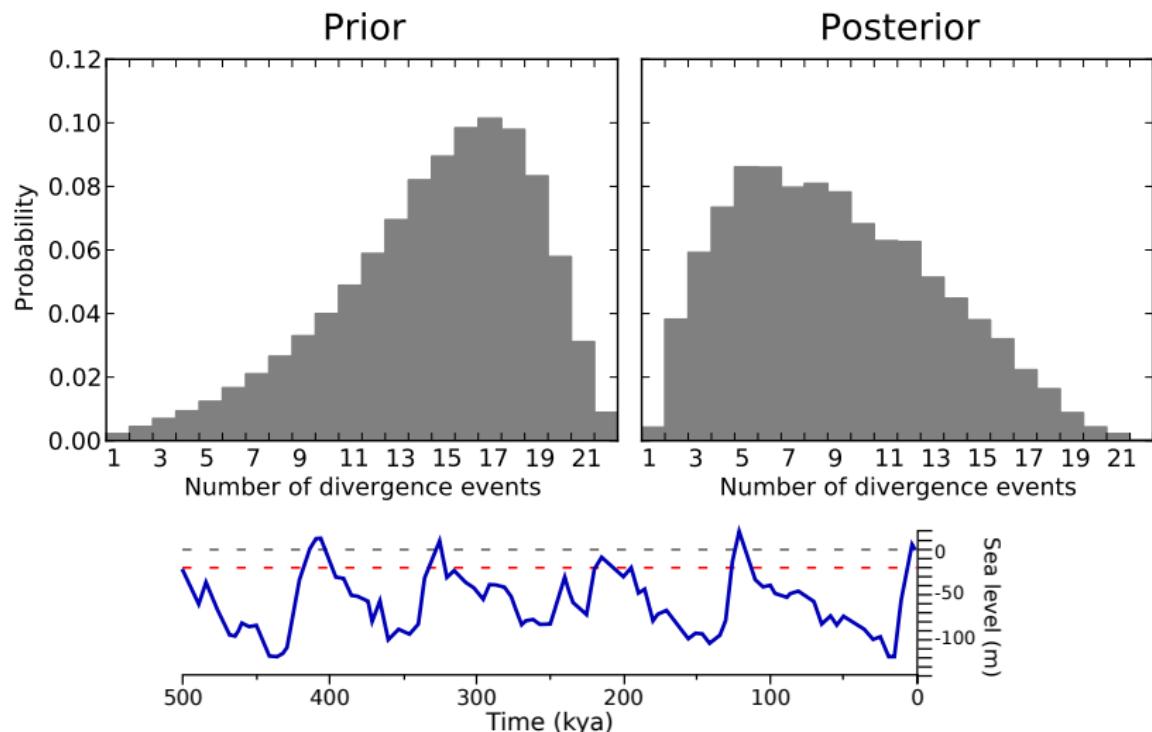
Species	n ₁	n ₂
Mammals		
<i>Crocidura beatus</i>	12	11
<i>Crocidura negrina-panayensis</i>	12	6
<i>Hipposideros obscurus</i>	19	9
<i>Hipposideros pygmaeus</i>	3	12
<i>Cynopterus brachyotis</i>	20	8
<i>Cynopterus brachyotis</i>	8	14
<i>Haplonycteris fischeri</i>	29	8
<i>Haplonycteris fischeri</i>	9	21
<i>Macroglossus minimus</i>	19	4
<i>Macroglossus minimus</i>	8	10
<i>Ptenochirus jagori</i>	4	7
<i>Ptenochirus jagori</i>	8	8
<i>Ptenochirus minor</i>	30	9
Squamates		
<i>Cyrtodactylus gubaot-sumuroi</i>	29	6
<i>Cyrtodactylus annulatus</i>	14	3
<i>Cyrtodactylus philippinicus</i>	6	14
<i>Gekko mindorensis</i>	8	11
<i>Insulasaurus arborens</i>	22	10
<i>Pinoyscincus jagori</i>	8	8
<i>Dendrelaphis marenae</i>	6	6
Anurans		
<i>Limnonectes leytensis</i>	4	2
<i>Limnonectes magnus</i>	2	3



dpp-msbayes: Philippine diversification



dpp-msbayes: Philippine diversification



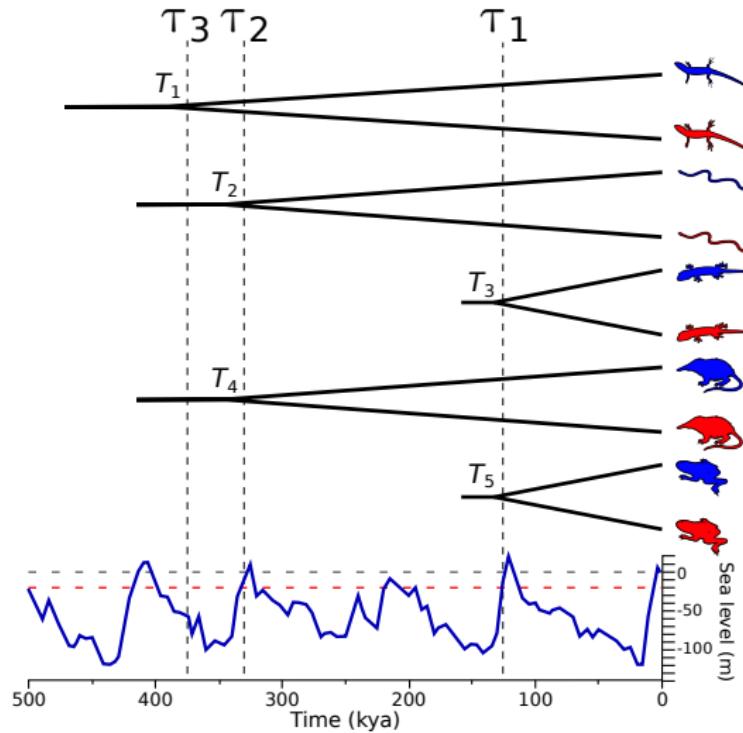
Conclusions

- ▶ Our new approximate-Bayesian method of phylogeographical model choice shows improved behavior
 - ▶ Improved accuracy, robustness, and power
 - ▶ More “honest” estimates regarding uncertainty

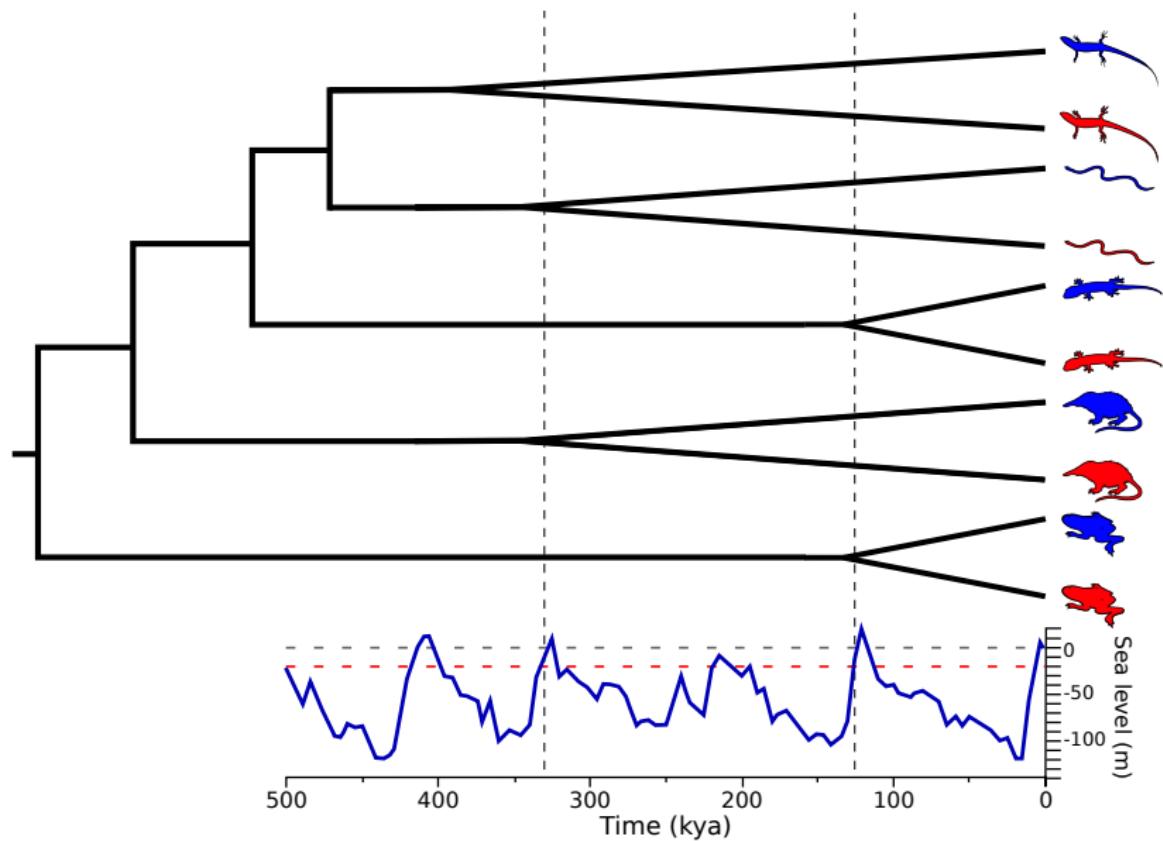
Conclusions

- ▶ Our new approximate-Bayesian method of phylogeographical model choice shows improved behavior
 - ▶ Improved accuracy, robustness, and power
 - ▶ More “honest” estimates regarding uncertainty
- ▶ Philippine climate-driven diversification model?
 - ▶ Results consistent with prediction of clustered divergences
 - ▶ Results suggest multiple co-divergences
 - ▶ However, there is a lot of uncertainty

Future directions: Full-Bayesian phylogenetic framework



Future directions: Full-Bayesian phylogenetic framework



Software

Everything is on GitHub...

- ▶ dpp-msbayes: <https://github.com/joaks1/dpp-msbayes>
- ▶ PyMsBayes: <https://github.com/joaks1/PyMsBayes>
- ▶ ABACUS: Approximate BAyesian C UtilitieS.
<https://github.com/joaks1/abacus>

Open Notebook Science

Everything is on GitHub...

- ▶ `msbayes-experiments`:
<https://github.com/joaks1/msbayes-experiments>
- ▶ `joaks1@gmail.com`

Acknowledgments

Ideas and feedback:

- ▶ KU Herpetology
- ▶ Holder Lab
- ▶ Melissa Callahan
- ▶ Mike Hickerson
- ▶ Laura Kubatko
- ▶ My committee

Computation:

- ▶ KU ITTC
- ▶ KU Computing Center
- ▶ iPlant

Funding:

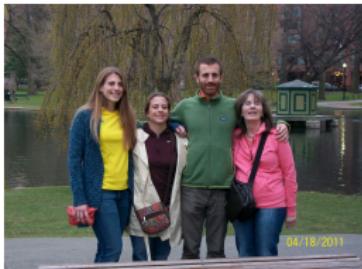
- ▶ NSF
- ▶ KU Grad Studies, EEB & BI
- ▶ SSB
- ▶ Sigma Xi

Photo credits:

- ▶ Rafe Brown, Cam Siler, & Jake Esselstyn
- ▶ FMNH Philippine Mammal Website:
 - ▶ D.S. Balete, M.R.M. Duya, & J. Holden

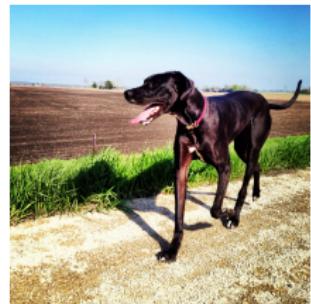
Acknowledgments

Friends & Family

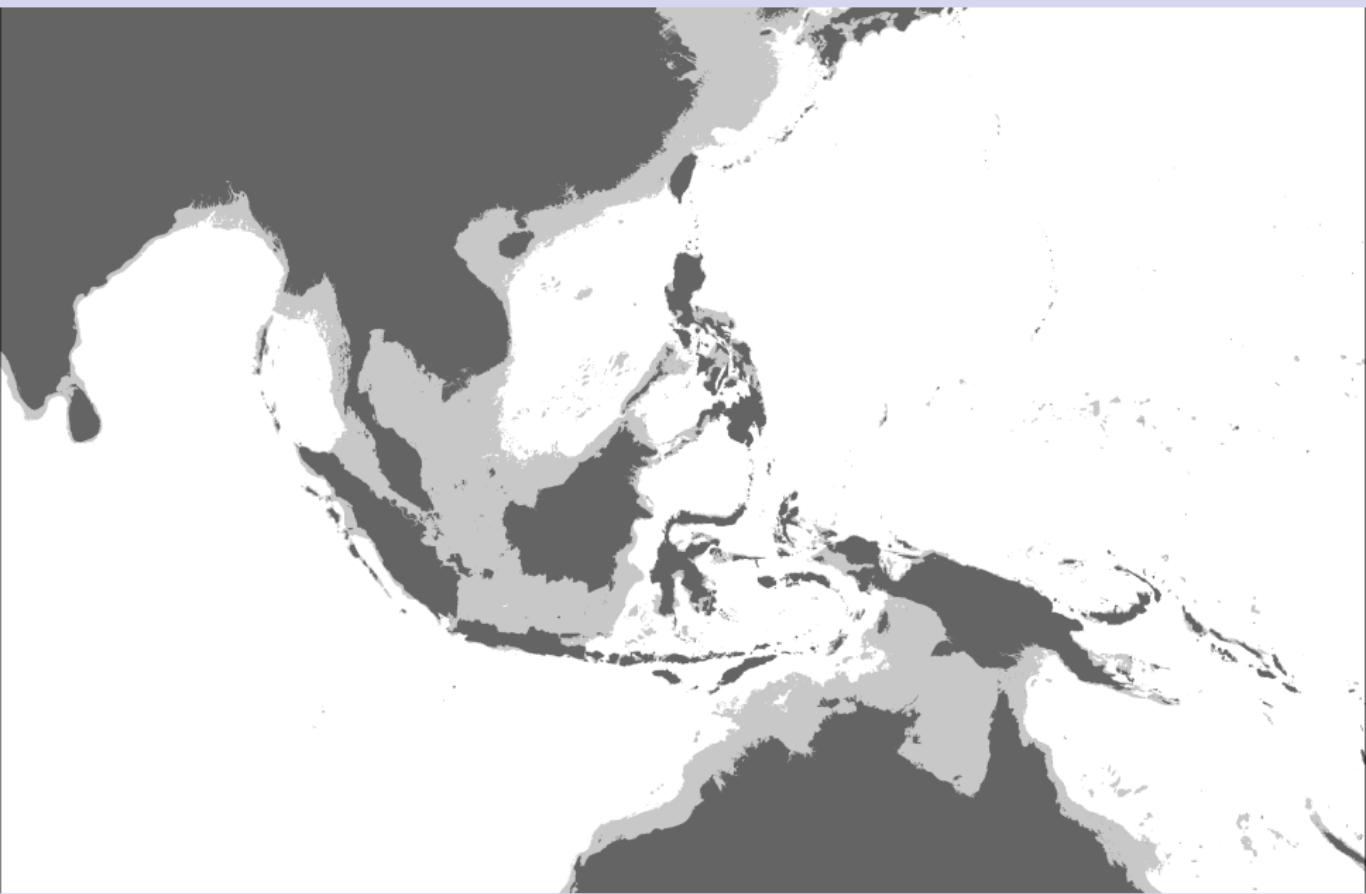


Acknowledgments

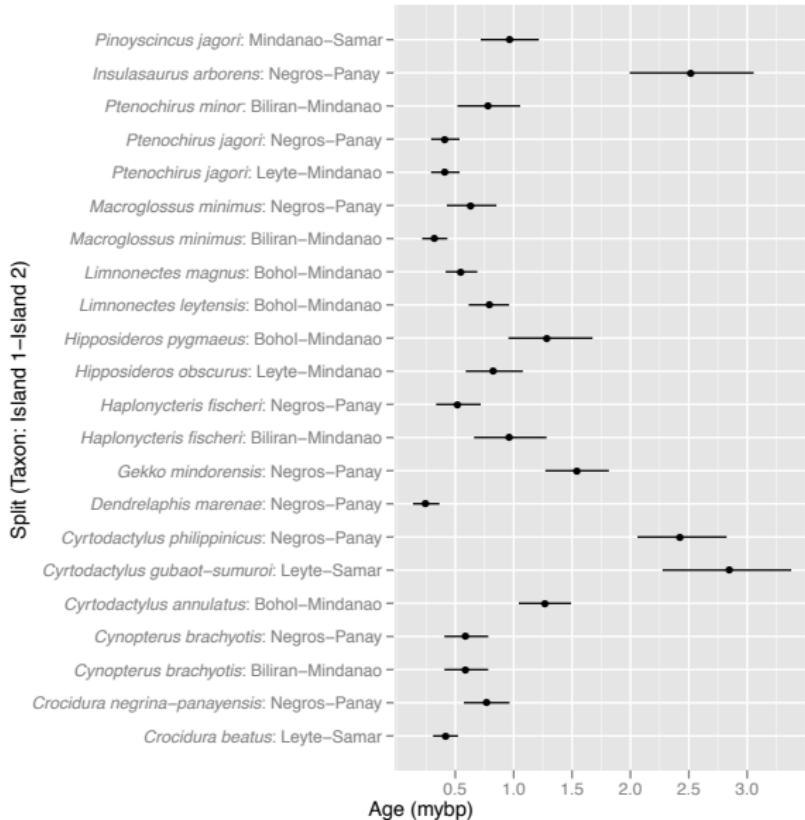
Friends & Family



Questions?

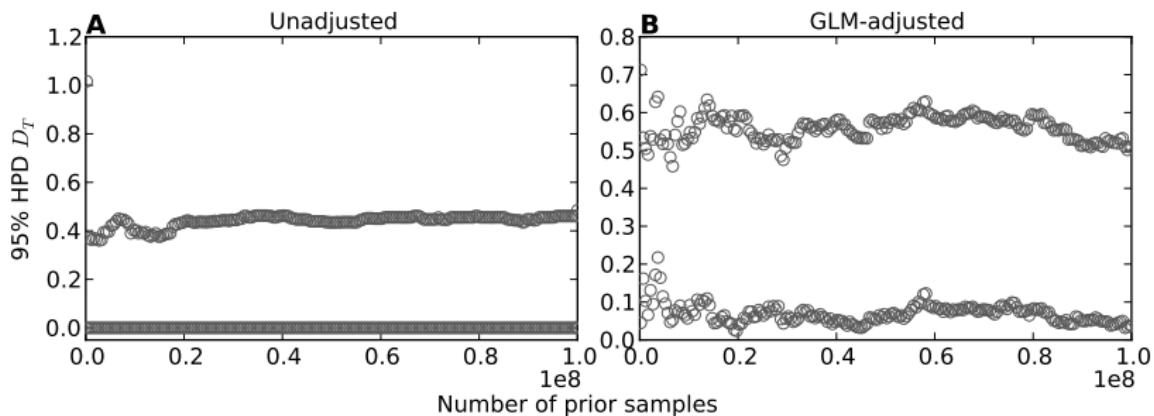


Gene tree divergences



Causes of bias: Insufficient sampling

- ▶ Models with more parameter space are less densely sampled
- ▶ Could explain bias toward small models in extreme cases
- ▶ **Predicts large variance in posterior estimates**
 - ▶ We explored empirical and simulation-based analyses with 2, 5, and 10 million prior samples, and estimates were very similar



Geological history

