

Workshop: Phylo with burst divergences

Jamie Oaks

Auburn University

phyletica.org



[joaks1](#) & [phyletica](#)



[@jamieoaks.bsky.social](#)



aub.ie/ssb-26-slides

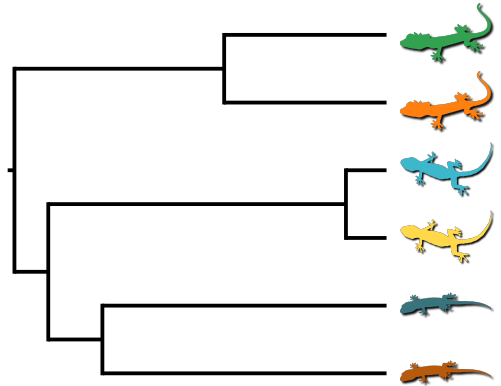


© 2007 Boris Kulikov boris-kulikov.blogspot.com

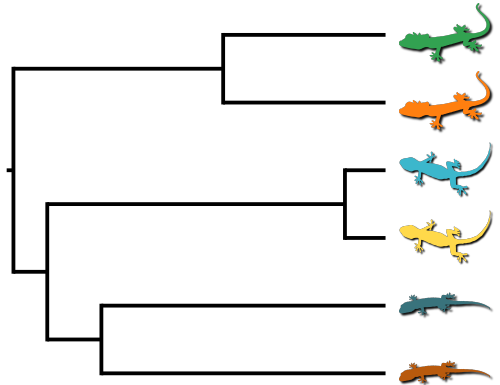
Icebreaker

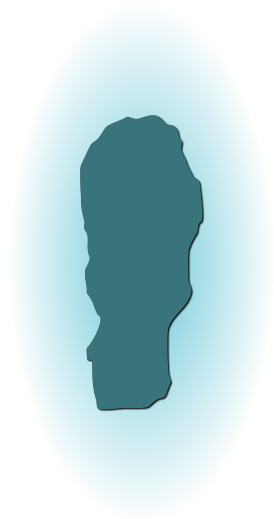
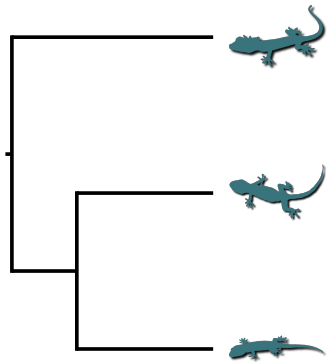
Please tell us:

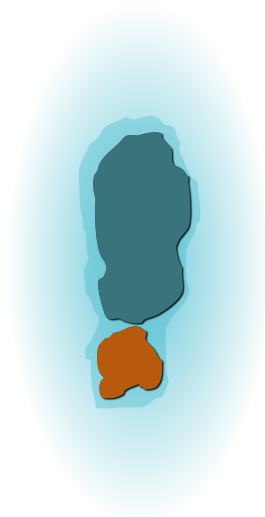
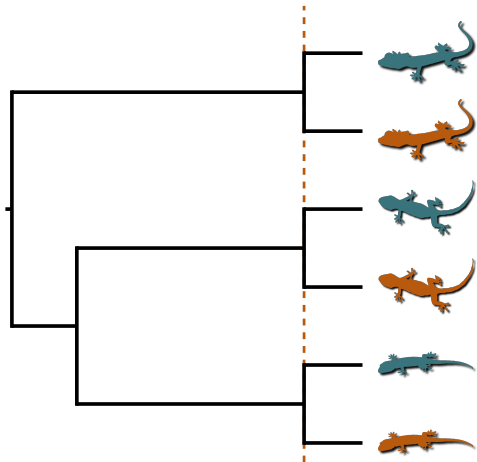
- ▶ Your name
- ▶ Where are you joining us from?
- ▶ Would you choose the power of *flight* or *invisibility*?
 - ▶ Stolen from [This American Life](#)

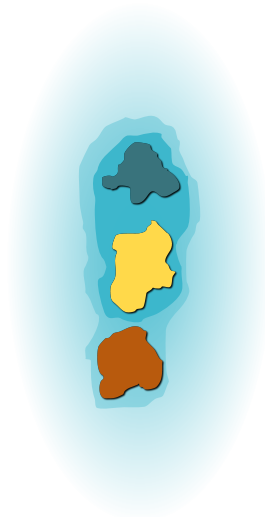
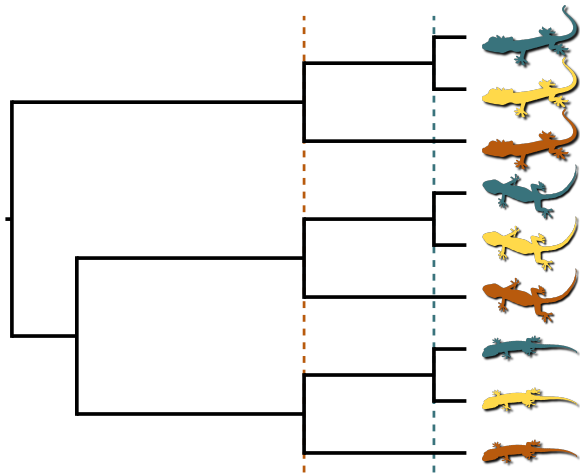


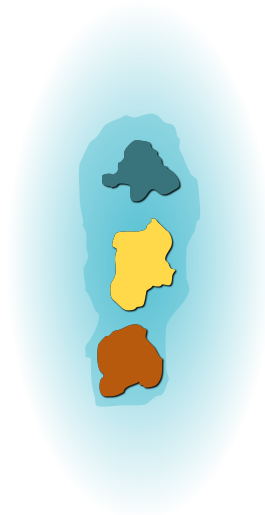
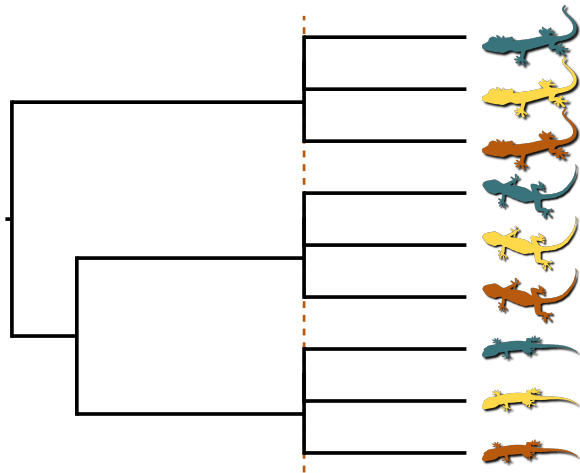
- **Assumption:** All processes of diversification affect each lineage independently

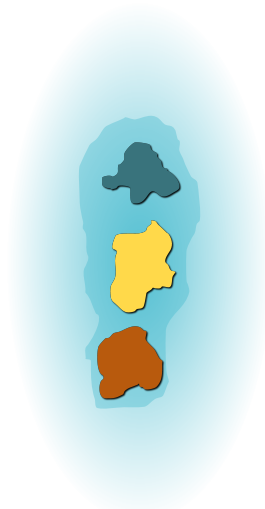
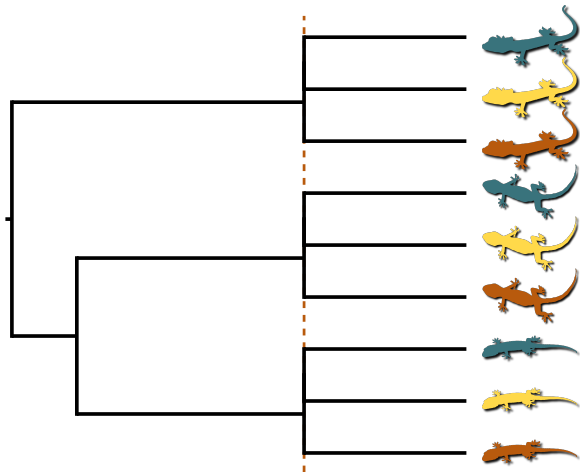










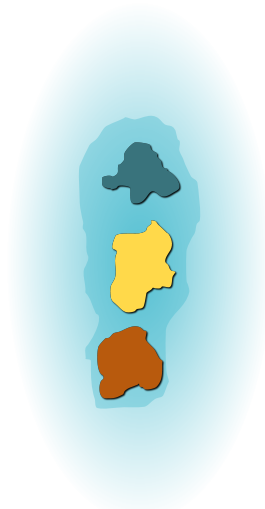
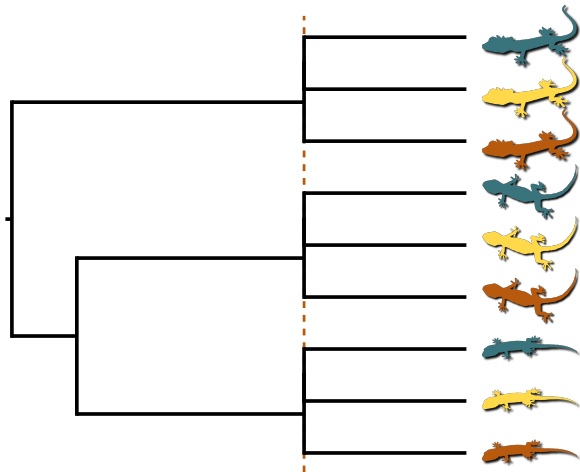


*These “**shared**” or “**burst**” divergences violate assumption of independent divergences*

Digressive Poll

I think polytomies/multifurcations are:

1. Useful model for representing phylo uncertainty, but not evolutionary patterns
2. Useful model of evolutionary patterns predicted by some diversification processes
3. Rubbish



These “**shared**” or “**burst**” divergences violate assumption of independent divergences

Biogeography

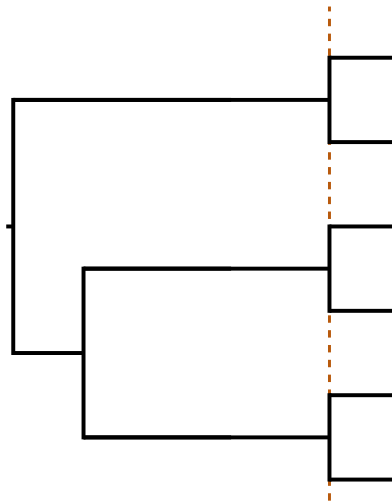
- ▶ Environmental changes that affect whole communities of species

Epidemiology

- ▶ Transmission at social gatherings

Genome evolution

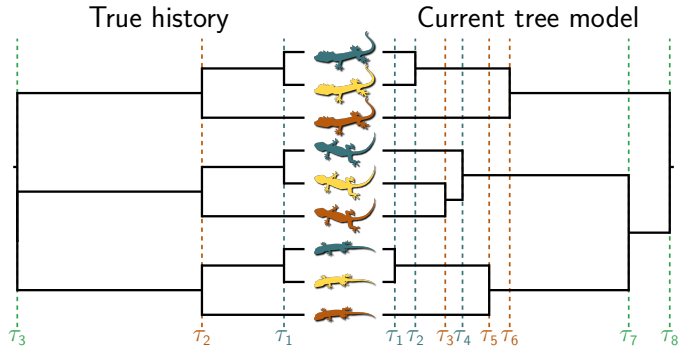
- ▶ Duplication of a chromosome segment harboring a gene family



Why account for shared divergences?

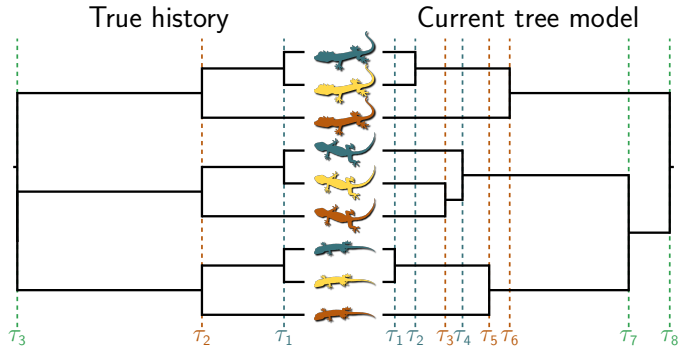
Why account for shared divergences?

1. Improve inference



Why account for shared divergences?

1. Improve inference
2. **Provide a framework for studying processes of co-diversification**



Biogeography

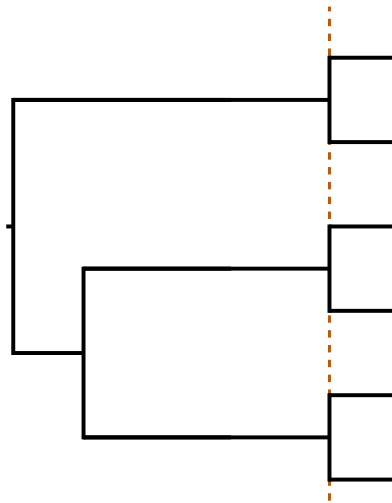
- ▶ Environmental changes that affect whole communities of species

Epidemiology

- ▶ Transmission at social gatherings

Genome evolution

- ▶ Duplication of a chromosome segment harboring a gene family



Biogeography

- ▶ Environmental changes that affect whole communities of species

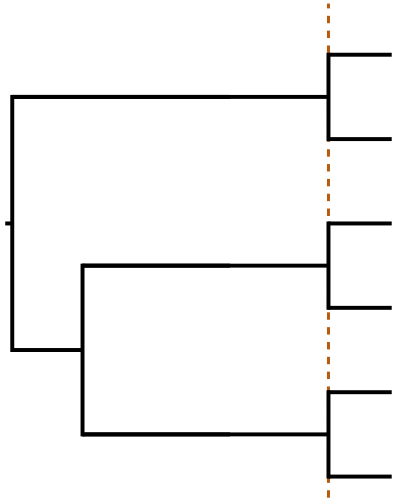
Epidemiology

- ▶ Transmission at social gatherings

Genome evolution

- ▶ Duplication of a chromosome segment harboring a gene family

These processes are interesting!



Our approach:

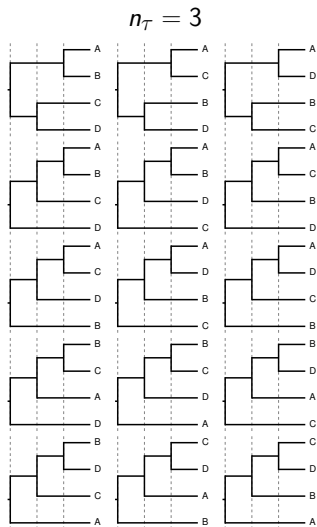
Generalize the space of trees considered during phylogenetic inference



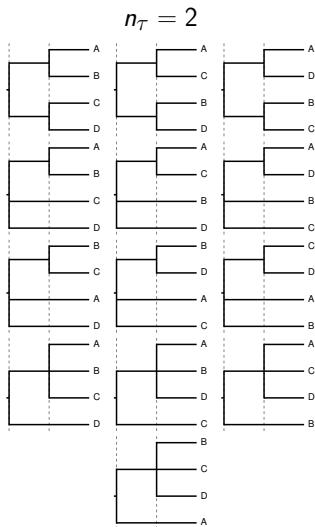
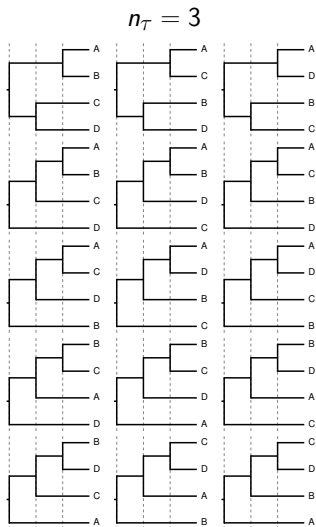
Dr. Perry Wood, Jr.

Generalizing tree space

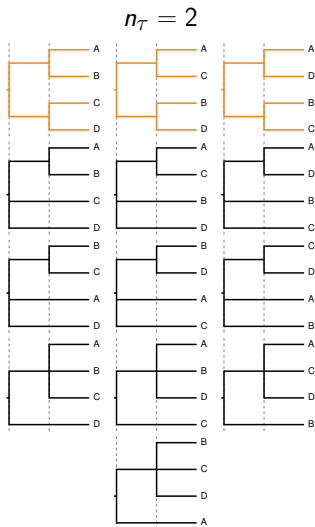
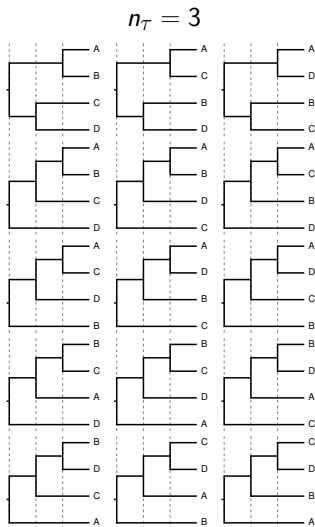
Generalizing tree space



Generalizing tree space



Generalizing tree space

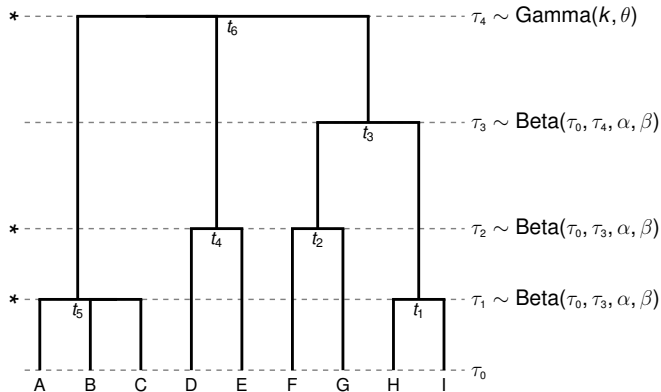


$n_T = 1$

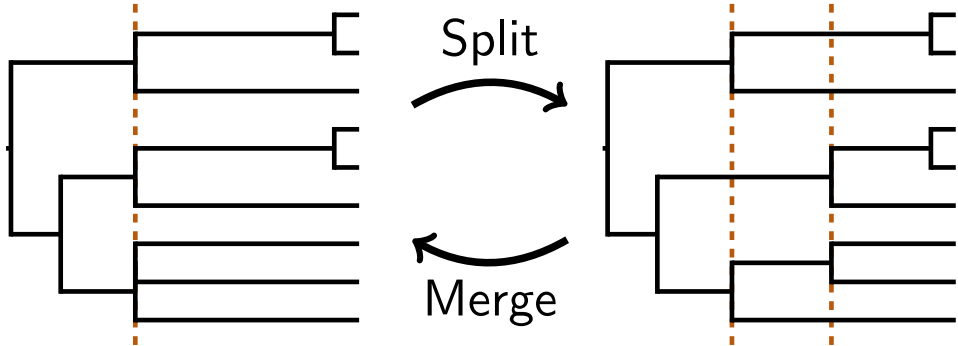


Generalized tree distribution

- ▶ All topologies equally probable
- ▶ Parametric distribution on age of root
- ▶ Beta distributions on other div times

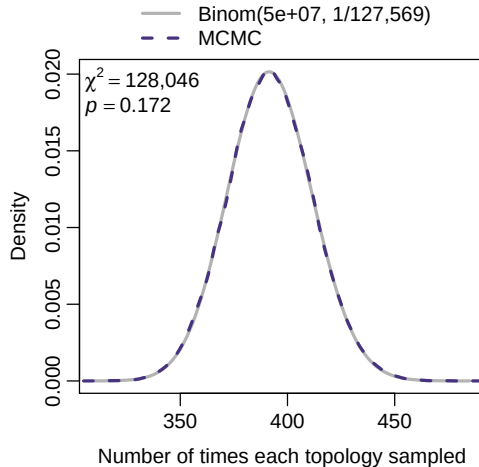


Inferring trees with shared divergences



Reversible-jump MCMC

Validating rjMCMC with 7-leaf tree



The rjMCMC algorithms sample the expected generalized tree distribution

PhycoEval

Phylogenetic coevality

J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Ecoevolity

Estimating evolutionary coevality

J. R. Oaks (2019). *Systematic Biology* 68: 371–395

- ▶ **Tree model**

- ▶ rjMCMC sampling of generalized tree distribution

PhycoEval

Phylogenetic coevality

J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Ecoevolity

Estimating evolutionary coevality

J. R. Oaks (2019). *Systematic Biology* 68: 371–395

- ▶ **Tree model**

- ▶ rjMCMC sampling of generalized tree distribution

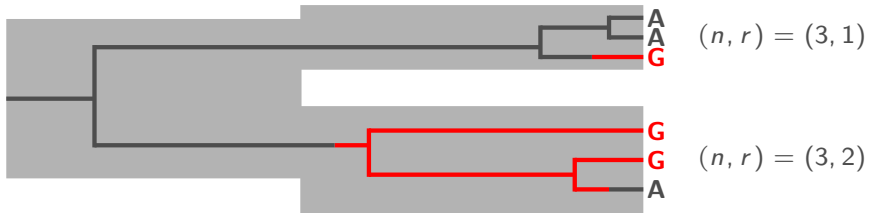
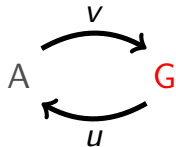
- ▶ **Likelihood model**

- ▶ CTMC model of characters evolving along genealogies
 - ▶ Infer species trees by analytically integrate over genealogies¹

¹ D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

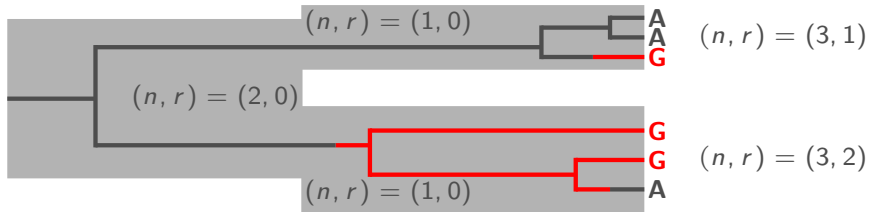
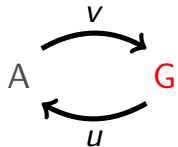
- ▶ **Tree model**
 - ▶ rjMCMC sampling of generalized tree distribution
- ▶ **Likelihood model**
 - ▶ CTMC model of characters evolving along genealogies
 - ▶ Infer species trees by analytically integrate over genealogies¹
- ▶ *Goal: Co-estimation of phylogeny and shared divergences from genomic data*

¹ D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



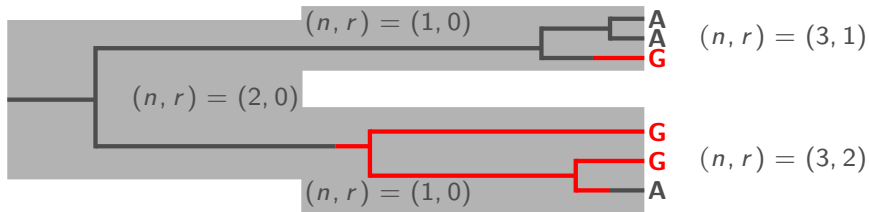
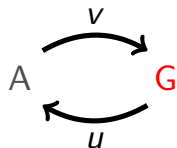
¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

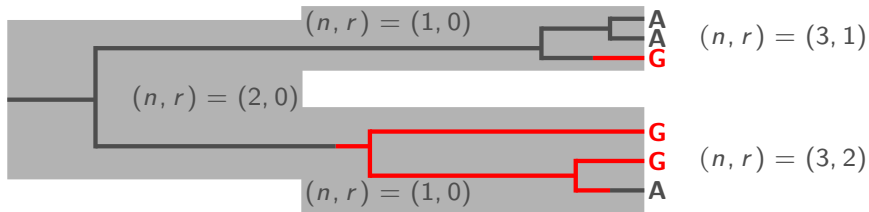
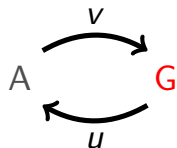


$Q =$

	$(1, 0)$	$(1, 1)$	$(2, 0)$	$(2, 1)$	$\dots$	(n, n)
$(1, 0)$	.	.	.	.		.
$(1, 1)$	.	.	.	.		.
$(2, 0)$	.	.	.	.		.
$(2, 1)$	.	.	.	.		.
$\vdots$						
(n, n)	.	.	.	.		.

¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

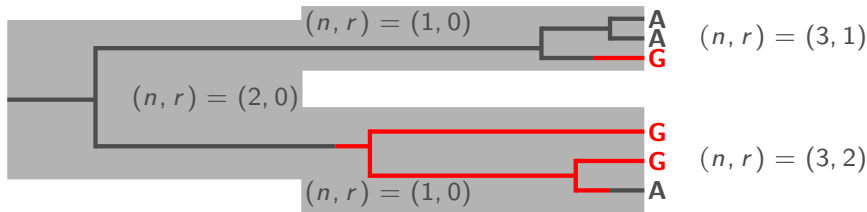
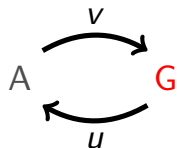


$$Q = \begin{array}{c|cccccc} & (1,0) & (1,1) & (2,0) & (2,1) & \cdots & (n,n) \\ \hline (1,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (1,1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \vdots & & & & & & \\ (n,n) & \cdot & \cdot & \cdot & \cdot & & \cdot \end{array}$$

$$\begin{aligned} Q_{(n,r);(n,r-1)} &= (n-r+1)v, && \text{mutation,} \\ Q_{(n,r);(n,r+1)} &= (r+1)u, && \text{mutation,} \\ Q_{(n,r);(n-1,r)} &= \frac{(n-1-r)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n-1,r-1)} &= \frac{(r-1)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n,r)} &= -\frac{(n-1)n}{2N_e(u+v)} - (n-r)v - ru. \end{aligned}$$

¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



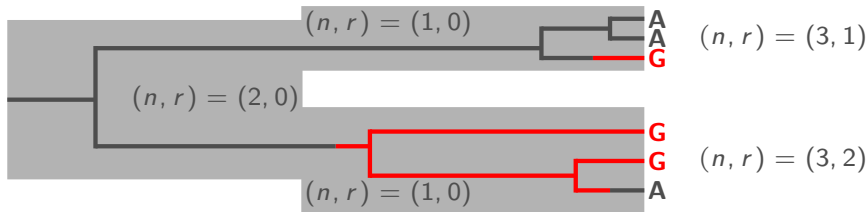
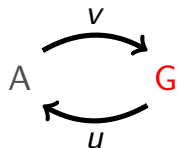
$$Q = \begin{array}{c|cccccc} & (1, 0) & (1, 1) & (2, 0) & (2, 1) & \cdots & (n, n) \\ \hline (1, 0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (1, 1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2, 0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2, 1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ \vdots & & & & & & \\ (n, n) & \cdot & \cdot & \cdot & \cdot & & \cdot \end{array}$$

$$\begin{aligned} Q_{(n,r);(n,r-1)} &= (n-r+1)v, && \text{mutation,} \\ Q_{(n,r);(n,r+1)} &= (r+1)u, && \text{mutation,} \\ Q_{(n,r);(n-1,r)} &= \frac{(n-1-r)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n-1,r-1)} &= \frac{(r-1)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n,r)} &= -\frac{(n-1)n}{2N_e(u+v)} - (n-r)v - ru. \end{aligned}$$

► e^{Qt} to keep track of all conditional probabilities along each branch (Carathéodory-Fejér method¹)

¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



	(1, 0)	(1, 1)	(2, 0)	(2, 1)	...	(n, n)	
$Q =$	(1, 0)	.	.	.	.	.	$Q_{(n,r);(n,r-1)} = (n-r+1)v$, mutation,
	(1, 1)	.	.	.	.	.	$Q_{(n,r);(n,r+1)} = (r+1)u$, mutation,
	(2, 0)	.	.	.	.	.	$Q_{(n,r);(n-1,r)} = \frac{(n-1-r)n}{2N_e(u+v)}$, coalescence,
	(2, 1)	.	.	.	.	.	$Q_{(n,r);(n-1,r-1)} = \frac{(r-1)n}{2N_e(u+v)}$, coalescence,
	$\vdots$						$Q_{(n,r);(n,r)} = -\frac{(n-1)n}{2N_e(u+v)} - (n-r)v - ru$.
	(n, n)	.	.	.	.	.	

- e^{Qt} to keep track of all conditional probabilities along each branch (Carathéodory-Fejér method¹)
- At root, get likelihood of population tree integrated over all possible gene trees and mutational histories²

¹ T. Schmelzer et al. (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

² D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

- ▶ **Tree model**

- ▶ rjMCMC sampling of generalized tree distribution

- ▶ **Likelihood model**

- ▶ CTMC model of characters evolving along genealogies
 - ▶ Infer species trees by analytically integrate over genealogies¹

- ▶ *Goal: Co-estimation of phylogeny and shared divergences from genomic data*

¹ D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

- ▶ **Tree model**

- ▶ rjMCMC sampling of generalized tree distribution

- ▶ **Likelihood model**

- ▶ CTMC model of characters evolving along genealogies
 - ▶ Infer species trees by analytically integrate over genealogies¹

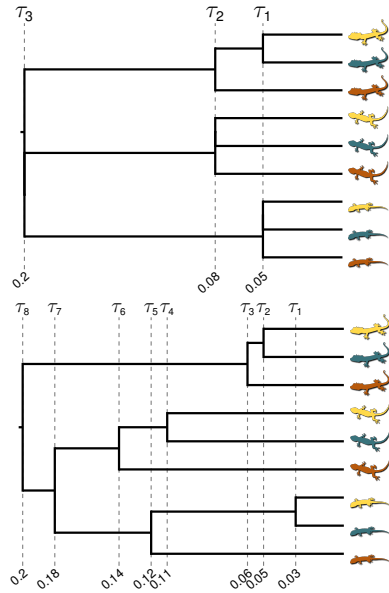
- ▶ *Goal: Co-estimation of phylogeny and shared divergences from genomic data*

- ▶ *Does it work?*

¹ D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932

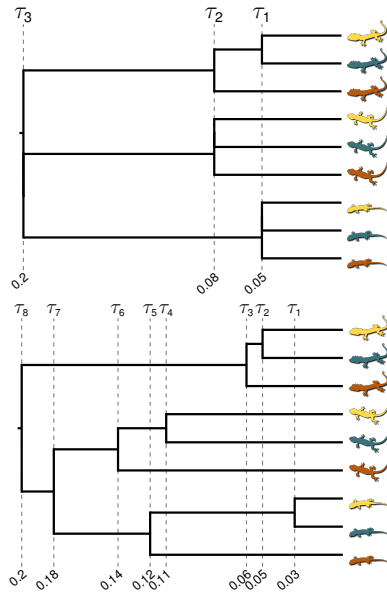
Methods: Simulations

- ▶ Simulated 100 data sets with 50,000 base pairs



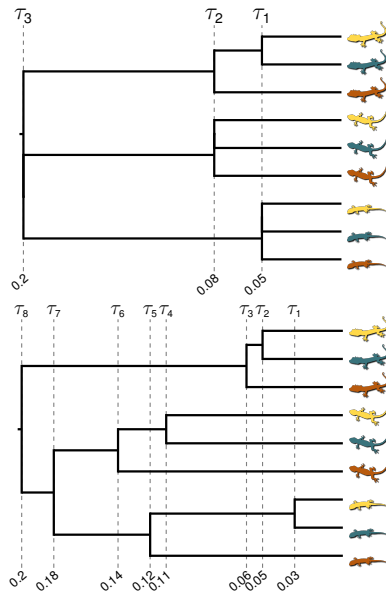
Methods: Simulations

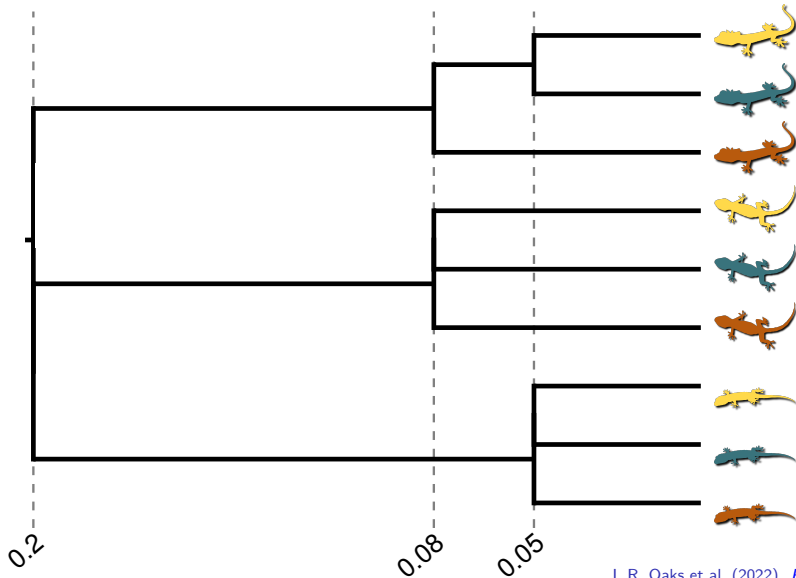
- ▶ Simulated 100 data sets with 50,000 base pairs
- ▶ Analyzed each data set with:
 - ▶ M_G = Generalized tree model
 - ▶ M_{IB} = Independent-bifurcating tree model

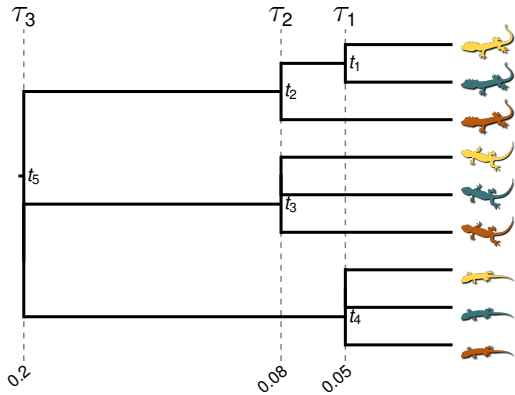


Methods: Simulations

- ▶ Simulated 100 data sets with 50,000 base pairs
- ▶ Analyzed each data set with:
 - ▶ M_G = Generalized tree model
 - ▶ M_{IB} = Independent-bifurcating tree model
- ▶ Simulated 100 data sets where topology and div times randomly drawn from M_G and M_{IB}

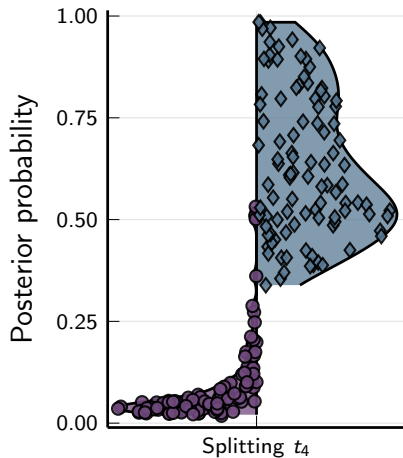
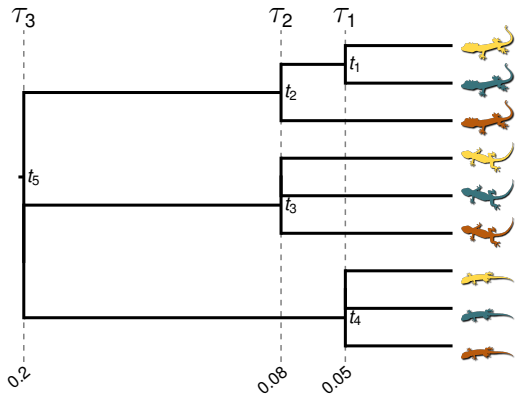






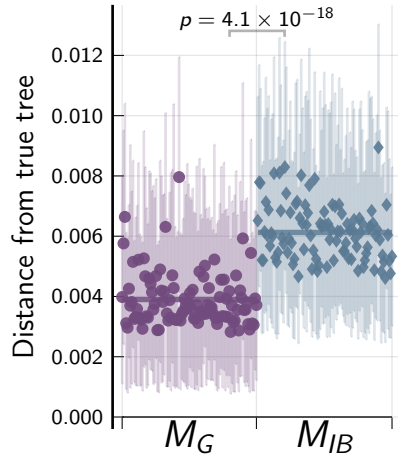
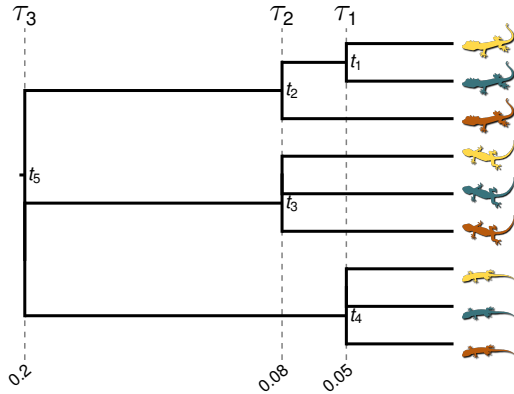
● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model



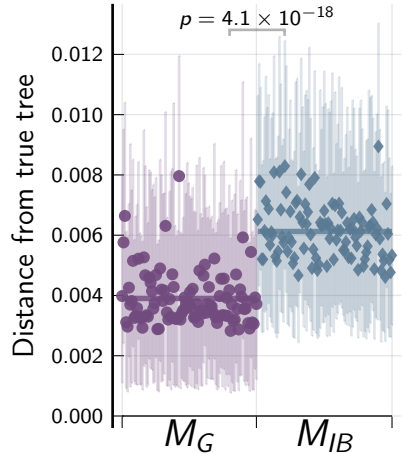
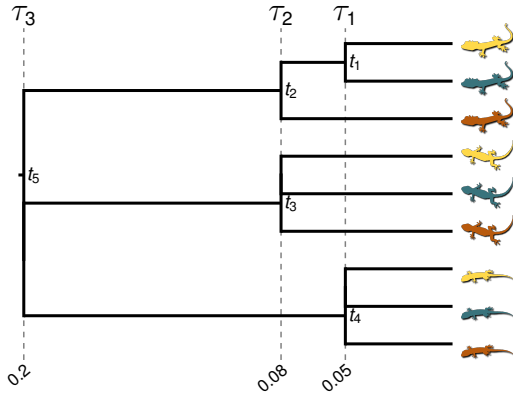
● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model



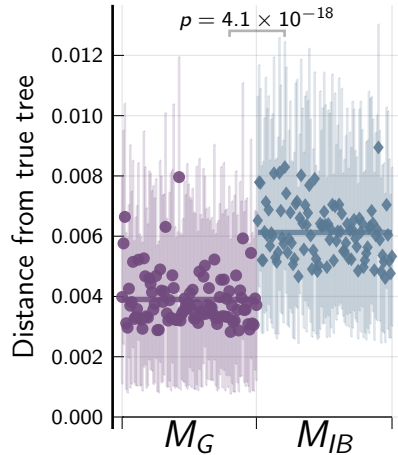
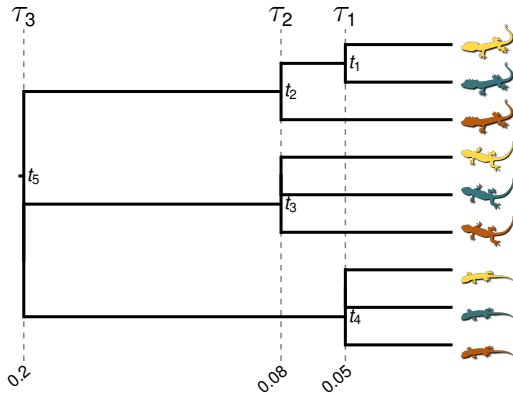
● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model



● M_G = Generalized model

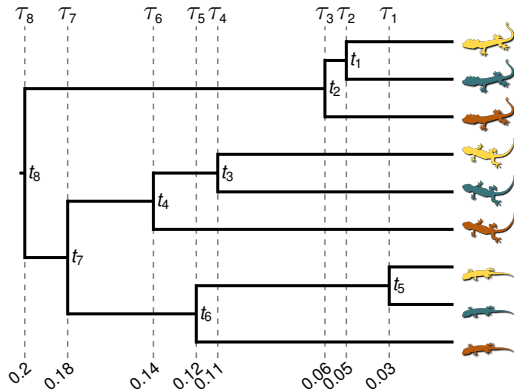
◆ M_{IB} = Independent-bifurcating model



M_G significantly better at inferring trees with shared divergences

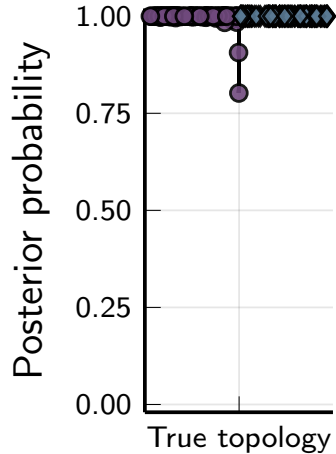
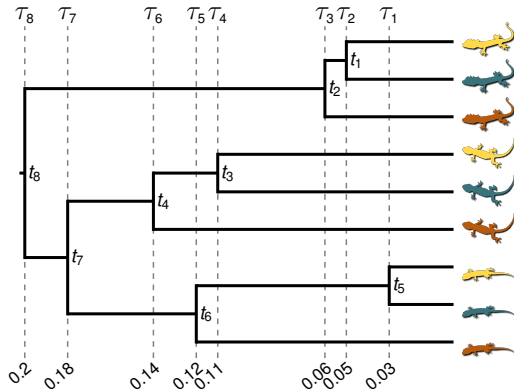
● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model



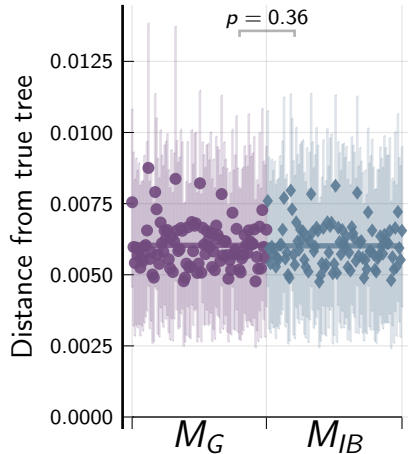
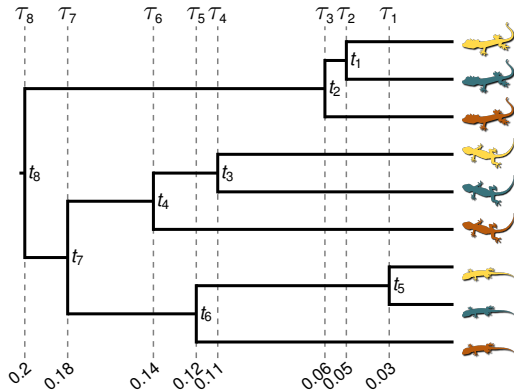
● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model

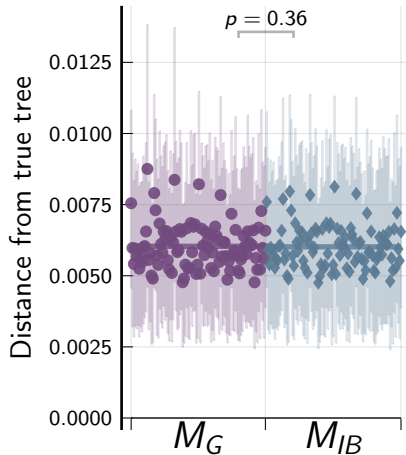
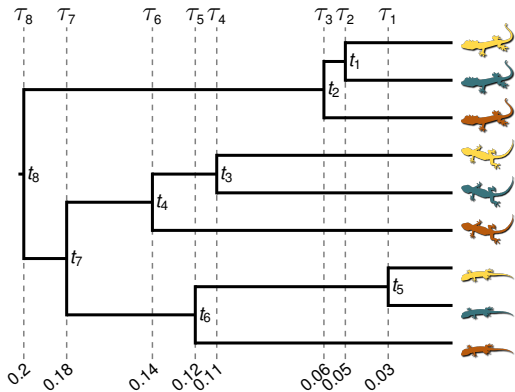


● M_G = Generalized model

◆ M_{IB} = Independent-bifurcating model

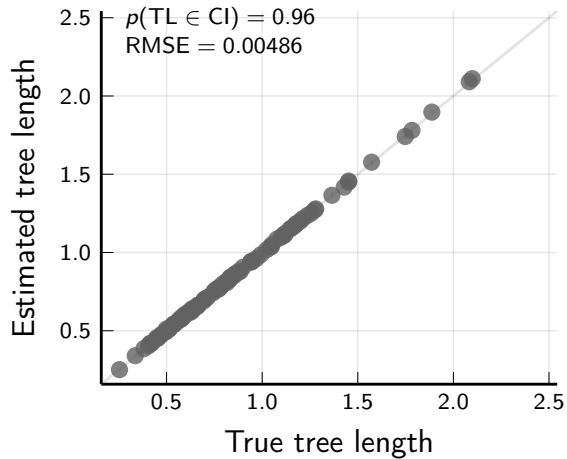


● M_G = Generalized model ◆ M_{IB} = Independent-bifurcating model

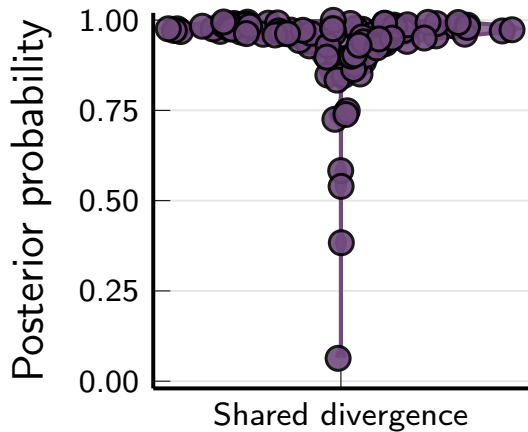
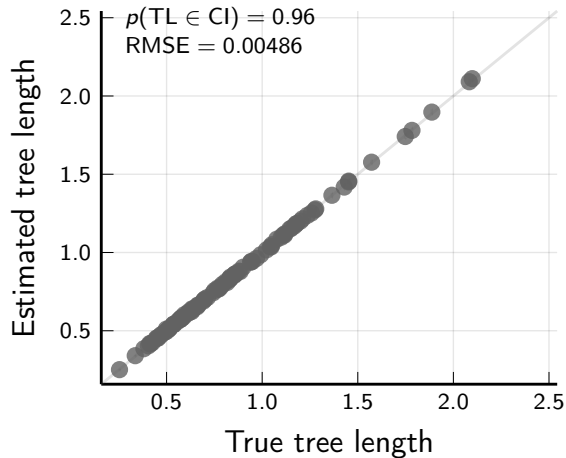


M_G performs as well as true model when divergences are independent

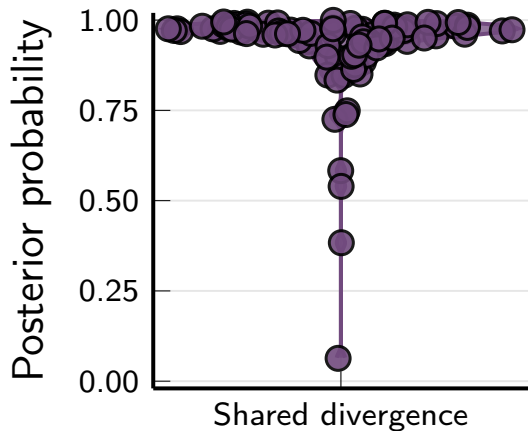
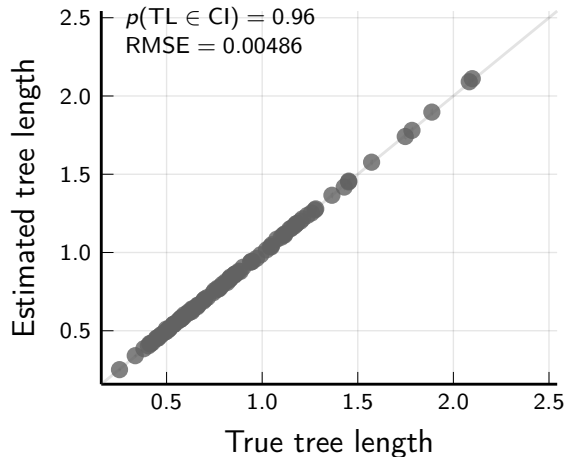
Results: random M_G trees



Results: random M_G trees

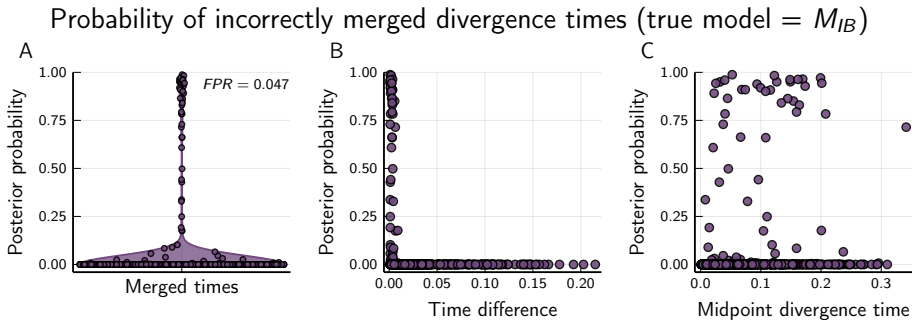


Results: random M_G trees

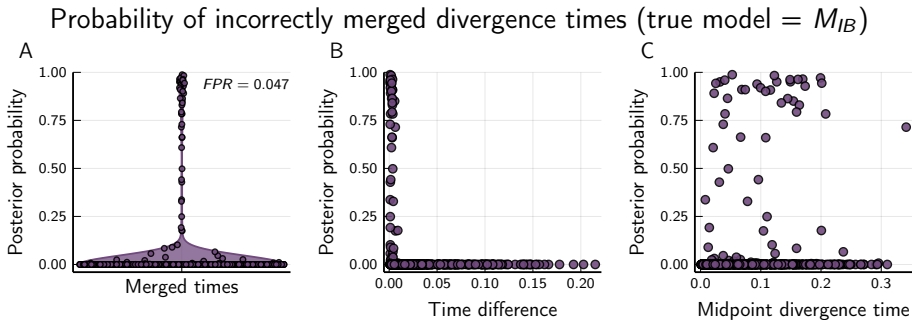


M_G performs well with data simulated on random trees with shared divergences

Results: random M_{IB} trees

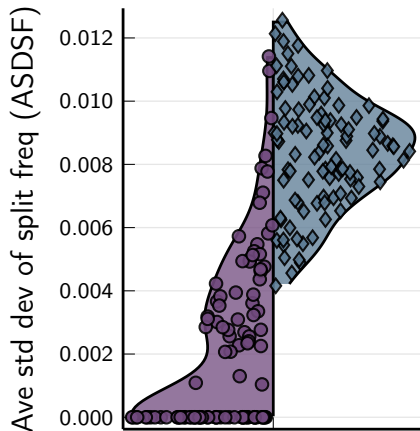


Results: random M_{IB} trees

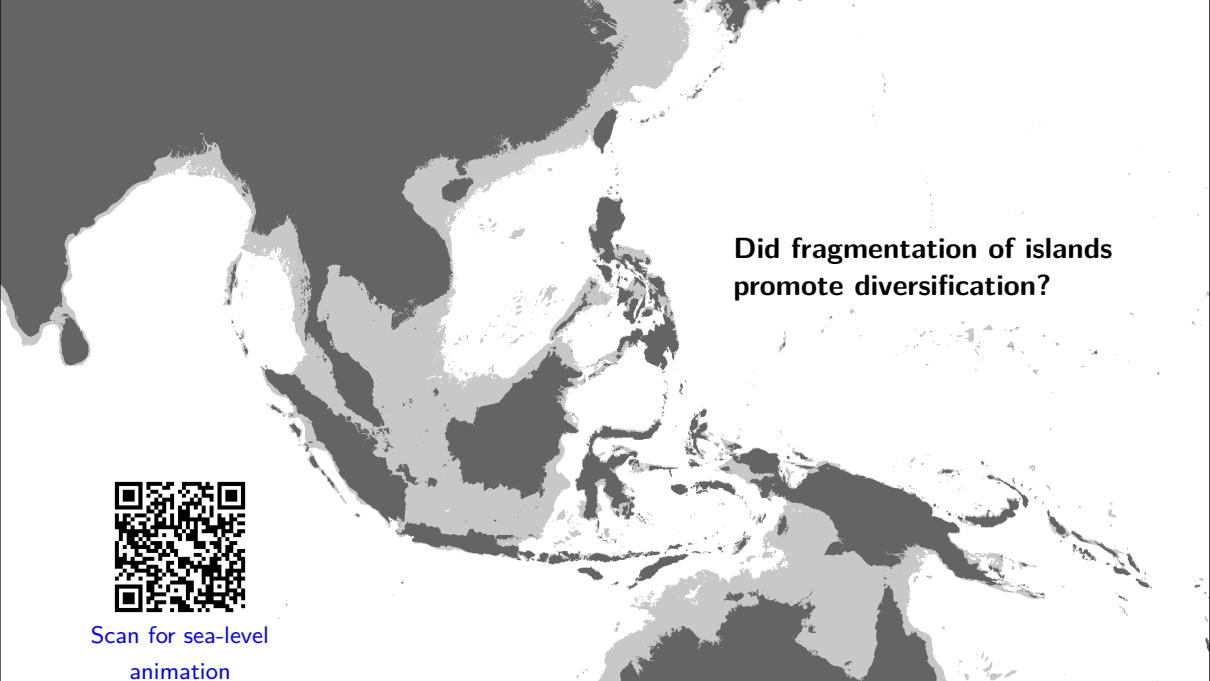


M_G has low false positive rate

● M_G = Generalized model ◆ M_{IB} = Independent-bifurcating model



Generalizing tree space improves MCMC convergence and mixing

A map of Southeast Asia, including the Malay Peninsula, Sumatra, Java, and the Indonesian archipelago. The map is color-coded to show different land levels: dark green for the highest land, light green for intermediate levels, and yellow for the lowest land. The yellow areas represent the submerged portions of the landmass, illustrating how rising sea levels would fragment the region into numerous smaller islands. The text "Did fragmentation of islands promote diversification?" is overlaid on the right side of the map.

**Did fragmentation of islands
promote diversification?**

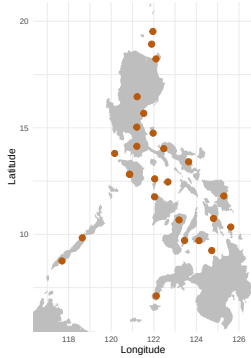


Scan for sea-level
animation

Cyrtodactylus



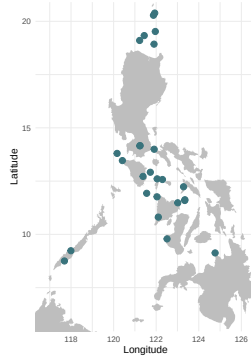
©Rafe M. Brown



Gekko



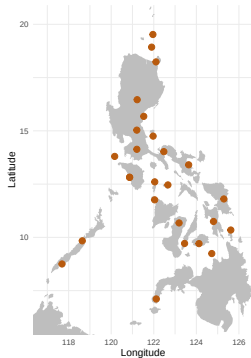
©Rafe M. Brown



Cyrtodactylus



©Rafe M. Brown

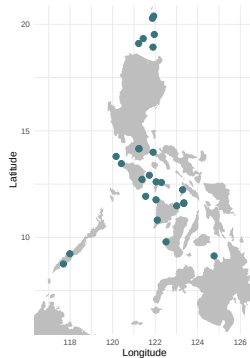


1702 loci
155,887 sites

Gekko

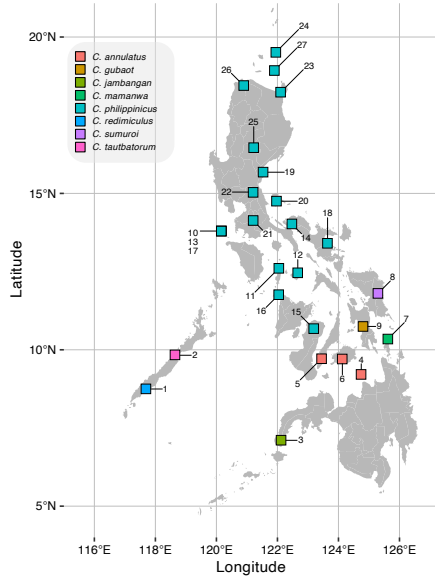
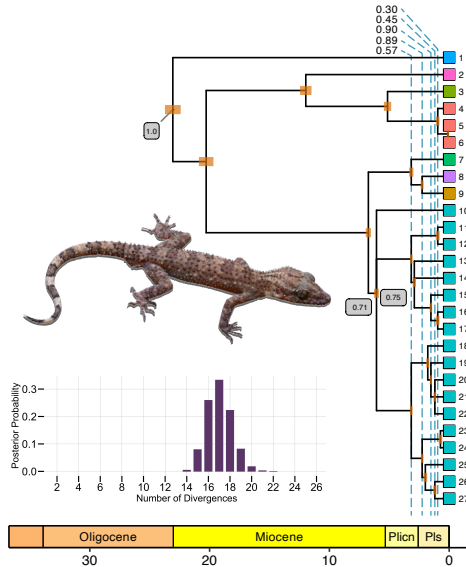


©Rafe M. Brown



1033 loci
94,813 sites

Cyrtodactylus



Take-home points

- ▶ We can accurately infer phylogenies with shared divergences with moderately sized data sets

Take-home points

- ▶ We can accurately infer phylogenies with shared divergences with moderately sized data sets
- ▶ Generalizing tree space avoids spurious support and improves MCMC mixing

Take-home points

- ▶ We can accurately infer phylogenies with shared divergences with moderately sized data sets
- ▶ Generalizing tree space avoids spurious support and improves MCMC mixing
- ▶ Among Philippine gekkonids, we found support for shared divergences predicted by sea-level changes

Acknowledgments

Ideas and feedback:

- ▶ [Phyletica Lab](#) (the Phyleticians)
- ▶ Perry Wood, Jr.
- ▶ AUMNH
- ▶ Mark Holder
- ▶ Rafe Brown
- ▶ Cam Siler
- ▶ Lee Grismer
- ▶ Ammon Thompson
- ▶ Michael Landis
- ▶ Sebastian Höhna

Computation:

- ▶ Alabama Supercomputer Authority
- ▶ Auburn University Hopper Cluster

Funding:



AUBURN
UNIVERSITY

Photo credits:

- ▶ Rafe Brown
- ▶ Perry Wood, Jr.
- ▶ [PhyloPic](#)

Questions?

joaks@auburn.edu



phyletica.org

[joaks1](#) & [phyletica](#)



[@jamieoaks.bsky.social](#)

Slides:



aub.ie/ssb-26-slides



© 2007 Boris Kulikov boris-kulikov.blogspot.com

Paper



aub.ie/ssb-26-paper

Installation



aub.ie/ssb-26-install

Tutorial



aub.ie/ssb-26-tutorial