Generalizing phylogenetics to infer patterns predicted by processes of diversification

Jamie Oaks
Auburn University
phyletica.org
@jamoaks

Scan for slides:

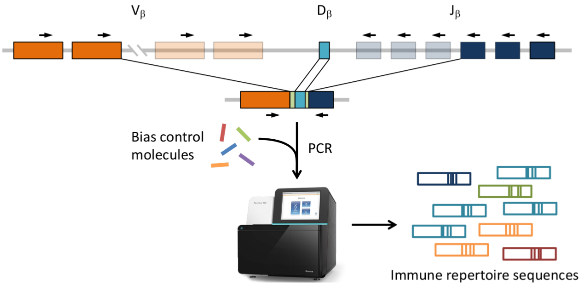
phyletica.org/slides/duke-cbb.pdf

© 2007 Boris Kulikov boris-kulikov.blogspot.com

AUBURN UNIVERSITY

Adaptive biotechnologies™

© 2007 Boris Kulikov boris-kulikov.blogspot.com



AUBURN UNIVERSITY

© 2007 Boris Kulikov boris-kulikov.blogspot.com

AUBURN
UNIVERSITY

© 2007 Boris Kulikov boris-kulikov.blogspot.com

Generalizing phylogenetics to infer patterns of shared evolutionary events

The Phyleticians

Postdocs

- Perry Wood, Jr
- *Brian Folt*
- *Jesse Grismer*

Graduate students

- Tashitso Anamza
- Matt Buehler
- Kerry Cobb
- Kyle David
- Saman Jahangiri
- Randy Klabacka
- Morgan Muell
- Tanner Myers
- Claire Tracy
- *Breanna Sipley*
- *Aundrea Westfall*

Undergraduate students

- Laura Lewis
- Mary Wells
- Hailey Whitaker
- Noah Yawn
- *Charlotte Benedict*
- *Eric Carbo*
- *Ryan Cook*
- *Andrew DeSana*
- *Miles Horne*
- *Jacob Landrum*
- *Nadia L'Bahy*
- *Jorge Lopez-Perez*
- *Holden Smith*
- *Virginia White*
- *Kayla Wilson*

- Phylogenetics is rapidly becoming the statistical foundation of biology

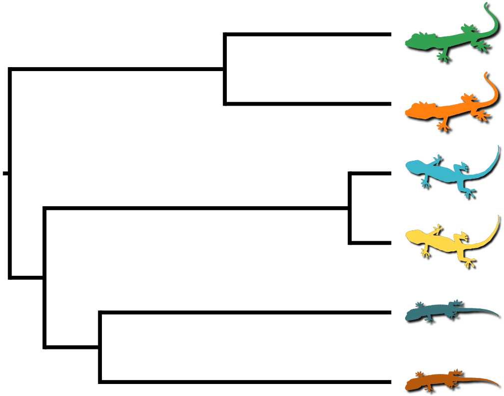
© 2007 Boris Kulikov boris-kulikov.blogspot.com

- Phylogenetics is rapidly becoming the statistical foundation of biology
- “Big data” present exciting possibilities and challenges

© 2007 Boris Kulikov boris-kulikov.blogspot.com

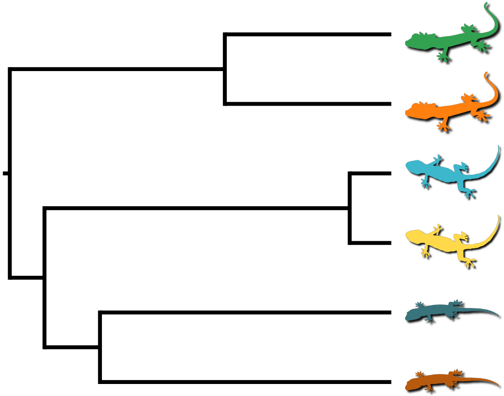
- Phylogenetics is rapidly becoming the statistical foundation of biology
- "Big data" present exciting possibilities and challenges
- Many opportunities to develop new ways to study biology in light of phylogeny

© 2007 Boris Kulikov boris-kulikov.blogspot.com

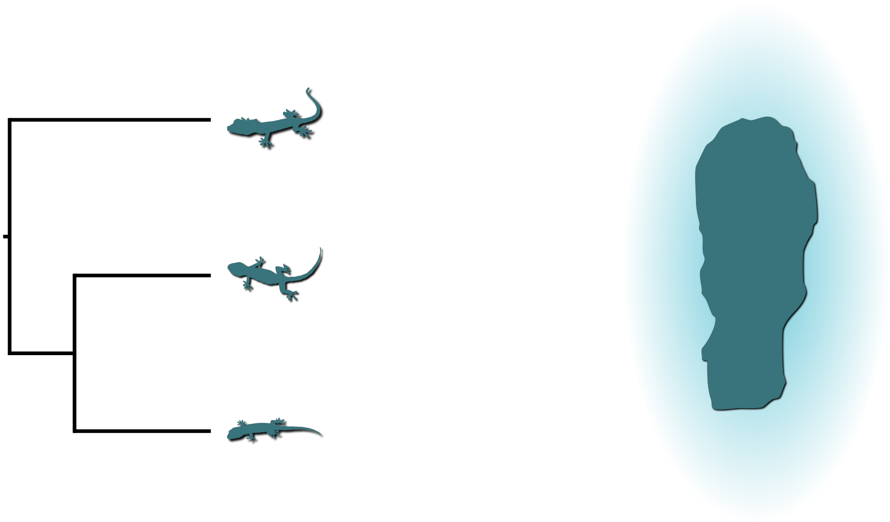


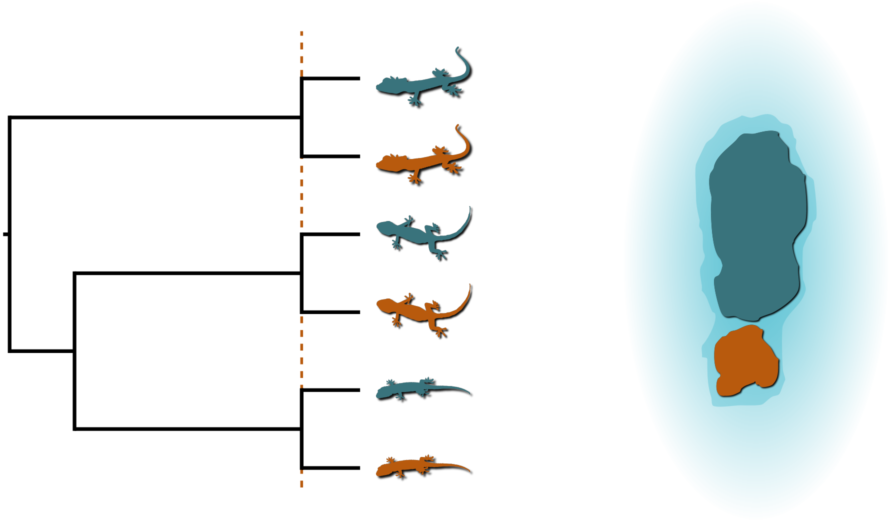
phylopic.org CC0

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



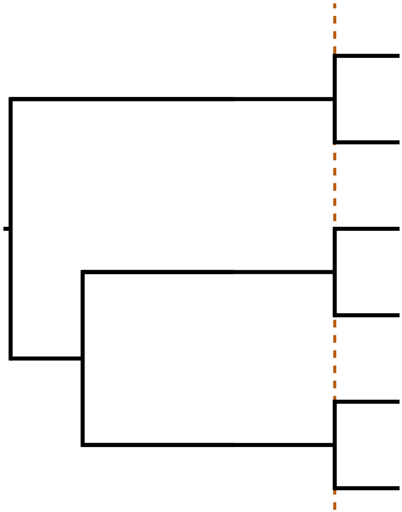
phylopic.org CC0





Biogeography

- Environmental changes that affect whole communities of species

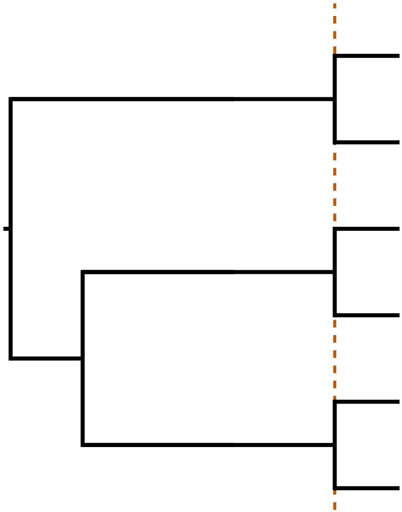


Biogeography

- Environmental changes that affect whole communities of species

Genome evolution

- Duplication of a chromosome segment harboring gene families



Biogeography

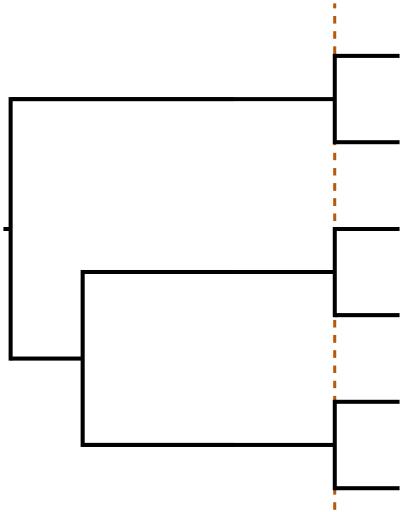
- Environmental changes that affect whole communities of species

Genome evolution

- Duplication of a chromosome segment harboring gene families

Epidemiology

- Transmission at social gatherings



Biogeography

- Environmental changes that affect whole communities of species

Genome evolution

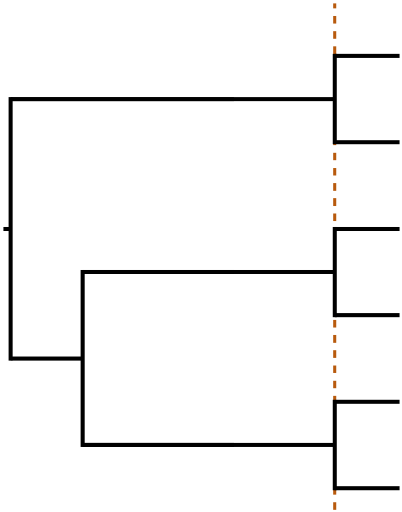
- Duplication of a chromosome segment harboring gene families

Epidemiology

- Transmission at social gatherings

Endosymbiont evolution (e.g., parasites, microbiome)

- Speciation of the host
- Co-colonization of new host species

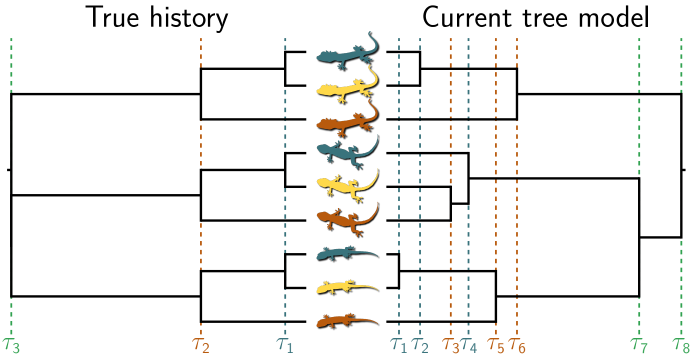


Why account for shared divergences?

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

Why account for shared divergences?

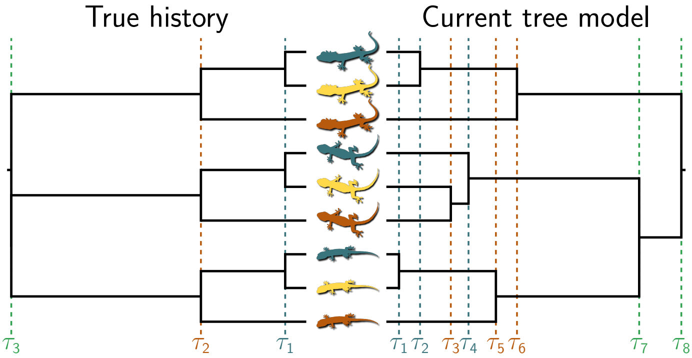
1. Improve inference



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

Why account for shared divergences?

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



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

Biogeography

- Environmental changes that affect whole communities of species

Genome evolution

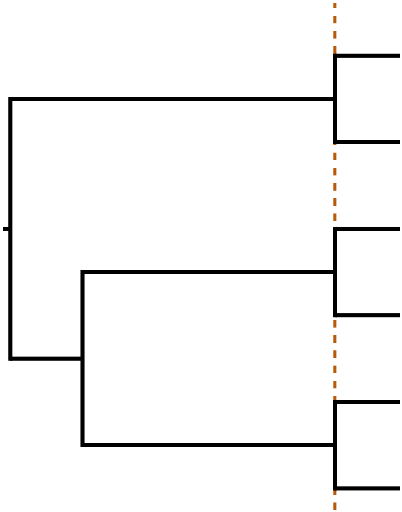
- Duplication of a chromosome segment harboring gene families

Epidemiology

- Transmission at social gatherings

Endosymbiont evolution (e.g., parasites, microbiome)

- Speciation of the host
- Co-colonization of new host species



Approaches to the problem

A pairwise approach (keep it "simple")

A fully phylogenetic approach

Tashitso Anamza

Tanner Myers

Randy Klabacka

Perry Wood, Jr.

Claire Tracy

Kerry Cobb

Matt Buehler

Nadia L'bahy

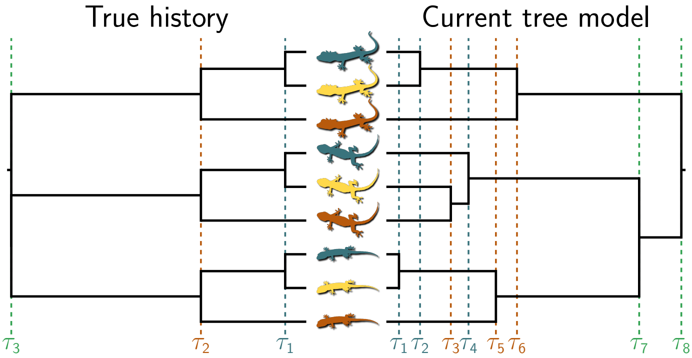
Approaches to the problem

A pairwise approach (keep it "simple")

A fully phylogenetic approach

Dr. Perry Wood, Jr.

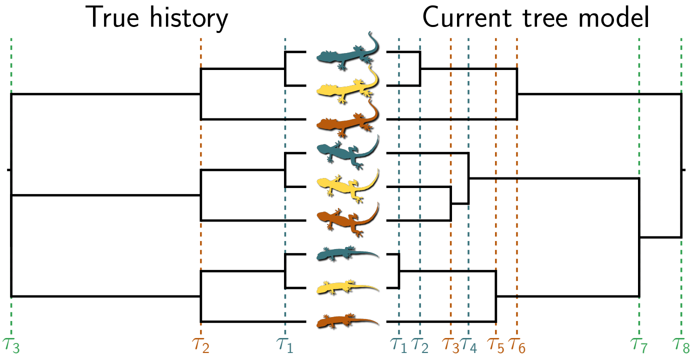
Challenges to accounting for shared divergences



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

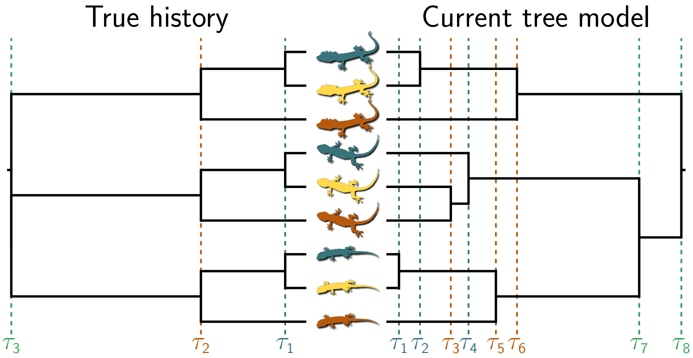
Challenges to accounting for shared divergences

1. Likelihood for genomic data is tricky

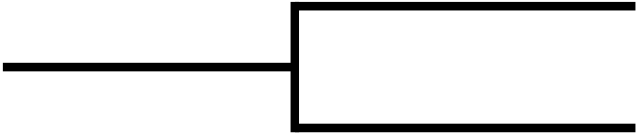


Challenges to accounting for shared divergences

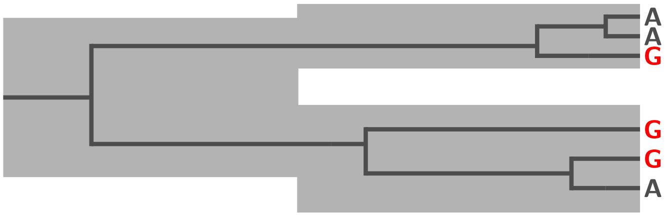
1. Likelihood for genomic data is tricky
2. Lots of possible trees of different dimensions

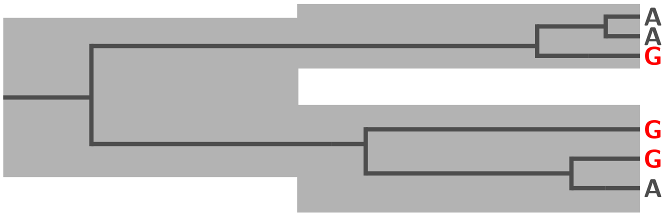


J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

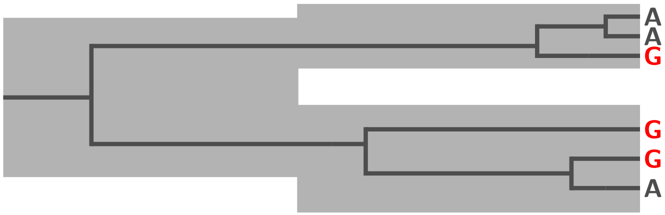



A
A
G
G
G
A

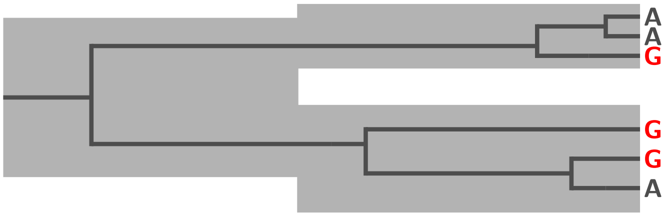

A
A
G
G
G
A



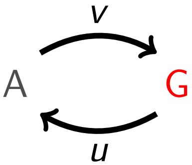
- Conditional on "population tree" (T), model "gene trees" (G) using coalescent

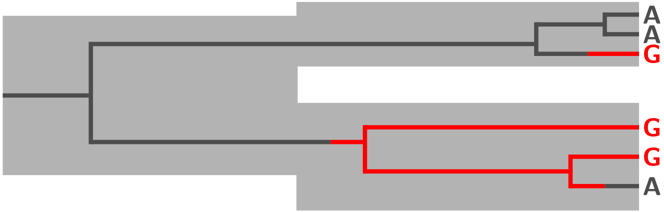


- Conditional on "population tree" (T), model "gene trees" (G) using coalescent
  - Coalescent is a stochastic model of shared inheritance (continuous-time Markov chain = CTMC)

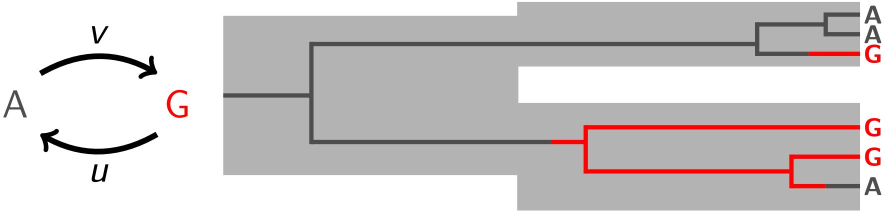


- Conditional on “population tree” (T), model “gene trees” (G) using coalescent
  - Coalescent is a stochastic model of shared inheritance (continuous-time Markov chain = CTMC)
  - Gene tree branching patterns are a function of population size

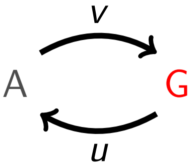


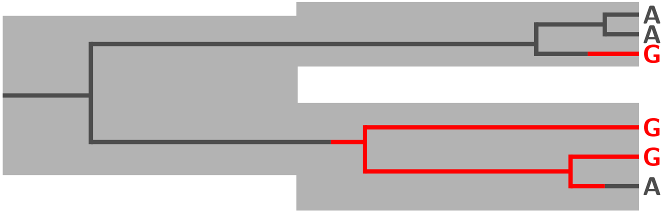


- Conditional on "population tree" (T), model "gene trees" (G) using coalescent
  - Coalescent is a stochastic model of shared inheritance (continuous-time Markov chain = CTMC)
  - Gene tree branching patterns are a function of population size
- Conditional on G, model mutation as a CTMC

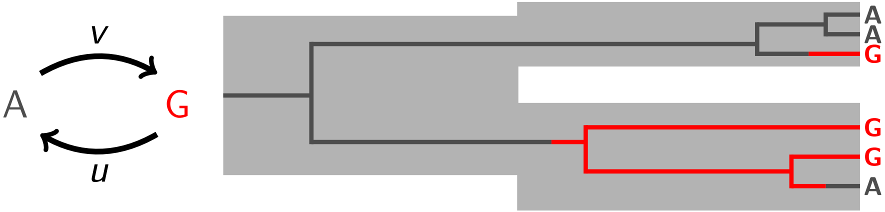


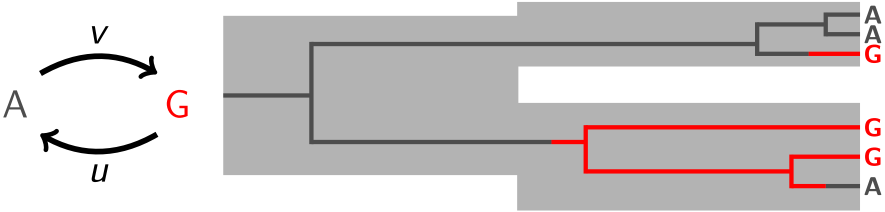
- Conditional on "population tree" (T), model "gene trees" (G) using coalescent
  - Coalescent is a stochastic model of shared inheritance (continuous-time Markov chain = CTMC)
  - Gene tree branching patterns are a function of population size
- Conditional on G, model mutation as a CTMC
- Genetic characters provide information about G



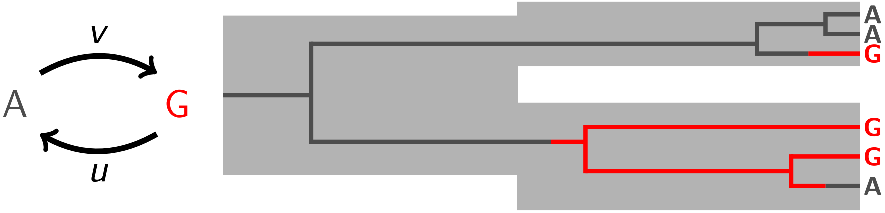


- Conditional on "population tree" (T), model "gene trees" (G) using coalescent
  - Coalescent is a stochastic model of shared inheritance (continuous-time Markov chain = CTMC)
  - Gene tree branching patterns are a function of population size
- Conditional on G, model mutation as a CTMC
- Genetic characters provide information about G
- G informs T (population sizes, divergence times, and relationships)

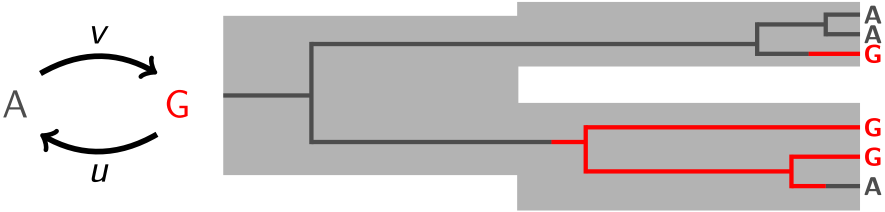

v
A
G
u
A
A
G
G
G
A



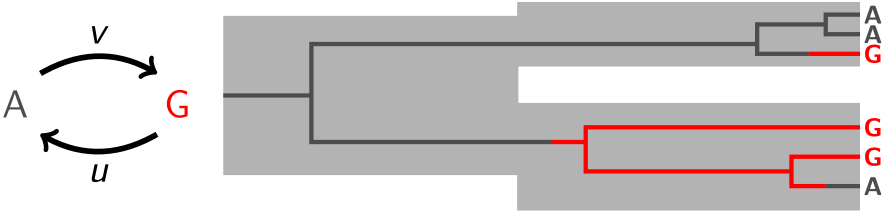
- “Standard” hierarchical approach



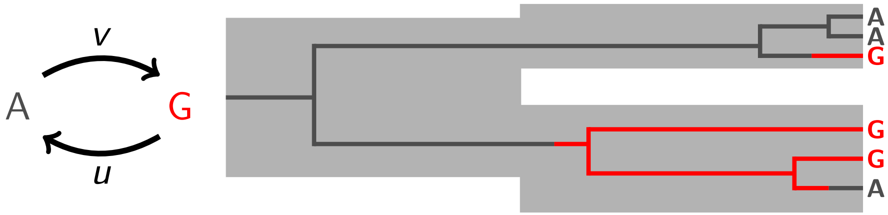
- “Standard” hierarchical approach
  - Calculate $p(\text{genetic data} \mid G) \times p(G \mid T)$



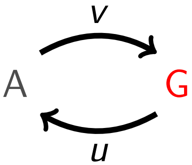
- "Standard" hierarchical approach
  - Calculate $p(\text{genetic data} \mid G) \times p(G \mid T)$
  - Use numerical integration (MCMC) to co-estimate both

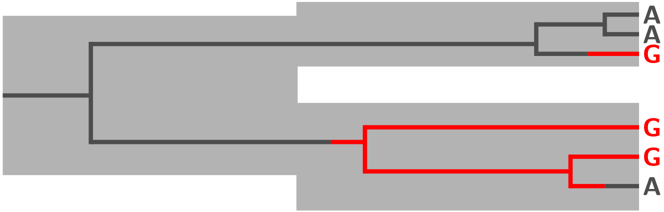


- "Standard" hierarchical approach
  - Calculate $p(\text{genetic data} \mid G) \times p(G \mid T)$
  - Use numerical integration (MCMC) to co-estimate both
- But, G and T are highly correlated

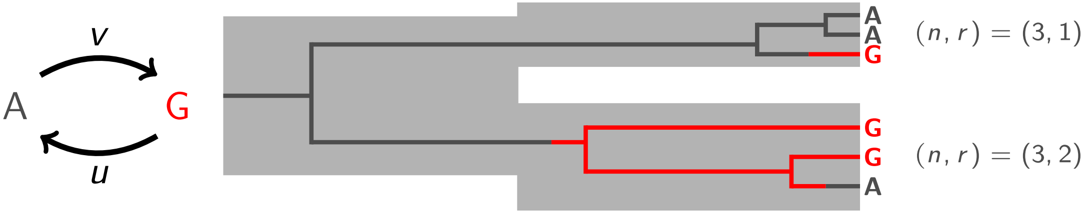


- "Standard" hierarchical approach
  - Calculate $p(\text{genetic data} \mid G) \times p(G \mid T)$
  - Use numerical integration (MCMC) to co-estimate both
- But, G and T are highly correlated
- As the number of loci (gene trees) increases, MCMC falls apart



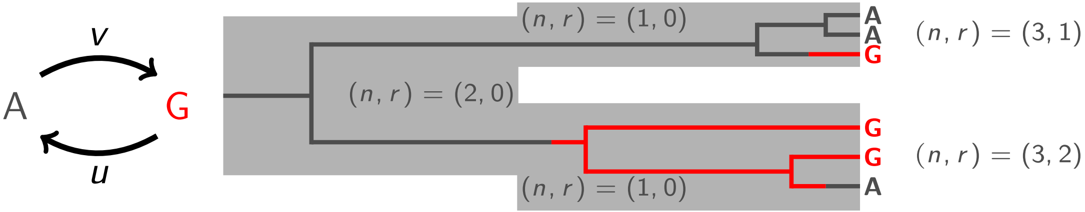


- "Standard" hierarchical approach
  - Calculate $p(\text{genetic data} \mid G) \times p(G \mid T)$
  - Use numerical integration (MCMC) to co-estimate both
- But, G and T are highly correlated
- As the number of loci (gene trees) increases, MCMC falls apart
- Can we integrate G analytically?



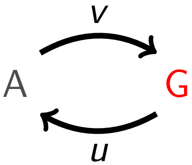
[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

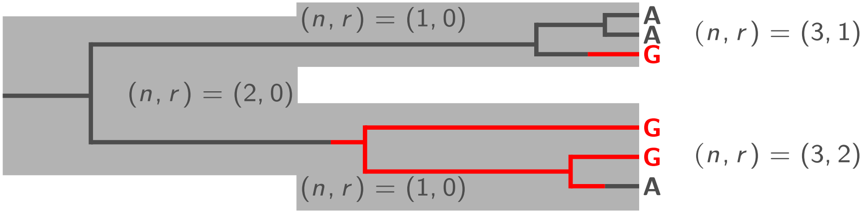
[2] D. Bryant et al. (2012). *Molecular Biology and Evolution* 29: 1917–1932



[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

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



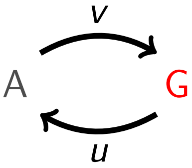


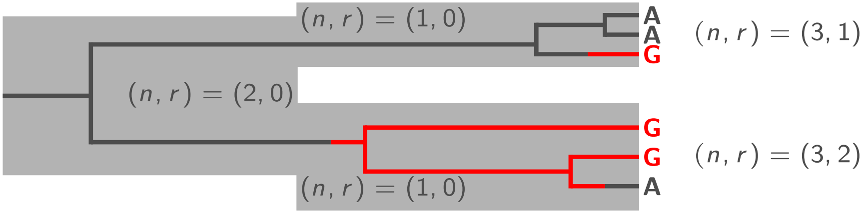
$Q =$

	(1, 0)	(1, 1)	(2, 0)	(2, 1)	⋯	**(n, n)**
(1, 0)	·	·	·	·		·
(1, 1)	·	·	·	·		·
(2, 0)	·	·	·	·		·
(2, 1)	·	·	·	·		·
⋮						
(n, n)	·	·	·	·		·

[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

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



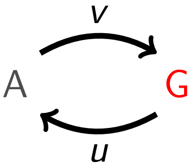


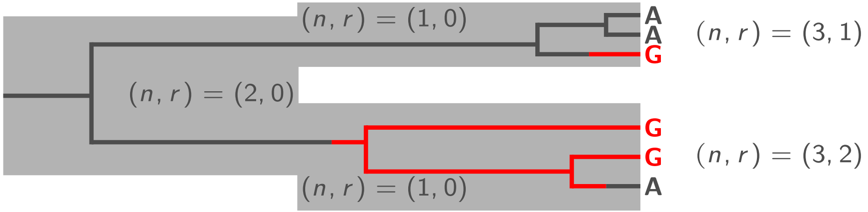
$$Q = \begin{array}{c|cccccc} & (1,0) & (1,1) & (2,0) & (2,1) & \cdots & (\mathbf{n},\mathbf{n}) \\ (1,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (1,1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,1) & \cdot & \cdot & \cdot & \cdot & & \cdot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

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

[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

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





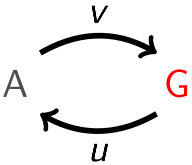
$$Q = \begin{array}{c|cccccc} & (1,0) & (1,1) & (2,0) & (2,1) & \cdots & (\mathbf{n},\mathbf{n}) \\ (1,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (1,1) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,0) & \cdot & \cdot & \cdot & \cdot & & \cdot \\ (2,1) & \cdot & \cdot & \cdot & \cdot & & \cdot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

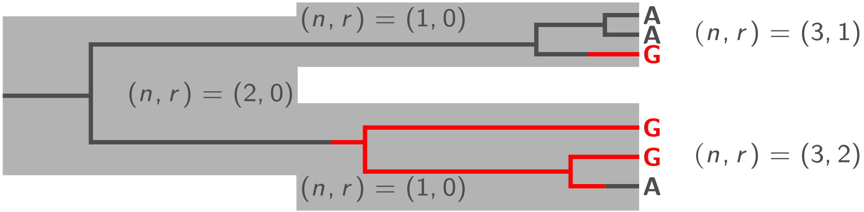
$$\begin{aligned} Q_{(n,r);(n,r-1)} &= (n-r+1)v, && \text{mutation,} \\ Q_{(n,r);(n,r+1)} &= (r+1)u, && \text{mutation,} \\ Q_{(n,r);(n-1,r)} &= \tfrac{(n-1-r)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n-1,r-1)} &= \tfrac{(r-1)n}{2N_e(u+v)}, && \text{coalescence,} \\ Q_{(n,r);(n,r)} &= -\tfrac{(n-1)n}{2N_e(u+v)}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

- e^{Qt} to keep track of all conditional probabilites along each branch (Carathéodory-Fejér method[1])

[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

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





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

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

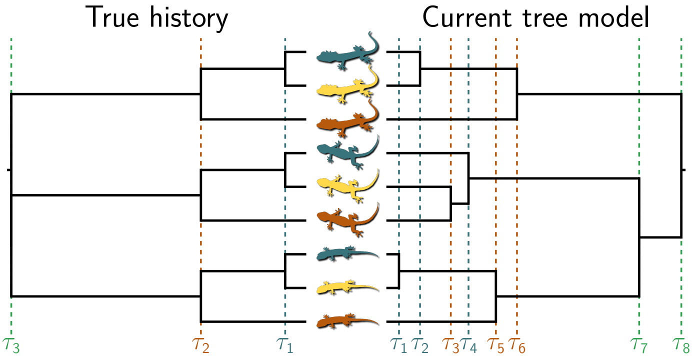
- e^{Qt} to keep track of all conditional probabilites along each branch (Carathéodory-Fejér method[1])
- At root, get likelihood of population tree integrated over all possible gene trees and mutational histories[2]

[1] T. Schmelzer and L. N. Trefethen (2007). *Electronic Transactions on Numerical Analysis* 29: 1–18

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

Challenges to accounting for shared divergences

1. Likelihood for genomic data is tricky
2. Lots of possible trees of different dimensions

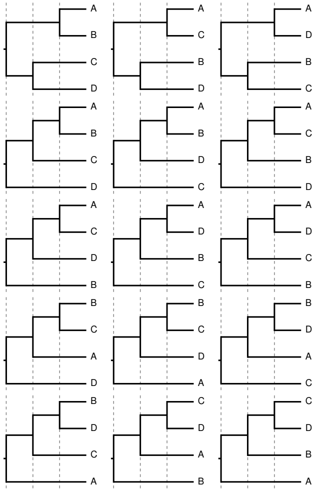


J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Generalizing tree space

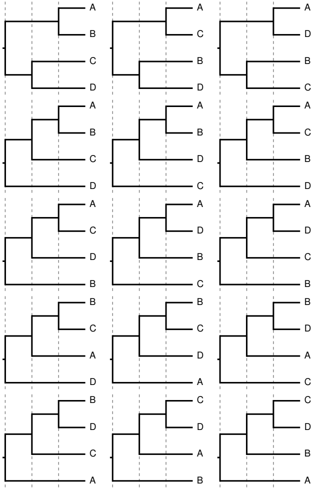
Generalizing tree space

$n_{\mathcal{T}} = 3$

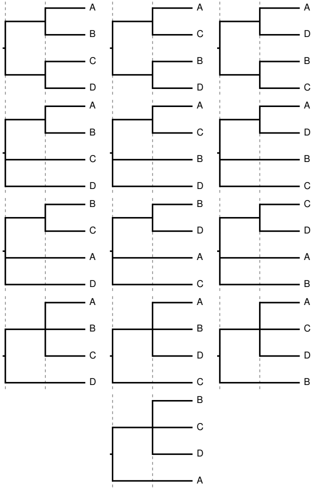


Generalizing tree space

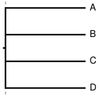
$n_{\mathcal{T}} = 3$



$n_{\mathcal{T}} = 2$

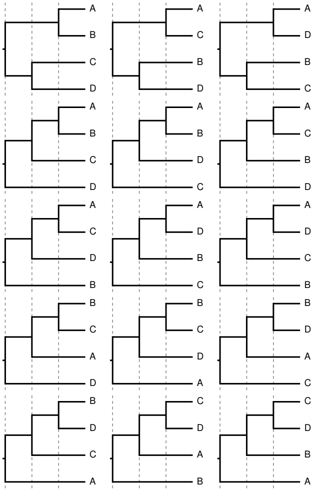


$n_{\mathcal{T}} = 1$

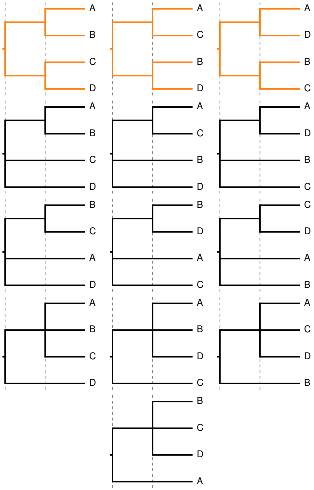


Generalizing tree space

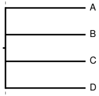
$n_{\mathcal{T}} = 3$



$n_{\mathcal{T}} = 2$

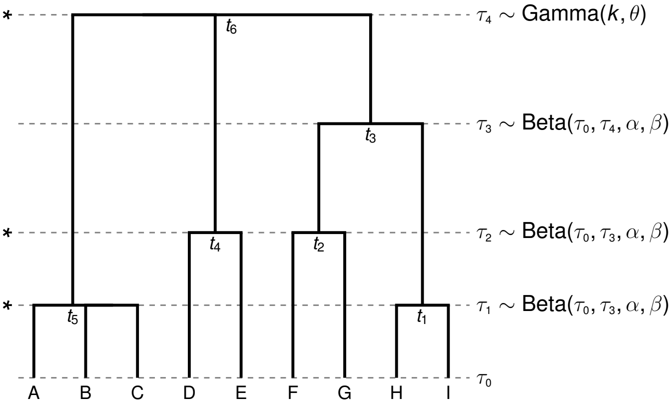


$n_{\mathcal{T}} = 1$



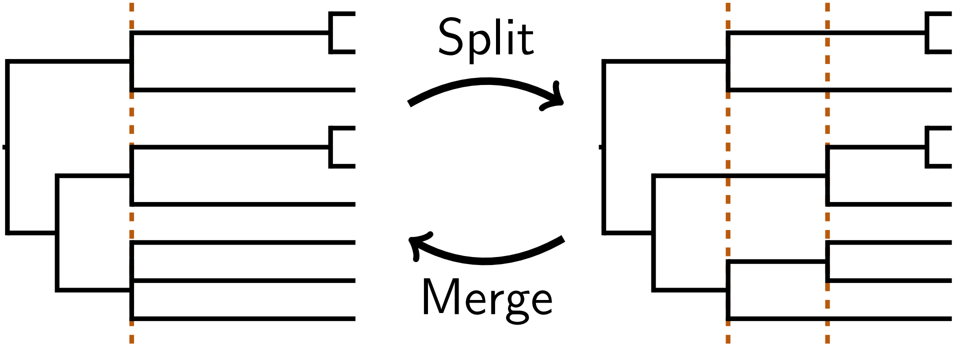
Generalized tree distribution

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



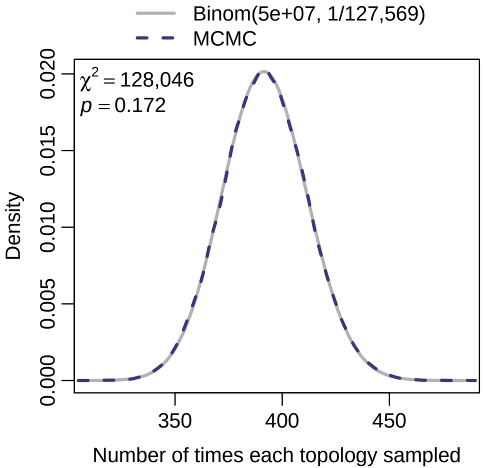
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Inferring trees with shared divergences



Reversible-jump MCMC

Validating rjMCMC with 7-leaf tree



The rjMCMC algorithms sample the expected generalized tree distribution

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

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

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

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 genealogies[1]

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

PhycoEval

Phylogenetic coevality

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

Ecoevolity

Estimating **evo**lutionary **coevality**

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

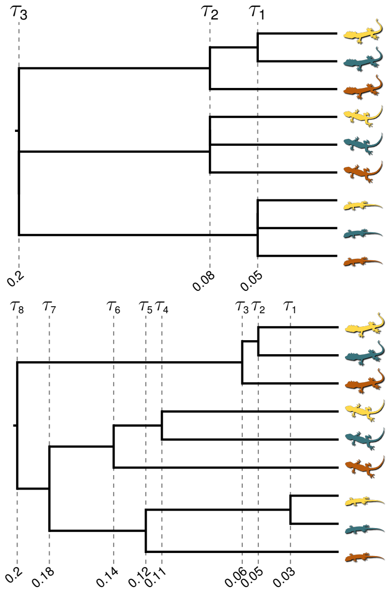
- **Tree model**
  - rjMCMC sampling of generalized tree distribution
- **Likelihood model**
  - CTMC model of characters evolving along genealogies
  - Infer species trees by analytically integrate over genealogies[1]
- *Goal: Co-estimation of phylogeny and shared divergences from genomic data*

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

Does it work?

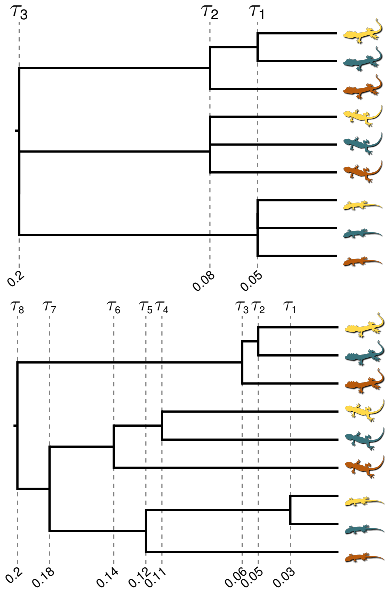
Methods: Simulations

- Simulated 100 data sets with 50,000 base pairs



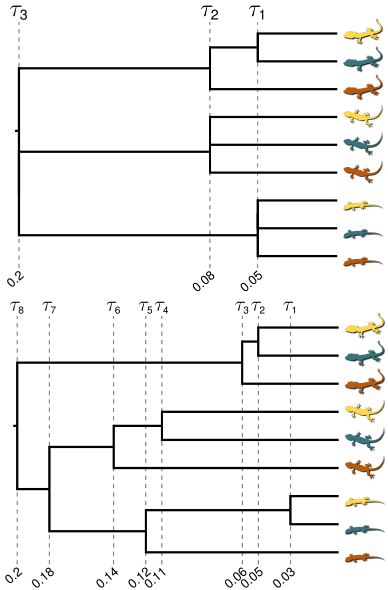
Methods: Simulations

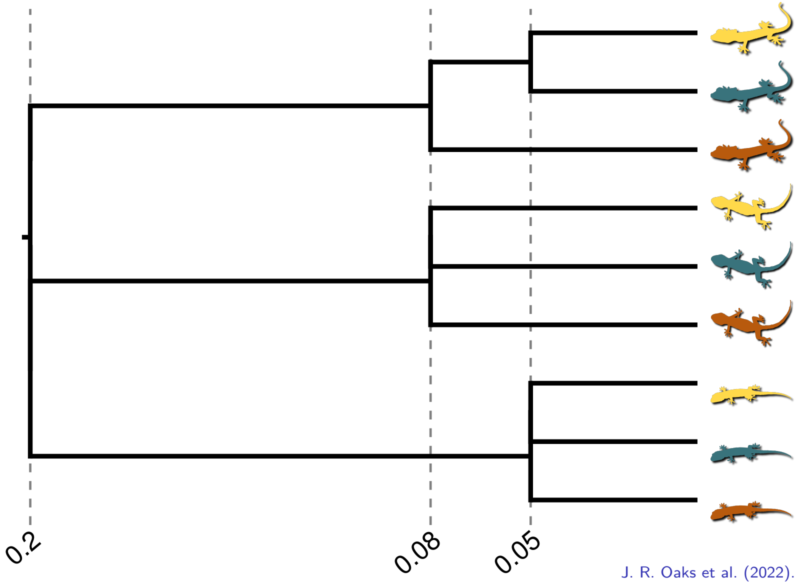
- Simulated 100 data sets with 50,000 base pairs
- Analyzed each data set with:
  - M_G = Generalized tree model
  - M_{IB} = Independent-bifurcating tree model



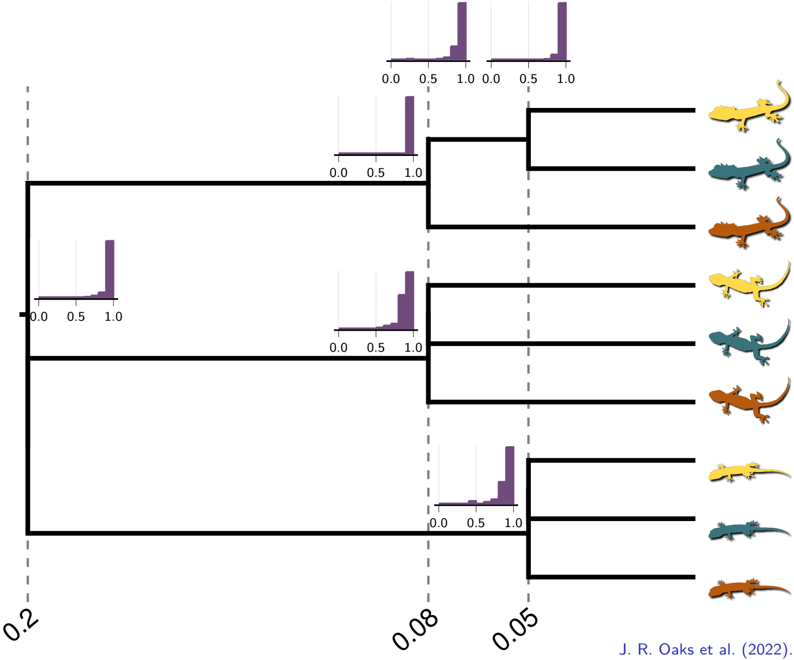
Methods: Simulations

- 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}

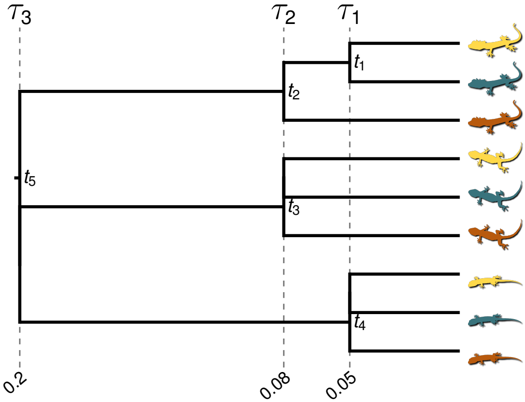




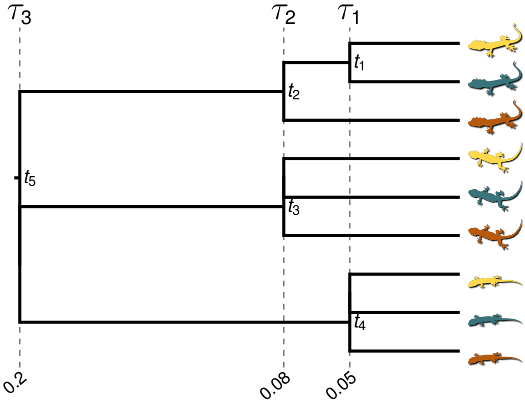
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

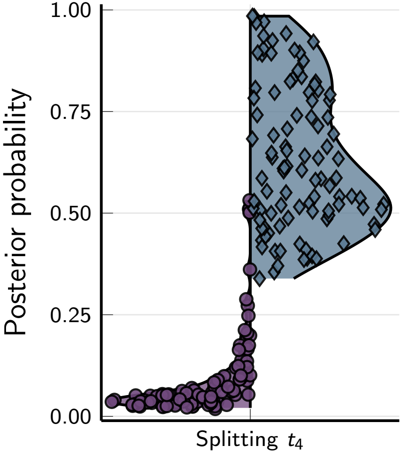


J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

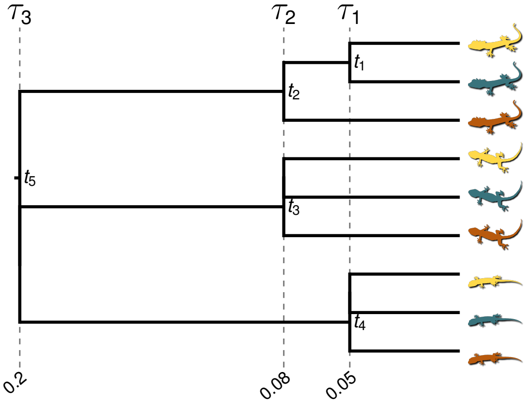

τ_3
τ_2
τ_1
t_1
t_2
t_3
t_4
t_5
0.2
0.08
0.05

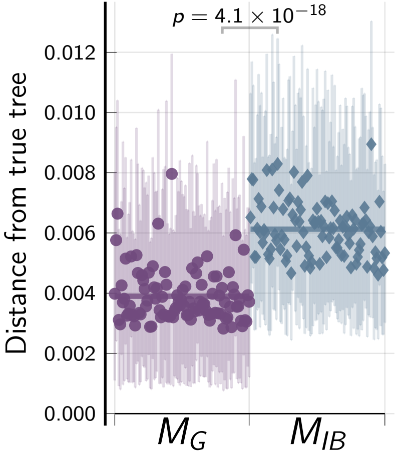
M_G = Generalized model M_{IB} = Independent-bifurcating model





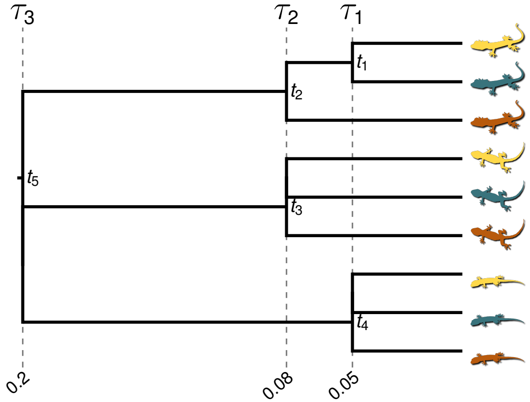
M_G = Generalized model M_{IB} = Independent-bifurcating model

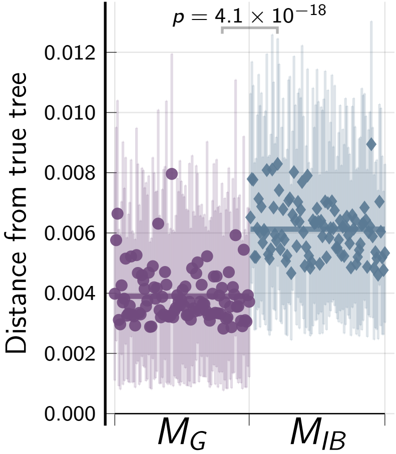




J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

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

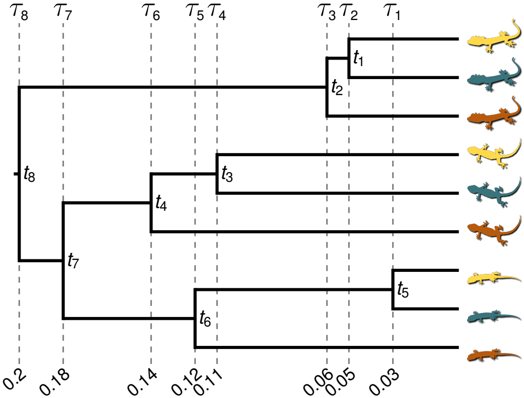




M_G significantly better at inferring trees with shared divergences

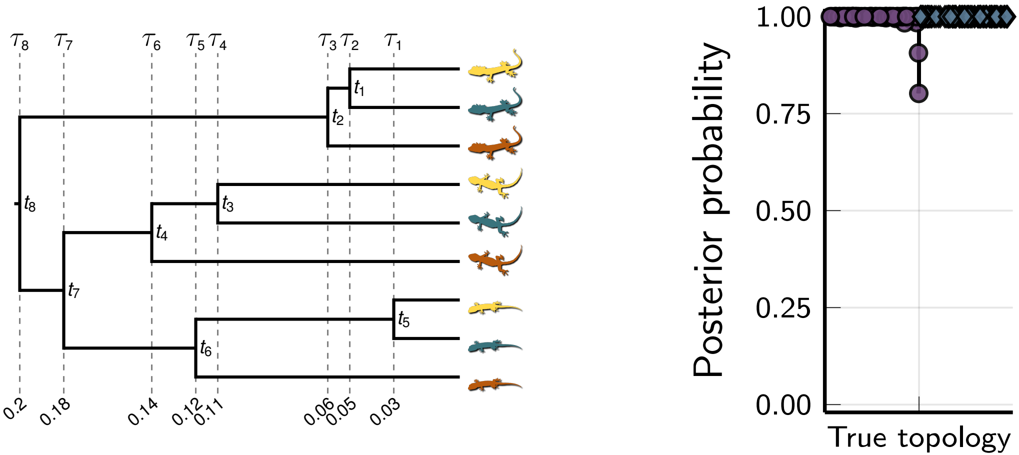
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

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



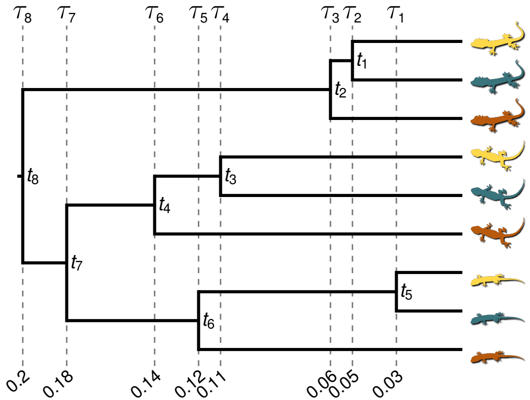
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

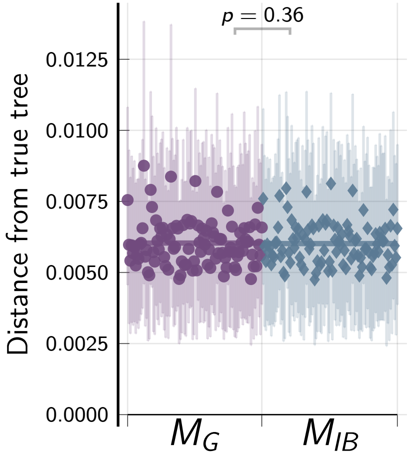
M_G = Generalized model M_{IB} = Independent-bifurcating model



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

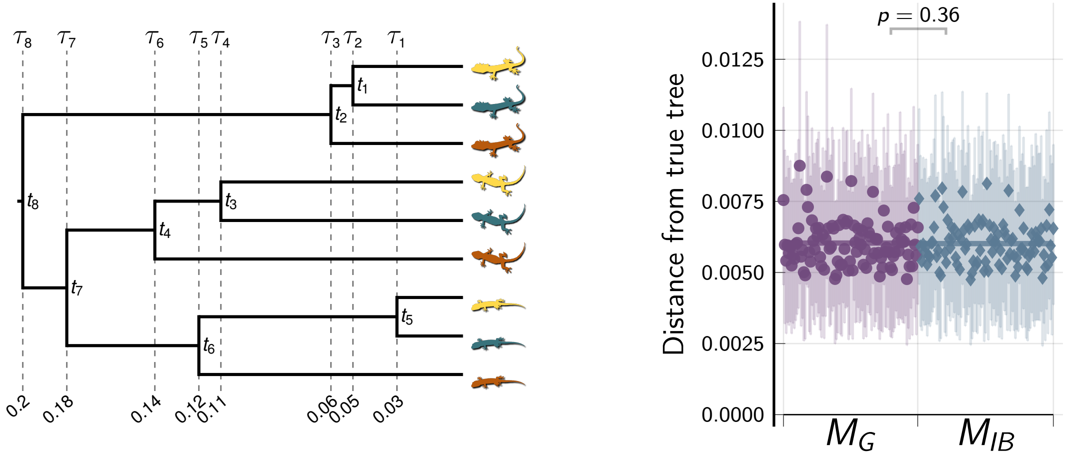
M_G = Generalized model M_{IB} = Independent-bifurcating model





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

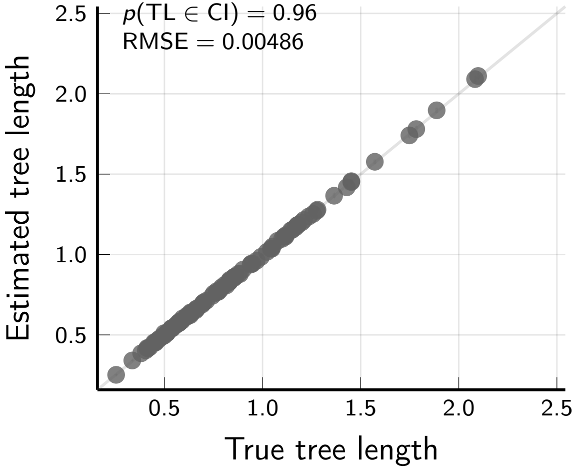
M_G = Generalized model M_{IB} = Independent-bifurcating model



M_G performs as well as true model when divergences are independent

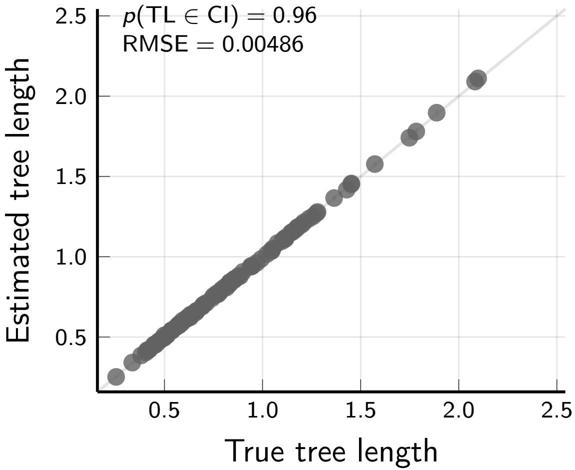
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

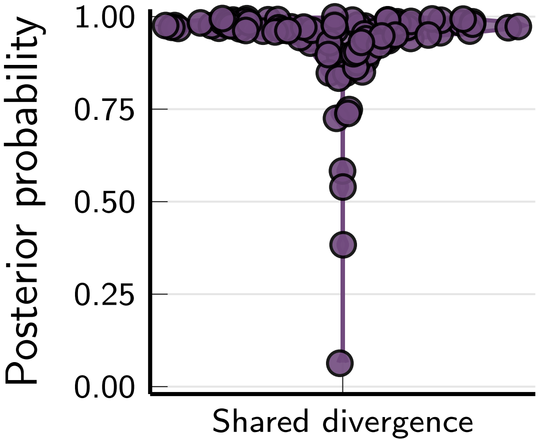
Results: random M_G trees



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

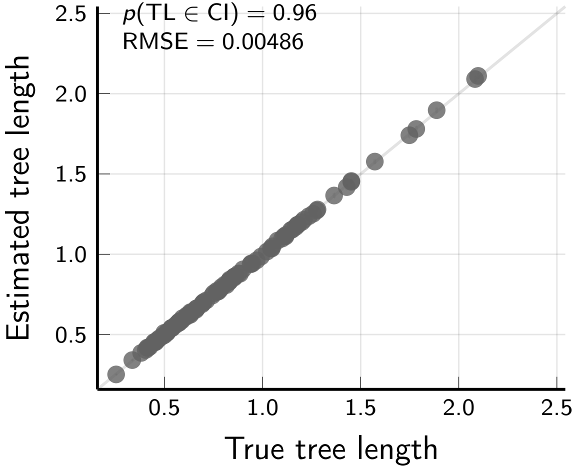
Results: random M_G trees

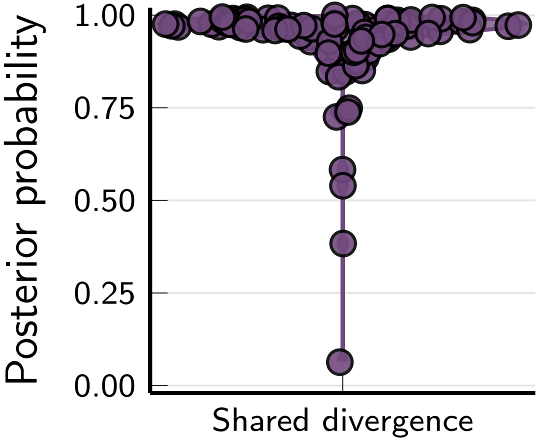




J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Results: random M_G trees

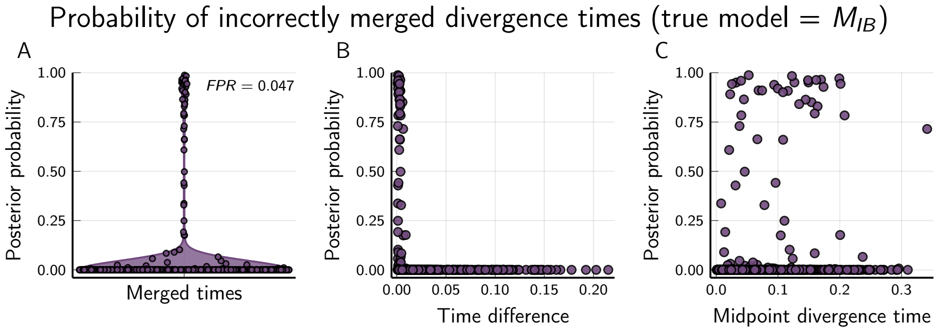




M_G performs well with data simulated on random trees with shared divergences

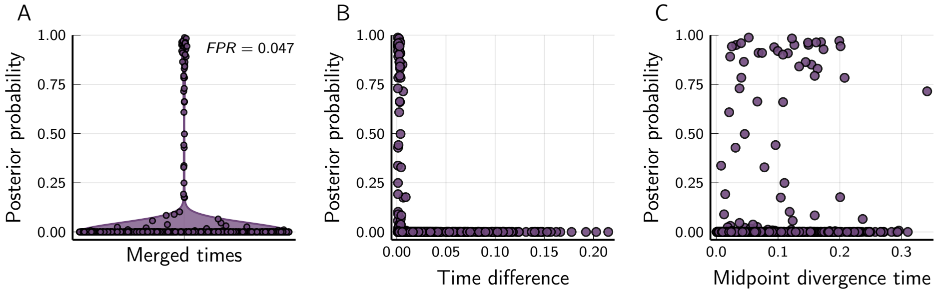
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Results: random M_{IB} trees

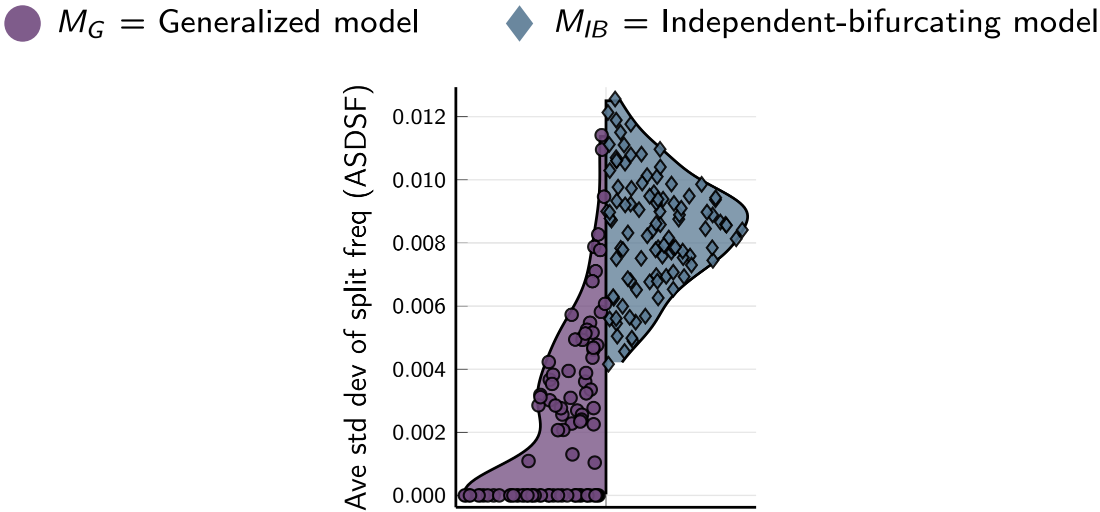


Results: random M_{IB} trees

Probability of incorrectly merged divergence times (true model = M_{IB})

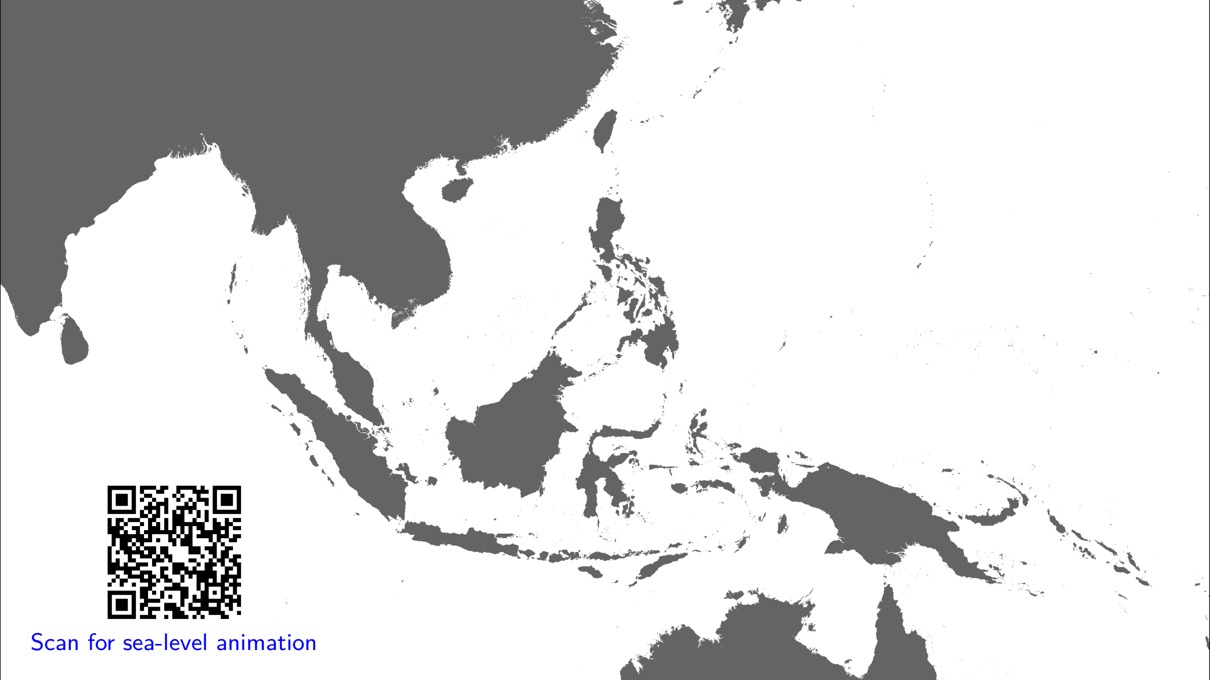


M_G has low false positive rate

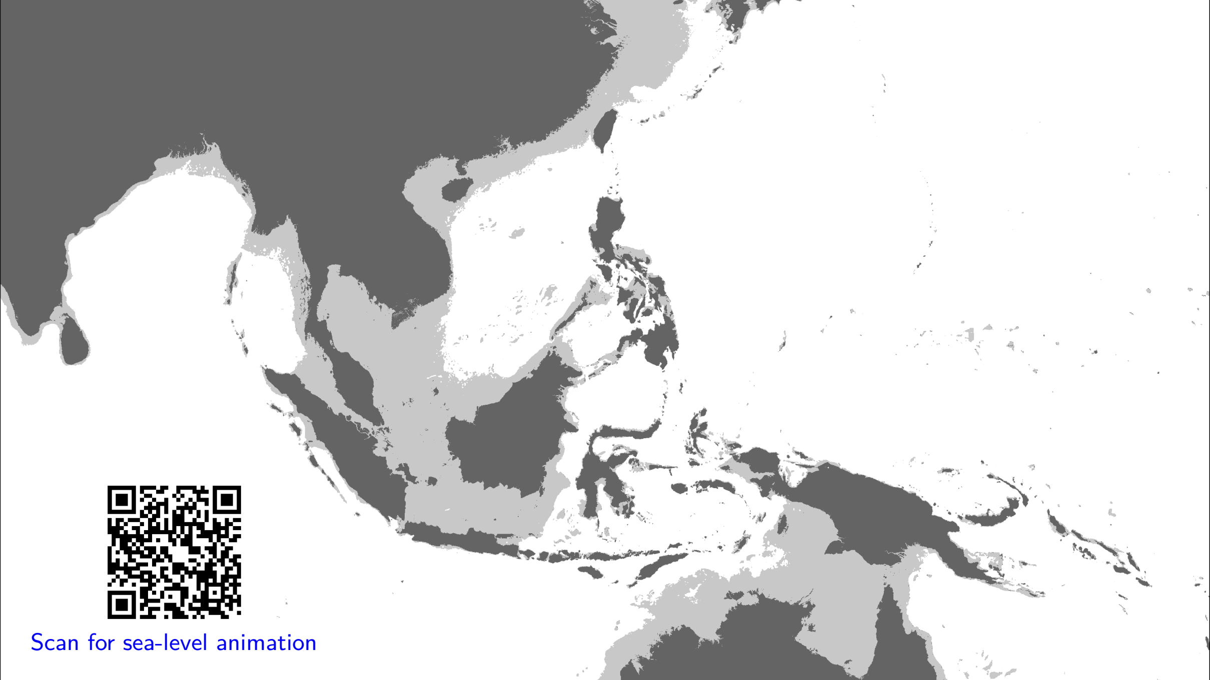


Generalizing tree space improves MCMC convergence and mixing

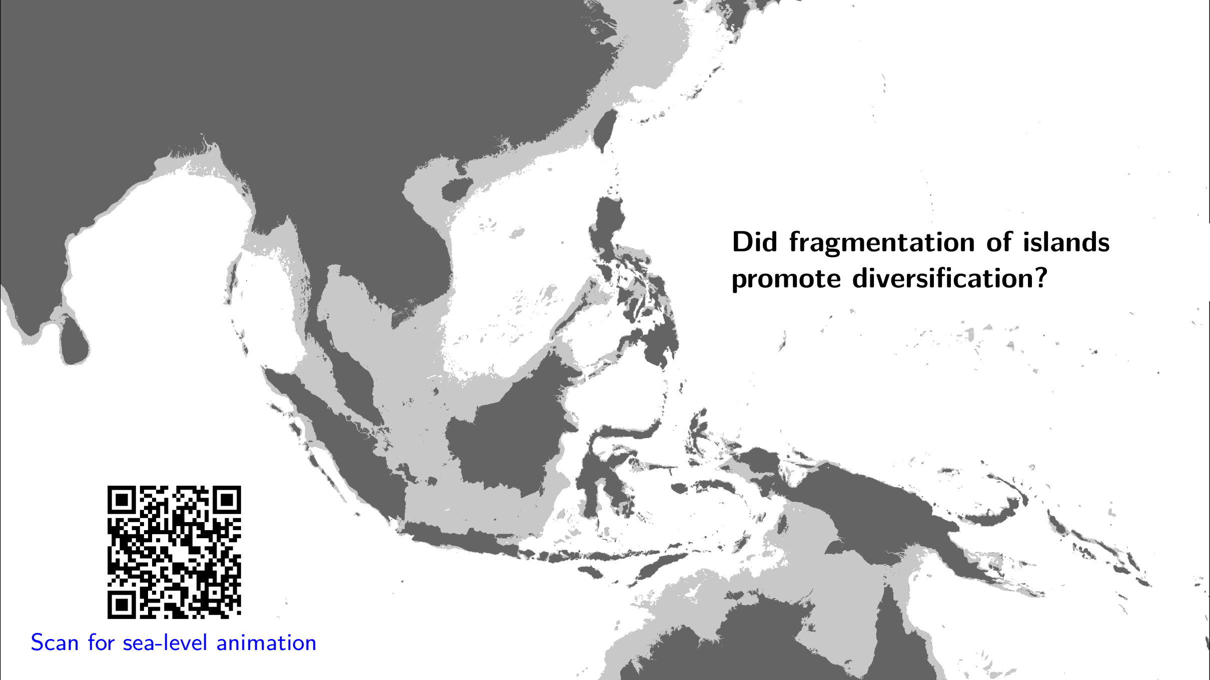
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119



Scan for sea-level animation



Scan for sea-level animation

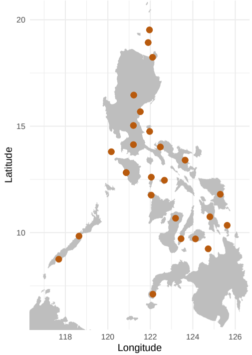


Did fragmentation of islands promote diversification?

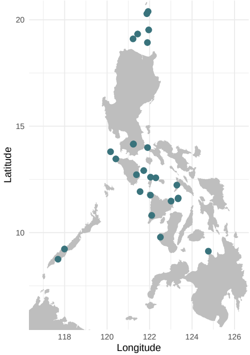
Scan for sea-level animation

Cyrtodactylus

©Rafe M. Brown



Gekko

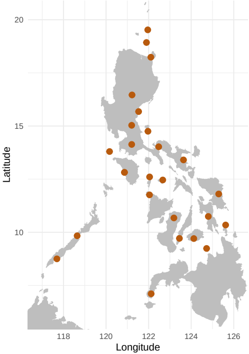


©Rafe M. Brown

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

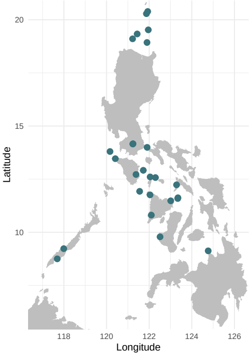
Cyrtodactylus

©Rafe M. Brown



1702 loci
155,887 sites

Gekko

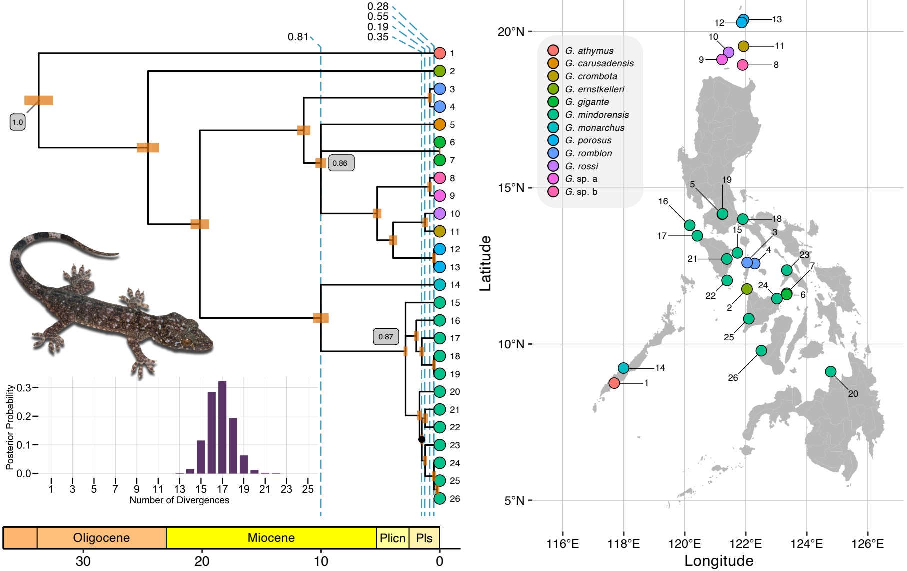


©Rafe M. Brown

1033 loci
94,813 sites

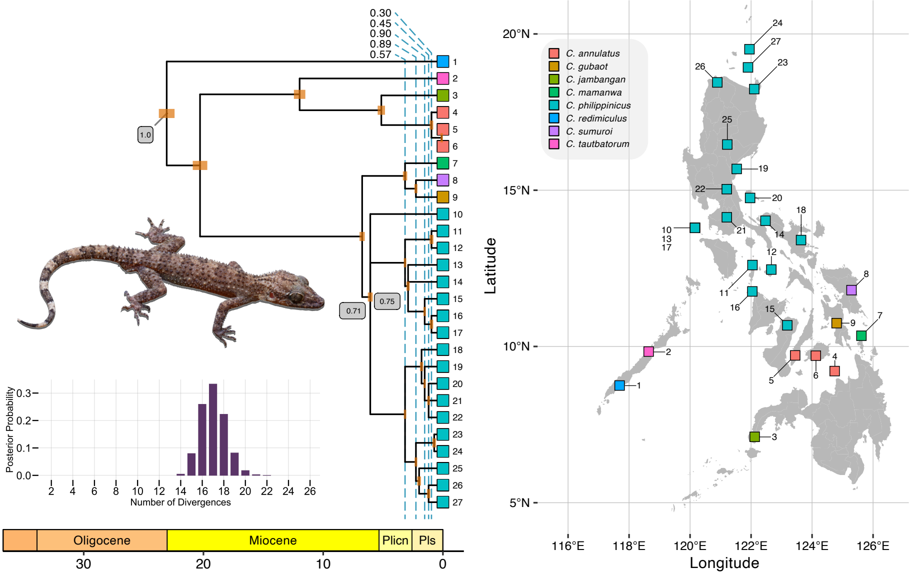
J. R. Oaks et al. (2022). *PNAS* 119: e2121036119

Gekko



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

Cyrtodactylus



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

Take-home points

- We can accurately infer phylogenies with shared divergences

Take-home points

- We can accurately infer phylogenies with shared divergences
- Generalizing tree space avoids spurious support incorrect relationships and improves MCMC mixing

Take-home points

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

Open science: everything is available. . .

Software:

- Phycoeval: github.com/phyletica/ecoevolity
  (release coming soon)

Open-Science Notebooks:

- Phycoeval analyses: github.com/phyletica/phycoeval-experiments
- Gecko RADseq: github.com/phyletica/gekgo

Moving forward: Theory/methods

Moving forward: Theory/methods

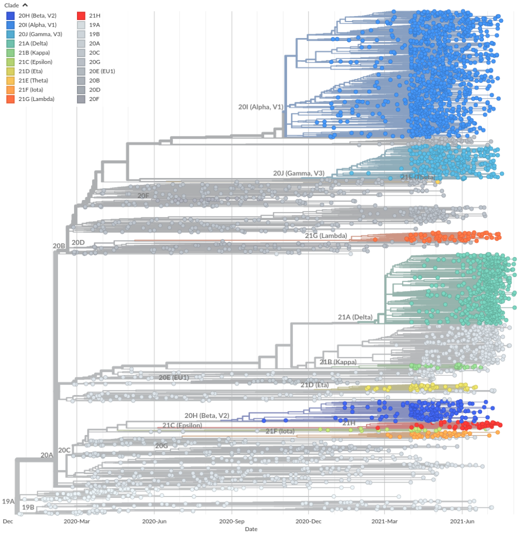
- Develop process-based and trait-dependent distributions over the space of generalized trees
  - "Birth-death-burst" model

Moving forward: Theory/methods

- Develop process-based and trait-dependent distributions over the space of generalized trees
  - "Birth-death-burst" model
- Extend generalized tree distribution to trees that are not ultrametric

Moving forward: Theory/methods

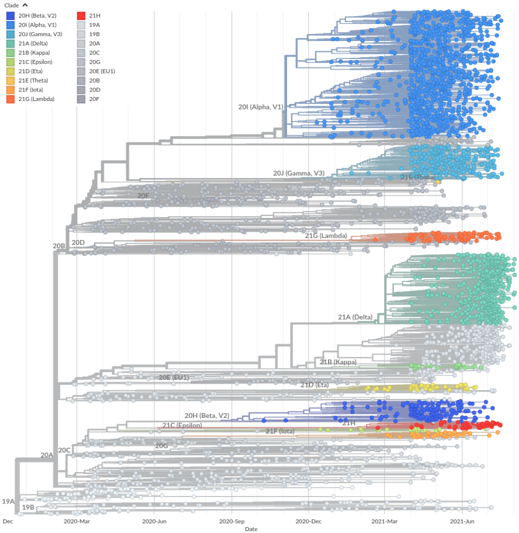
- Develop process-based and trait-dependent distributions over the space of generalized trees
  - "Birth-death-burst" model
- Extend generalized tree distribution to trees that are not ultrametric



nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

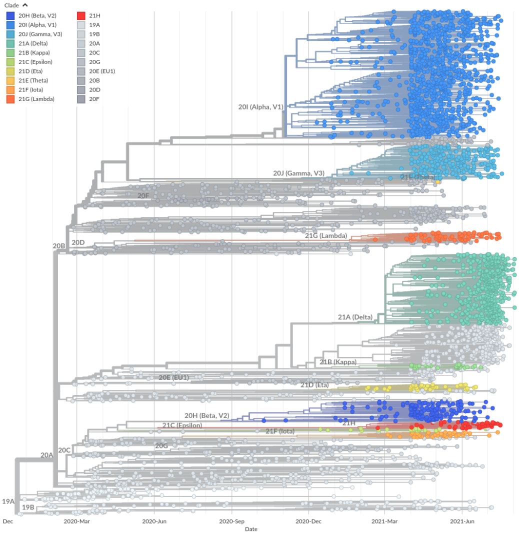
Moving forward: Theory/methods

- Develop process-based and trait-dependent distributions over the space of generalized trees
  - "Birth-death-burst" model
- Extend generalized tree distribution to trees that are not ultrametric
- Couple generalized tree distribution with other phylogenetic likelihood models



nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

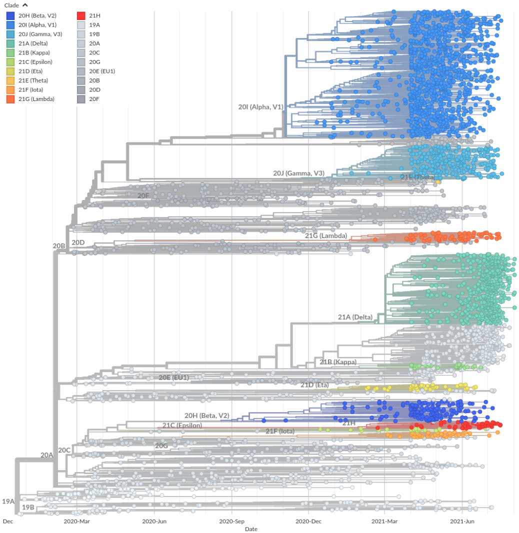
Moving forward: Applications



nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

Moving forward: Applications

Epidemiological dynamics of “super-spreading” events during the COVID-19 pandemic

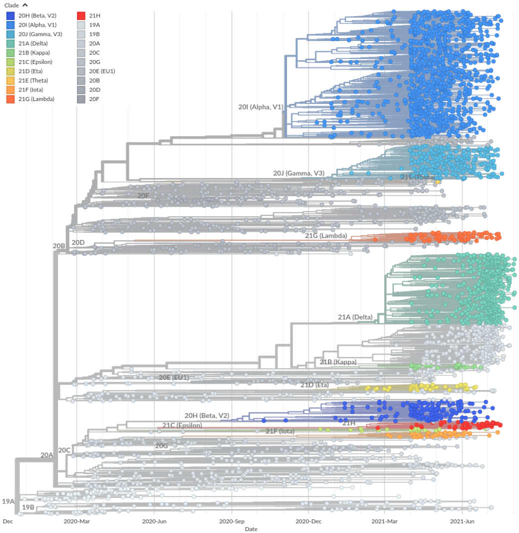


nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

Moving forward: Applications

Epidemiological dynamics of "super-spreading" events during the COVID-19 pandemic

- Spread at social gatherings creates shared and multifurcating divergences in the viral "transmission tree"

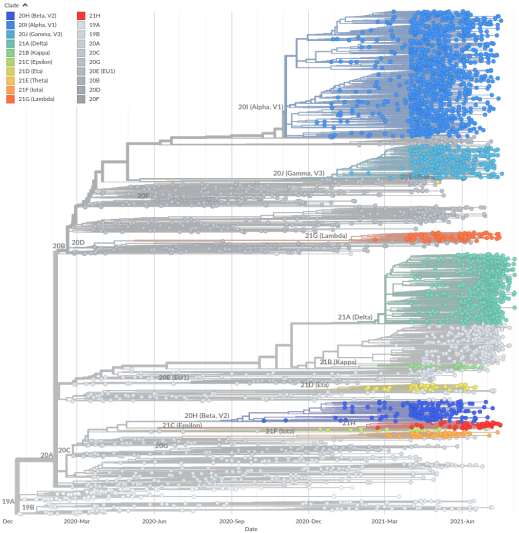


nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

Moving forward: Applications

Epidemiological dynamics of “super-spreading” events during the COVID-19 pandemic

- Spread at social gatherings creates shared and multifurcating divergences in the viral “transmission tree”
- Estimate rate of shared divergences as proxy for spread via social gatherings

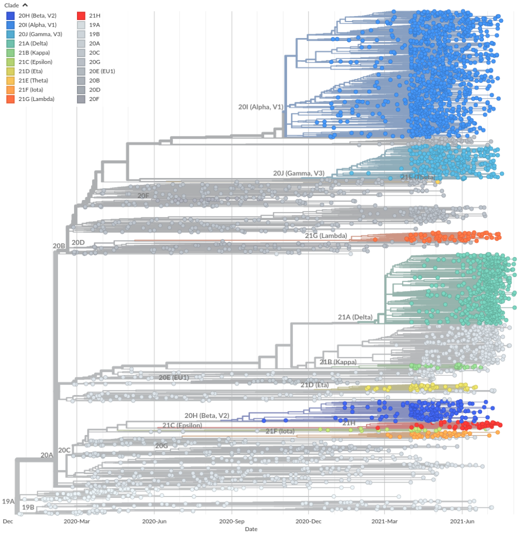


nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

Moving forward: Applications

Epidemiological dynamics of “super-spreading” events during the COVID-19 pandemic

- Spread at social gatherings creates shared and multifurcating divergences in the viral “transmission tree”
- Estimate rate of shared divergences as proxy for spread via social gatherings
- Test if this varies over time, among regions, and among variants of SARS-CoV-2

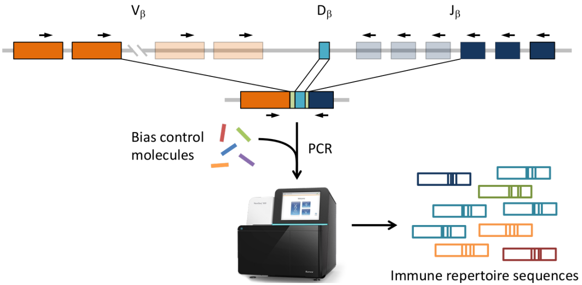


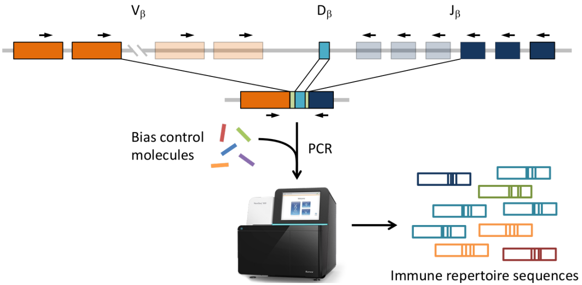
nextstrain.org J. Hadfield et al. (2018). *Bioinformatics* 34: 4121–4123

AUBURN UNIVERSITY

Adaptive biotechnologies™

© 2007 Boris Kulikov boris-kulikov.blogspot.com





V_β
D_β
J_β
Bias control molecules
PCR
Immune repertoire sequences

Stats & Algorithms

- Our team is very cross-functional
  - We solve statistical and computational challenges for many stakeholders across Adaptive

Erin Calfee
Computational Biologist I

Travers Ching
Sr Computational Biologist

Amy Ko
Computational Biologist II

Jamie Oaks
Sr Manager, Computational Bi...

© Copyright Adaptive Biotechnologies 2021. Confidential.

Adaptive biotechnologies™

Come work with us!

Adaptive Comp Bio Internships!

Scan to internship listing:

www.adaptivebiotech.com/career-listings

Adaptive biotechnologies™

Internships

Intern, BCR Discovery
Seattle

Intern, Cloud Engineer
Remote (WFH)

Intern, Computational Biology
Remote (WFH)

Intern, Computational Biology
Remote (WFH)

Intern, Computational Biology, Antigen Map
Remote (WFH)

Intern, Computational Biology (CRI)
Remote (WFH)

Intern, Computational Biology, Stats and Algorithms
Remote (WFH)

Thanks everyone!

- Thanks Devang and Duke GCB & CBB!

- Bryan Howie and my team at Adaptive
- Phyletica Lab (the Phyleticians)
- Mark Holder
- Rafe Brown
- Cam Siler
- Lee Grismer

Computation:

- Alabama Supercomputer Authority
- Auburn University Hopper Cluster

Funding:

NSF

AUBURN UNIVERSITY

Photo credits:

- Rafe Brown
- Perry Wood, Jr.
- PhyloPic

Questions?

joaks@auburn.edu

@jamoaks

phyletica.org

Scan for slides:

phyletica.org/slides/duke-cbb.pdf

© 2007 Boris Kulikov boris-kulikov.blogspot.com