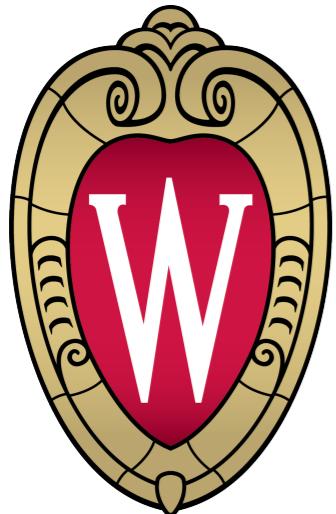


Estimating phylogenetic networks from multi locus sequence data

Claudia Solís-Lemus, PhD

University of Wisconsin-Madison
Wisconsin Institute for Discovery
Department of Plant Pathology



<https://solislemuslab.github.io/>



mstdn.social/@solislemuslab

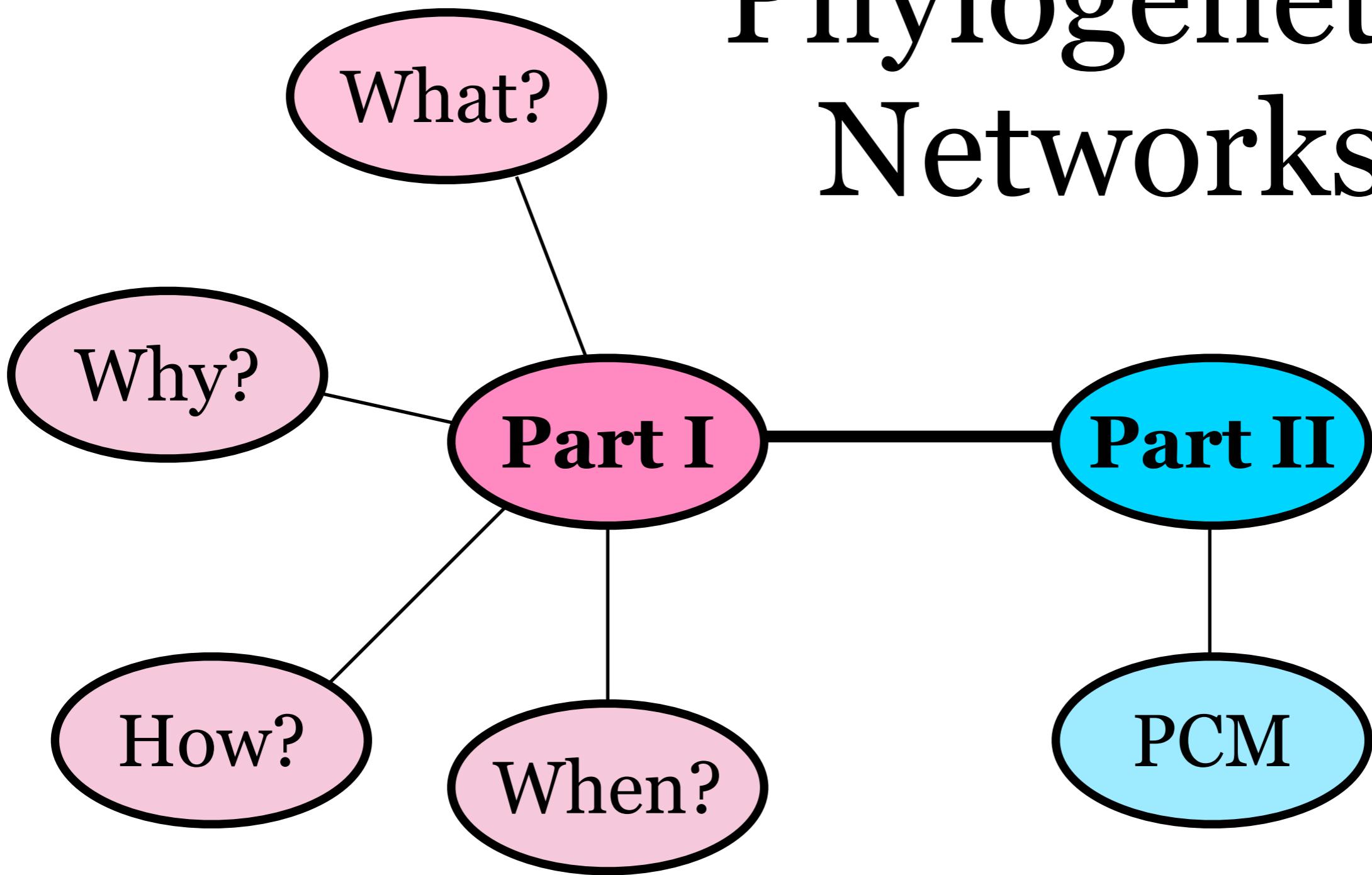


crsl4



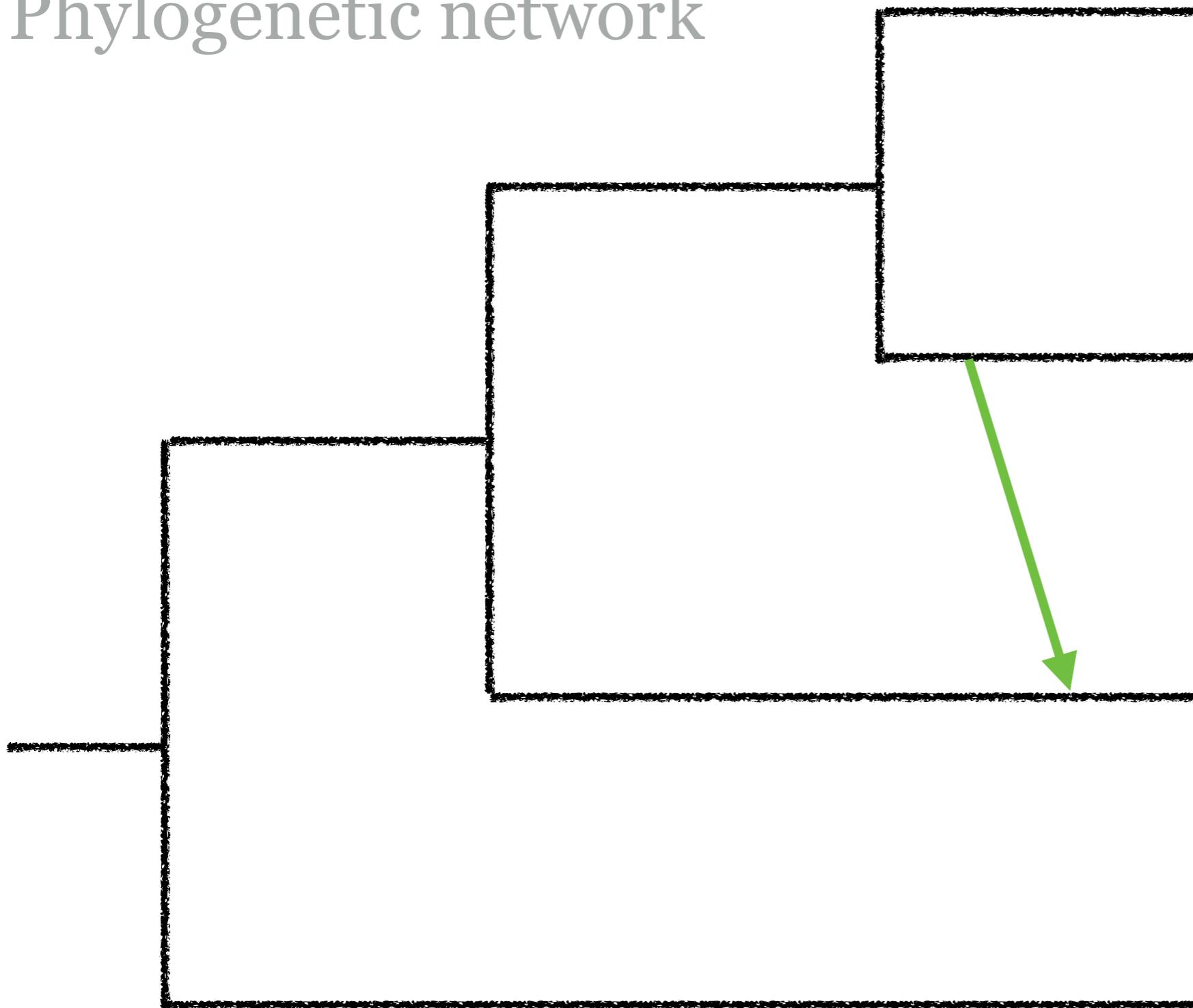
@thestatistician

Phylogenetic Networks



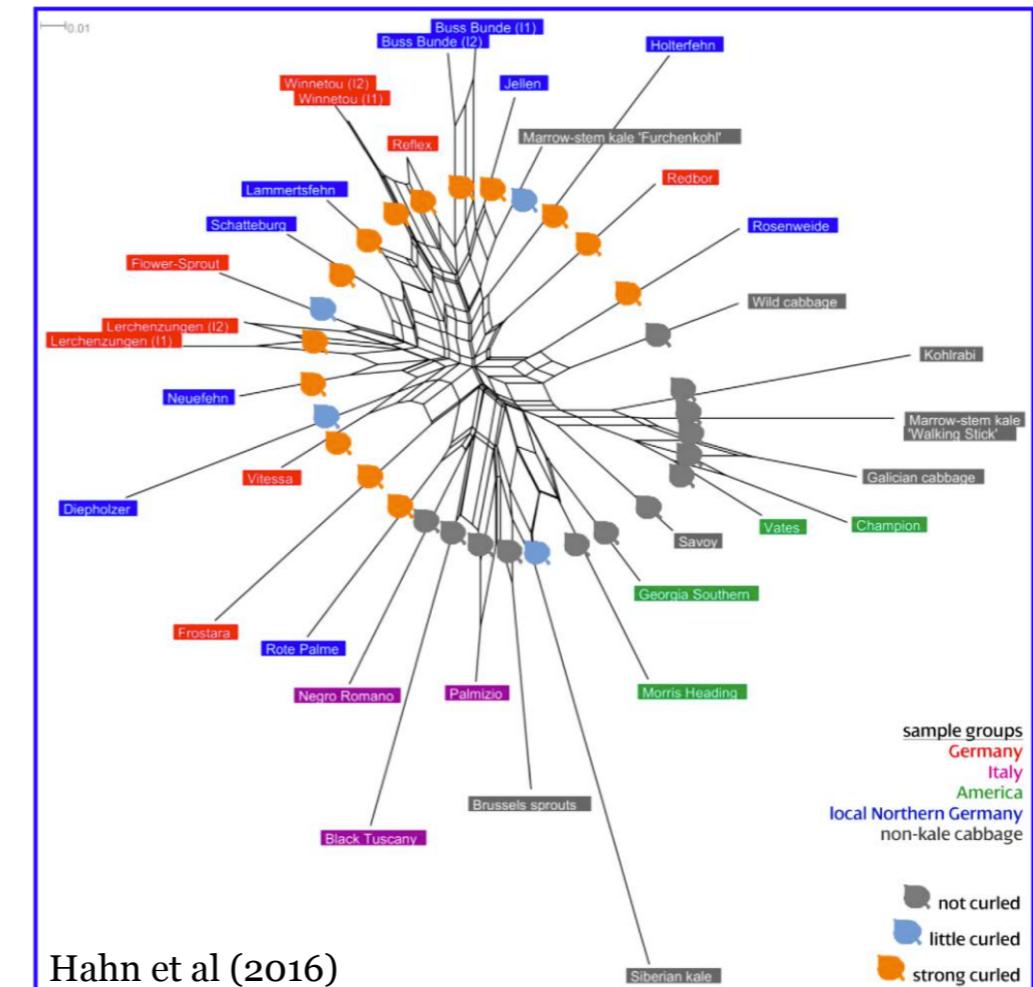
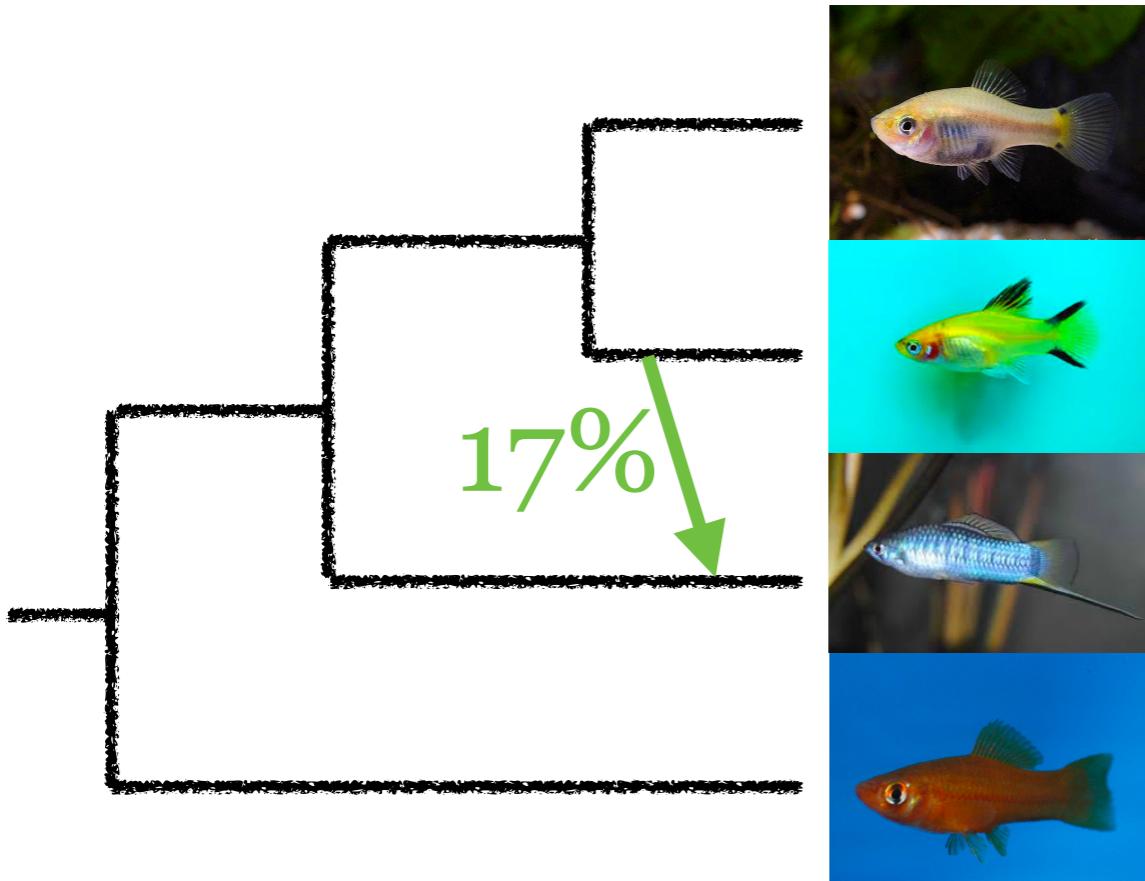
What?

Phylogenetic network



What?

Phylogenetic network

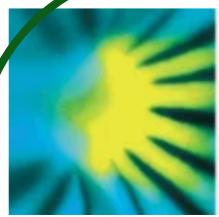


Explicit

Implicit

Why?

Phylogenetic network

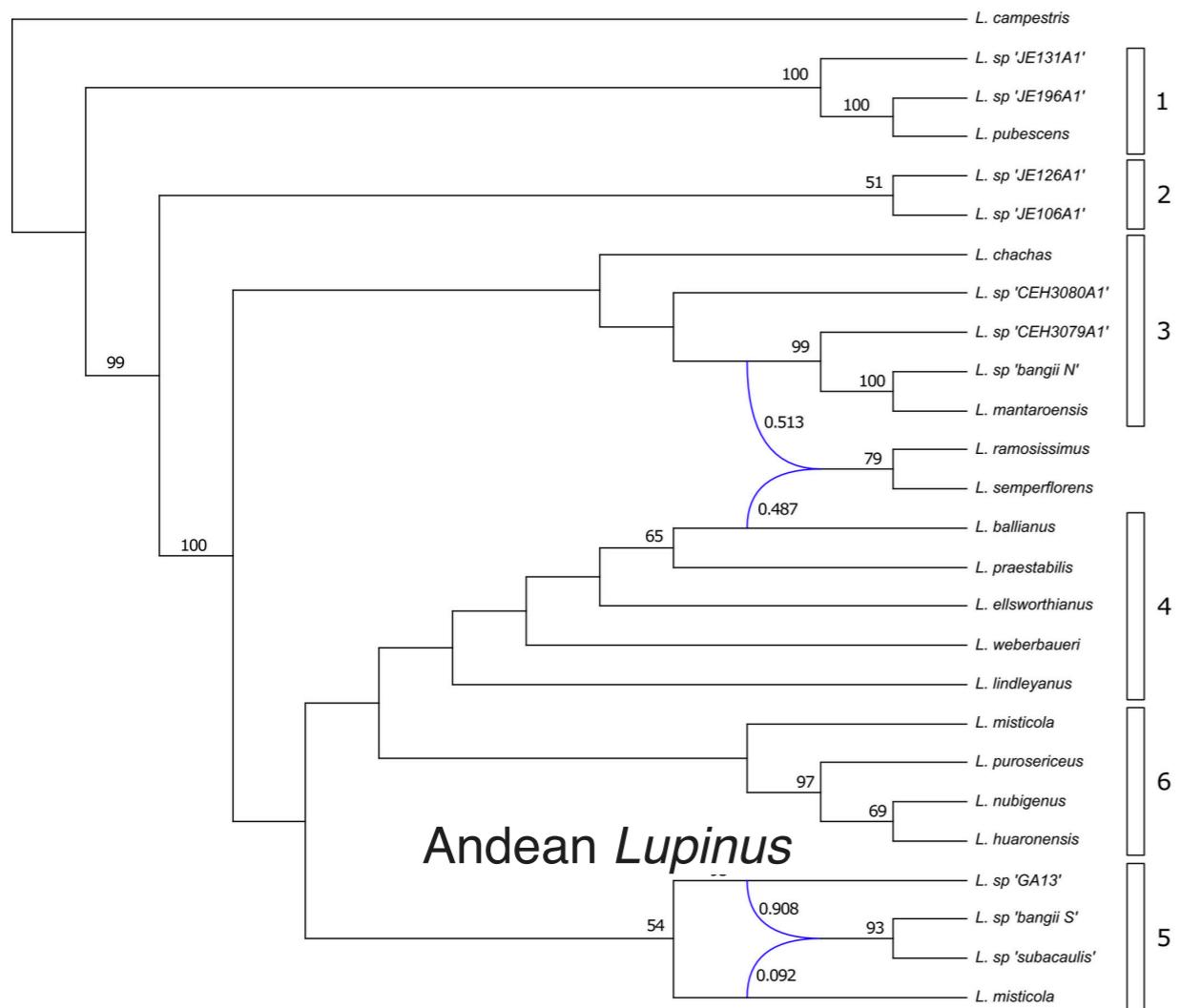


New Phytologist

Full Paper | Free Access

Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes

Bruno Nevado , Natalia Contreras-Ortiz, Colin Hughes, Dmitry A. Filatov



Current Biology

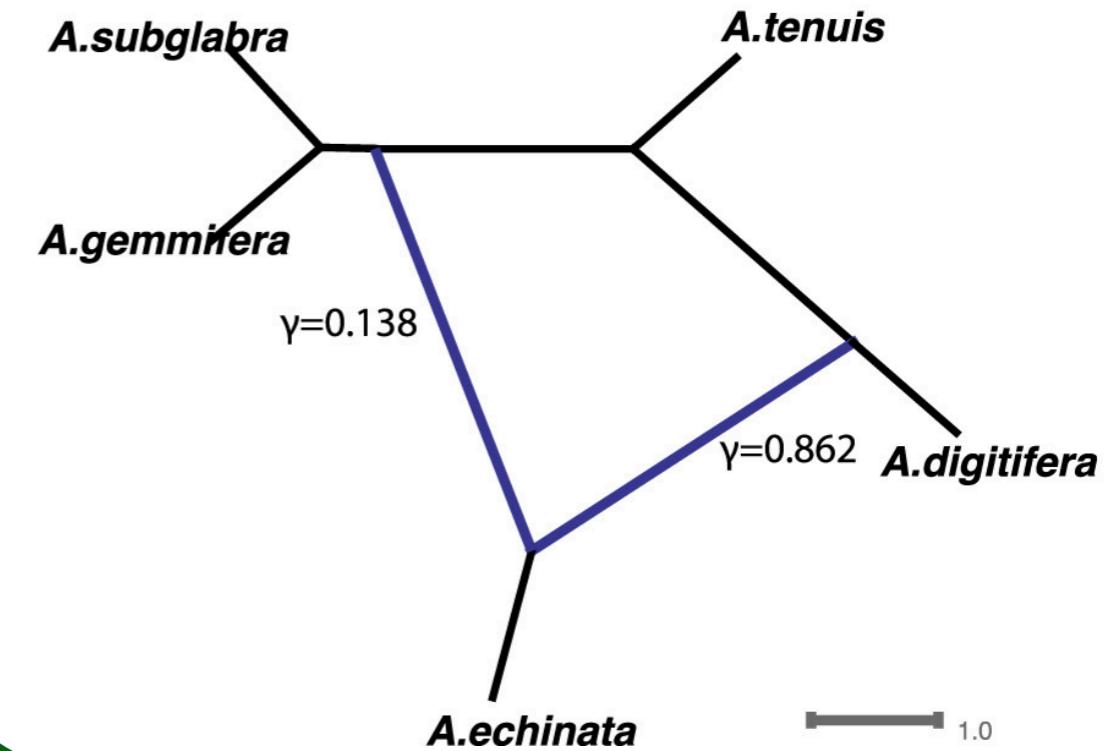
Volume 28, Issue 21, 5 November 2018, Pages 3373-3382.e5

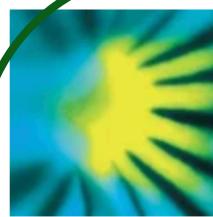


Article

The Roles of Introgression and Climate Change in the Rise to Dominance of *Acropora* Corals

Yafei Mao ^{1, 2, 3} , Evan P. Economo ² , Noriyuki Satoh ¹



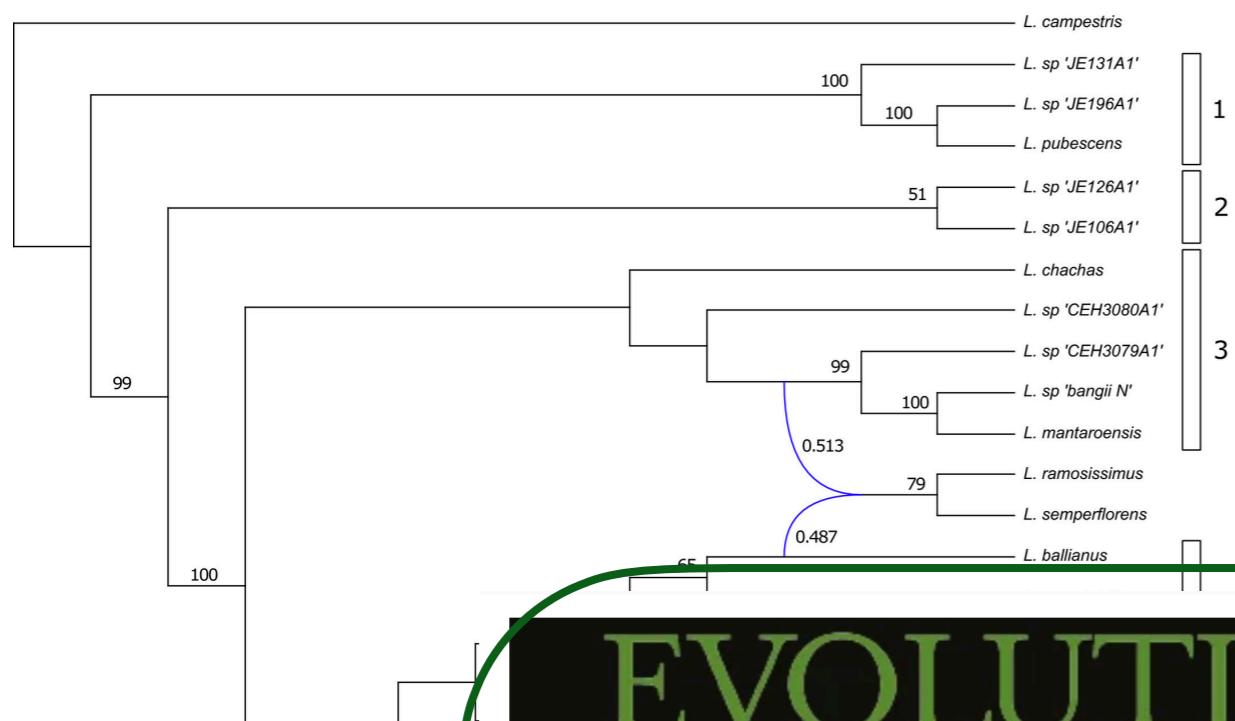


New Phytologist

Full Paper | Free Access

Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes

Bruno Nevado , Natalia Contreras-Ortiz, Colin Hughes, Dmitry A. Filatov



EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

ORIGINAL ARTICLE | Open Access |

Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean *Campanula* (Campanulaceae) clade

Andrew A. Crowl , Cody Myers, Nico Cellinese

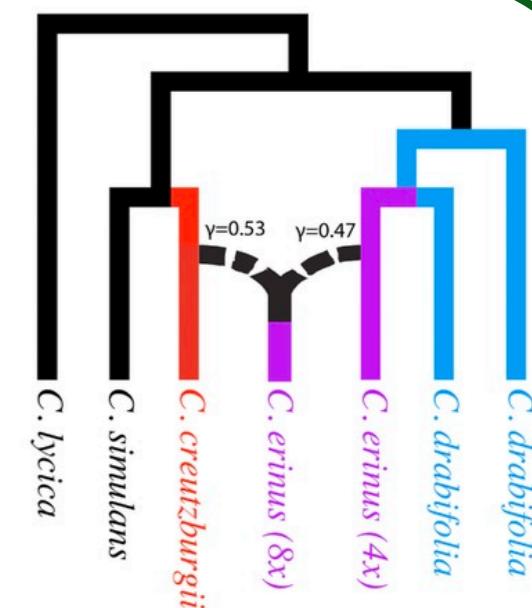
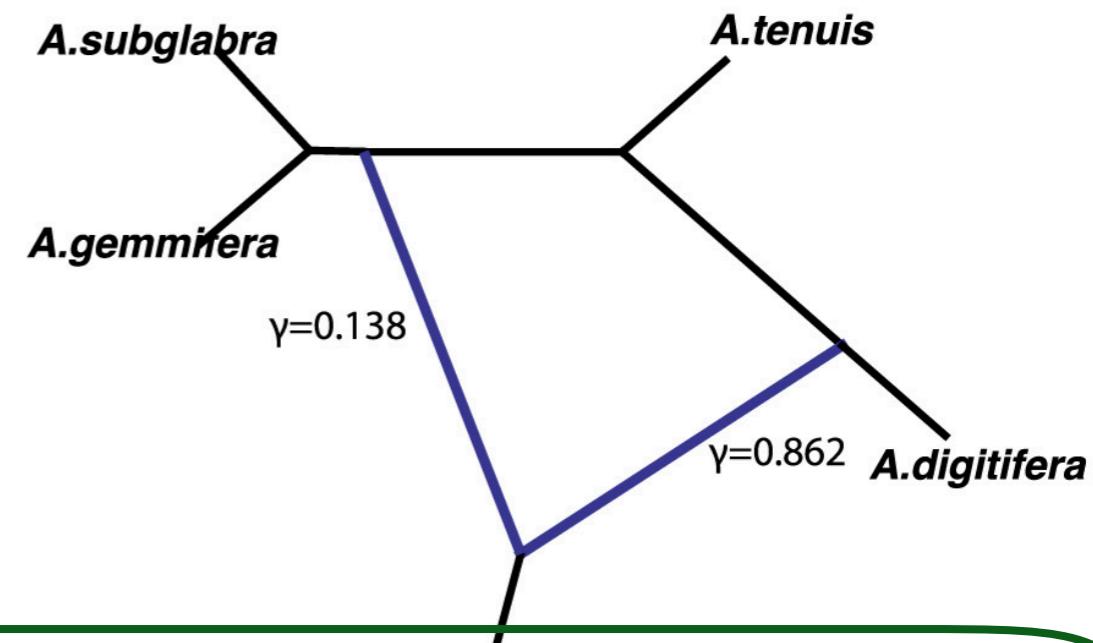
Current Biology

Volume 28, Issue 21, 5 November 2018, Pages 3373-3382.e5

Article

The Roles of Introgression and Climate Change in the Rise to Dominance of *Acropora* Corals

Yafei Mao ^{1, 2, 3} , Evan P. Economo ² , Noriyuki Satoh ¹



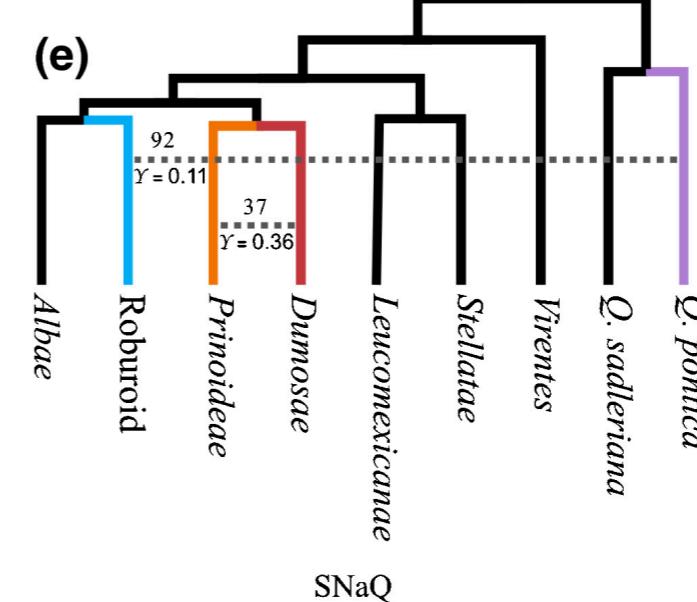


Full paper | Free Access

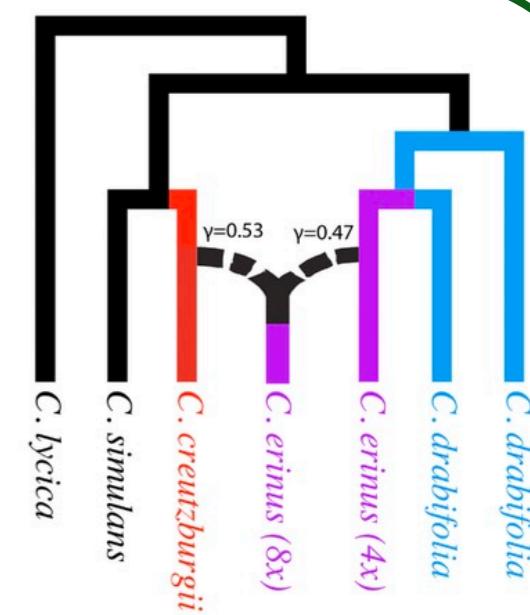
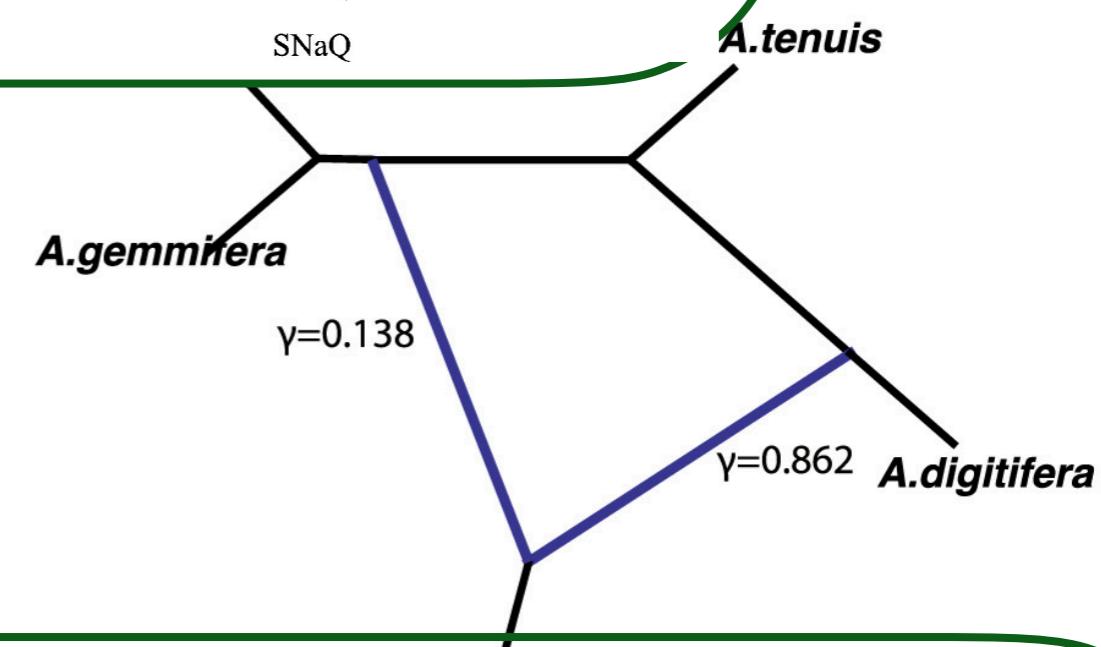
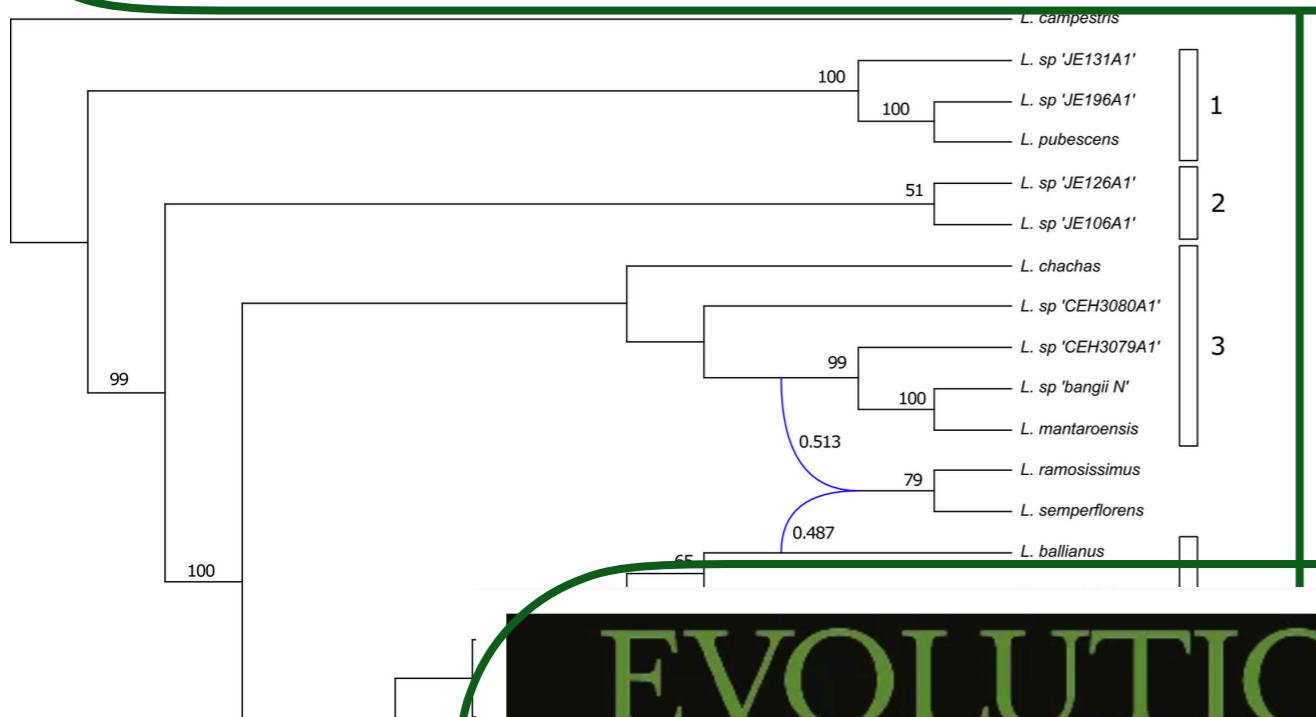
Uncovering the genomic signature of ancient introgression between white oak lineages (*Quercus*)

Andrew A. Crowl Paul S. Manos, John D. McVay, Alan R. Lemmon, Emily Moriarty Lemmon, Andrew L. Hipp ... See fewer authors

First published: 08 April 2019 | <https://doi.org/10.1111/nph.15842> | Citations: 32



imate Change in
Corals



EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

SSE
SOCIETY for the STUDY of EVOLUTION

ORIGINAL ARTICLE | Open Access |

Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean *Campanula* (Campanulaceae) clade

Andrew A. Crowl Cody Myers, Nico Cellinese



Full pa

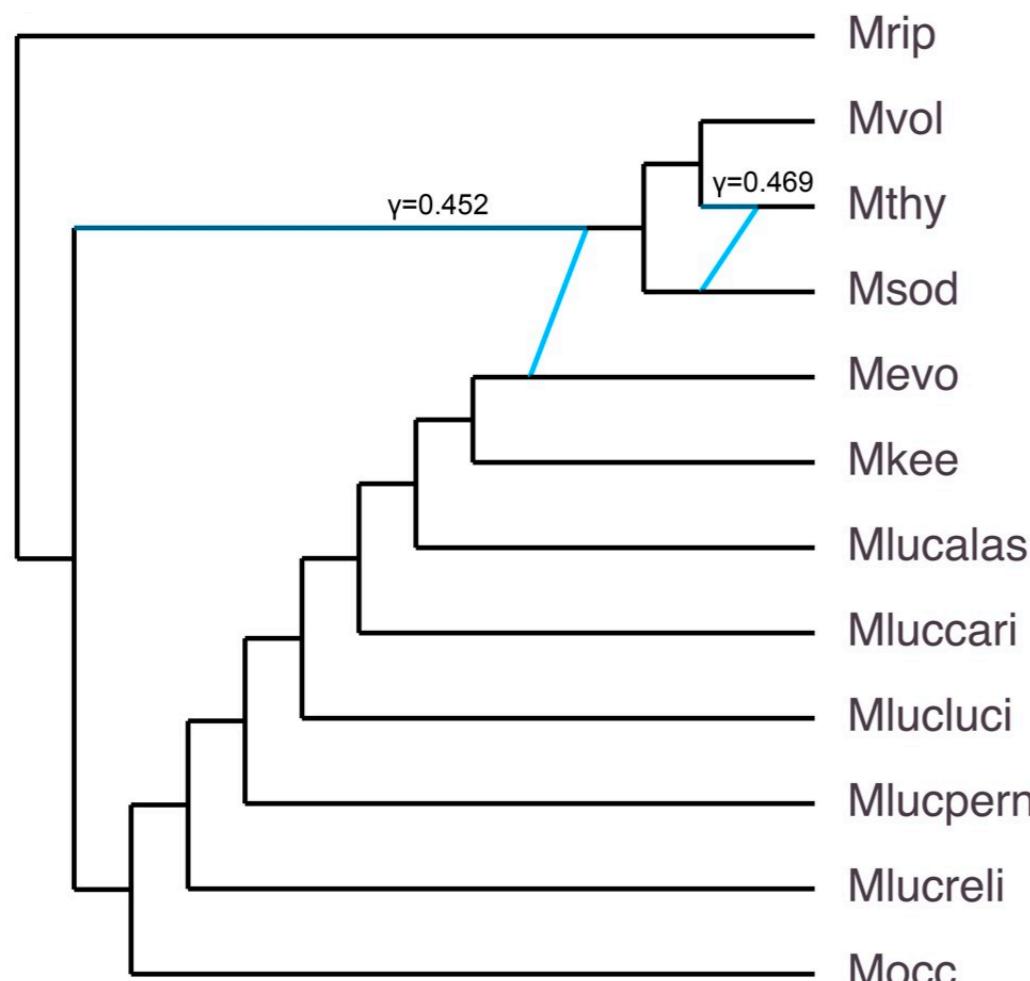
For

Evidence that *Myotis lucifugus* “Subspecies” are Five Nonsister Species, Despite Gene Flow

Ariadna E Morales , Bryan C Carstens

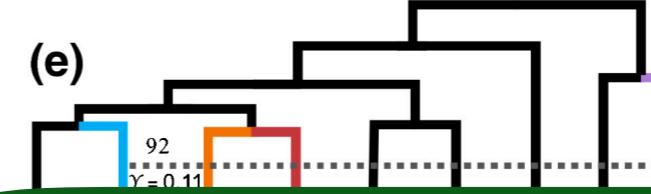
Systematic Biology, Volume 67, Issue 5, September 2018, Pages 756–769,
<https://doi.org/10.1093/sysbio/syy010>

Published: 15 February 2018 Article history ▾



Mediterranean Campanula (Campanulaceae)

Andrew A. Crowl , Cody Myers, Nico Cellinese



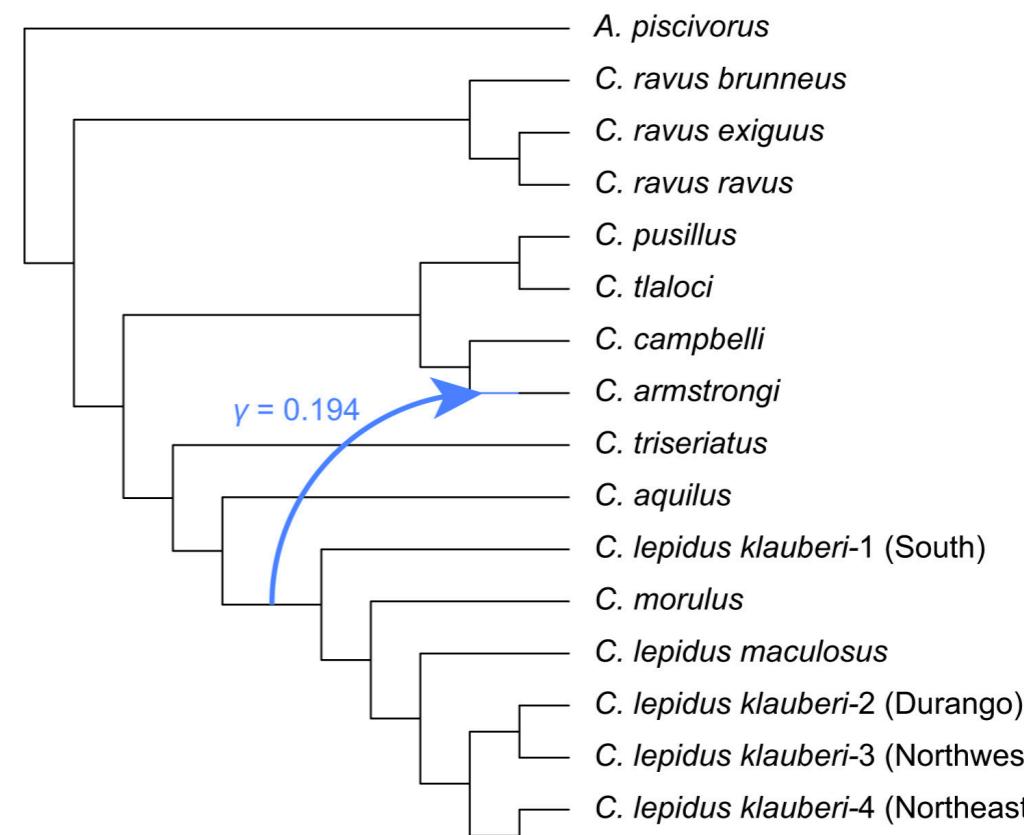
MOLECULAR ECOLOGY RESOURCES

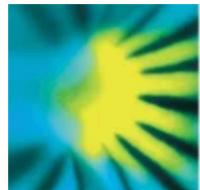
RESOURCE ARTICLE | Full Access

Cryptic diversity in the Mexican highlands: Thousands of UCE loci help illuminate phylogenetic relationships, species limits and divergence times of montane rattlesnakes (Viperidae: *Crotalus*)

Christopher Blair , Robert W. Bryson Jr, Charles W. Linkem, David Lazcano, John Klicka, John E. McCormack ... See fewer authors ^

First published: 22 November 2018 | <https://doi.org/10.1111/1755-0998.12970> | Citations: 20





Full pa

For

Evidence that *Myotis lucifugus* "Subspecies" are Five Nonsister Species, Despite Gene Flow

Ariadna E Morales , Bryan C Carstens

Systematic

https://doi.org/10.1111/nph.14920

Publ

Phylogenomic Signatures of Ancient Introgression in a Rogue Lineage of Darters (Teleostei: Percidae)

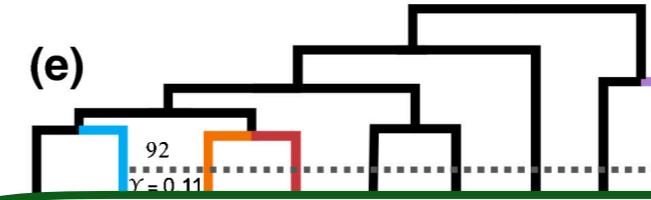
Daniel J MacGuigan , Thomas J Near

Systematic Biology, Volume 68, Issue 2, March 2019, Pages 329–346,
<https://doi.org/10.1093/sysbio/syy074>

Published: 03 December 2018 Article history ▾

Mediterranean Campanula (Campanulaceae)

Andrew A. Crowl , Cody Myers, Nico Cellinese

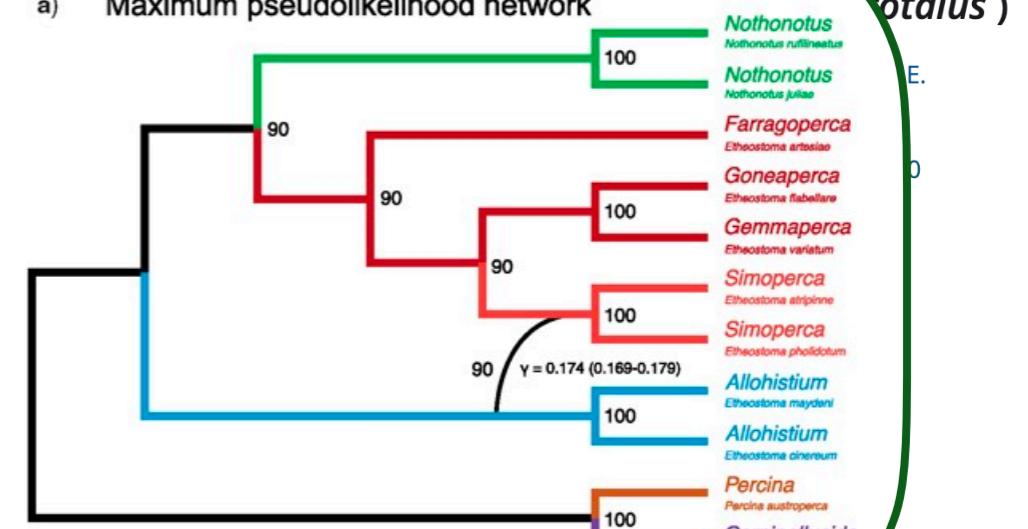


MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE | Full Access

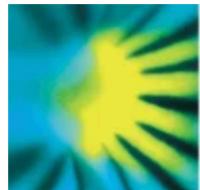
Cryptic diversity in the Mexican highlands: Thousands of UCE loci

a) Maximum pseudolikelihood network



- C. triseriatus*
- C. aquilus*
- C. lepidus klauberi-1* (South)
- C. morulus*
- C. lepidus maculosus*
- C. lepidus klauberi-2* (Durango)
- C. lepidus klauberi-3* (Northwest)
- C. lepidus klauberi-4* (Northeast)
- C. lepidus lepidus*

drabifolia
burgii
bifolia
sp. (8x)
sp. (4x)



Full paper | Article history | Supplements | Cite this article

Evidence that *Myotis lucifugus* "Subspecies" are Five Nonsister Species, Despite Gene Flow

Ariadna E Morales , Bryan C Carstens

Systematic Biology
https://doi.org/10.1093/sysbio/syy074
Published online December 2018

Phylogenomic Signatures of Ancient Introgression in a Rogue Lineage of Darters (Teleostei: Percidae)

Daniel J MacGuigan , Thomas J Near

Systematic Biology, Volume 68, Issue 2, March 2019, Pages 329–346,
<https://doi.org/10.1093/sysbio/syy074>

Published: 03 December 2018 Article history ▾

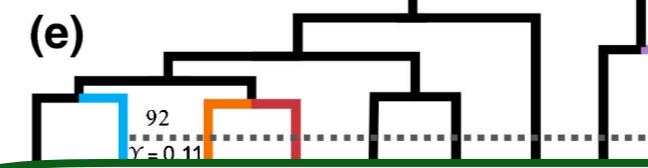
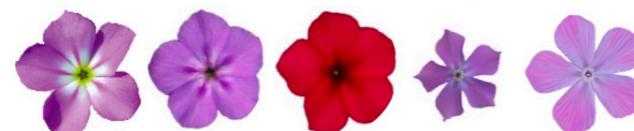
MOLECULAR ECOLOGY

Original Article |

Genomic evidence of gene flow during reinforcement in Texas *Phlox*

Federico Roda , Fábio K. Mendes, Matthew W. Hahn, Robin Hopkins

First published: 31 January 2017 | <https://doi.org/10.1111/mec.14041> | Citations: 15



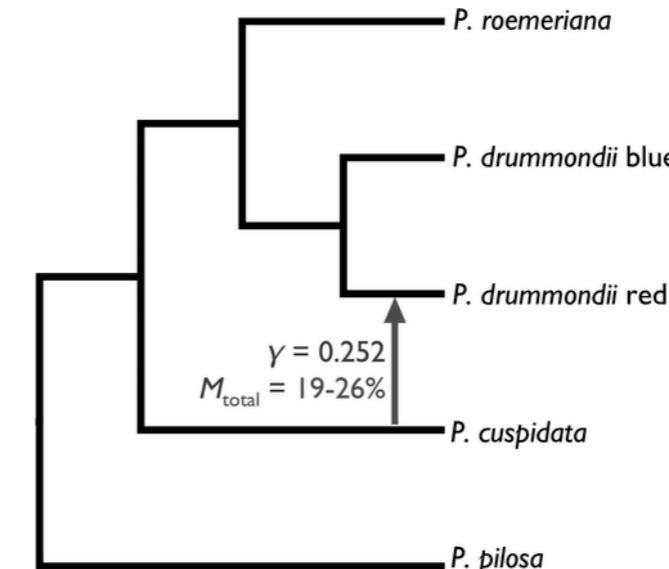
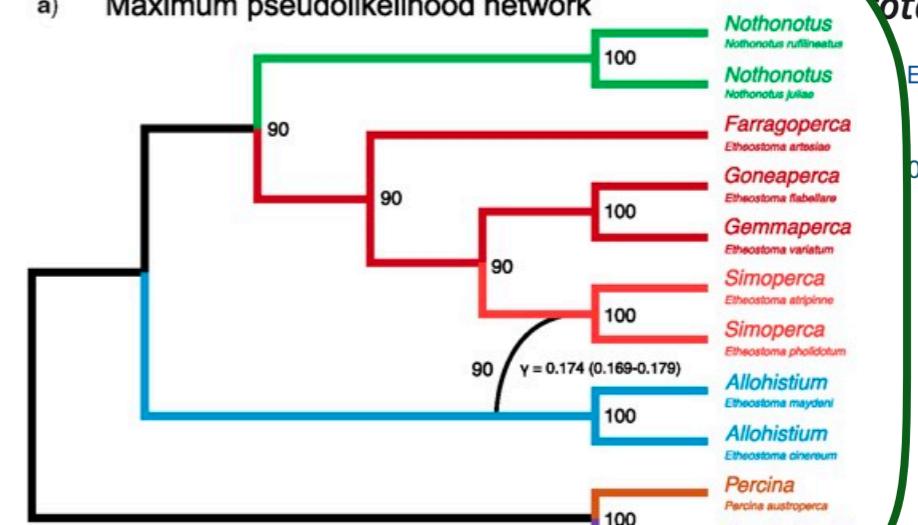
MOLECULAR ECOLOGY RESOURCES

RESOURCE ARTICLE |

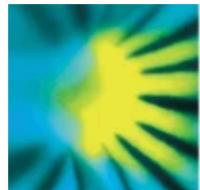
Cryptic diversity in the Mexican highlands: Thousands of UCE loci

in *Etheostoma* (Teleostei: Percidae) reveal cryptic lineages and species-level support for *Etheostoma* (*Nothonotus*) *totalis*)

a) Maximum pseudolikelihood network



(Phlox (4x))
drabifolia
bifolia



Full paper

Free

Evidence that *Myotis lucifugus* "Subspecies" Are Nonsister Species, Despite Gene Flow

Ariadna E Morales ✉, Bryan C Carstens

Systematic Biology

https://doi.org/10.1093/sysbio/syz073

Published

Phylogenomic Signature in a Rogue Lineage of *Myotis*

FREE

Daniel J MacGuigan ✉, Thomas J Near

Systematic Biology, Volume 68, Issue 3, May 2019, Pages 462–478,

https://doi.org/10.1093/sysbio/syy073

Published: 03 December 2018 Article history ▾

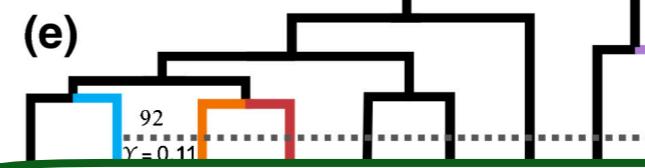
MOLECULAR

Original Article | Full Access

Genomic evidence of gene flow between *Myotis* subspecies in *Phlox*

Federico Roda ✉, Fábio K. Mendes, Matthew W. Hahn, Robin Hopkins

First published: 31 January 2017 | https://doi.org/10.1111/mec.14041 | Citations: 15



MOLECULAR ECOLOGY RESOURCES

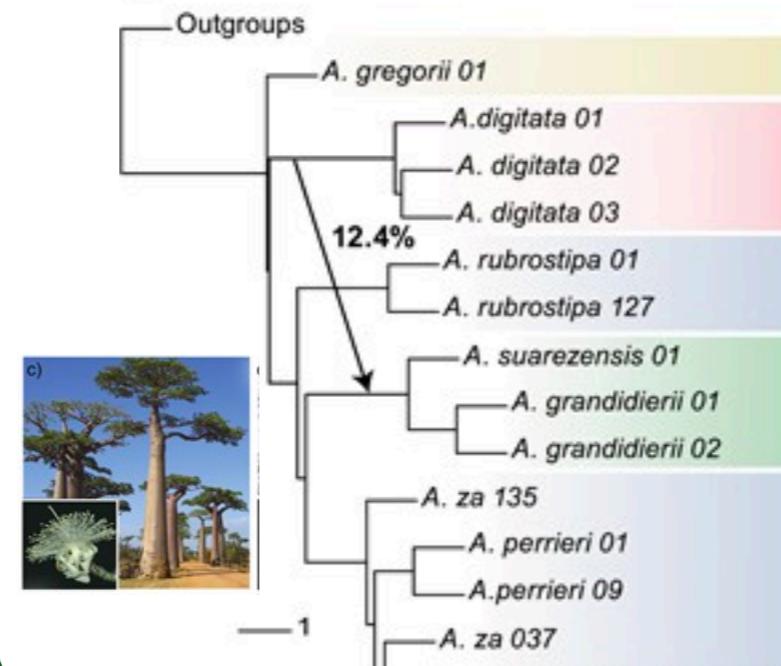
Reticulate Evolution Helps Explain Apparent Homoplasy in Floral Biology and Pollination in Baobabs (*Adansonia*; Bombacoideae; Malvaceae)

Nisa Karimi ✉, Corinne E Grover, Joseph P Gallagher, Jonathan F Wendel, Cécile Ané, David A Baum ✉

Systematic Biology, Volume 69, Issue 3, May 2020, Pages 462–478,

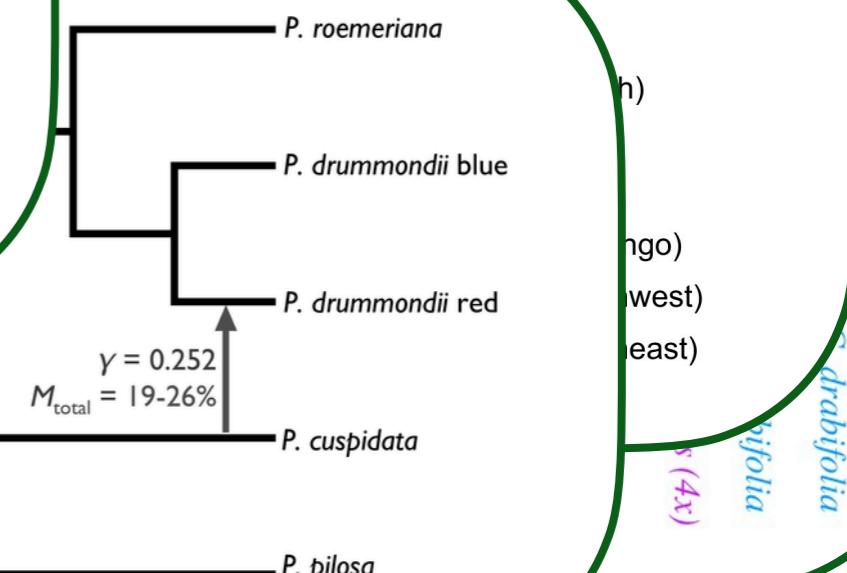
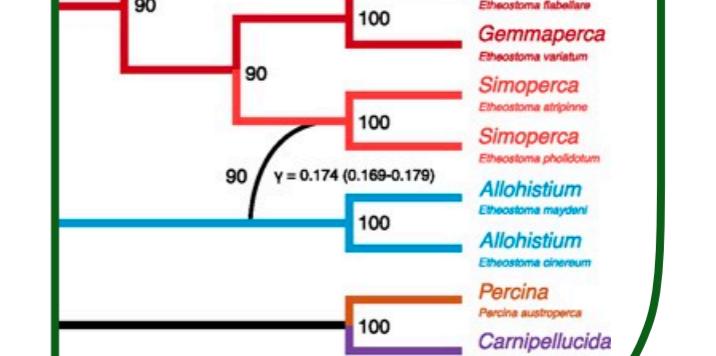
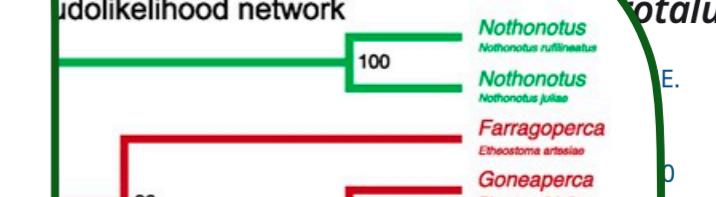
https://doi.org/10.1093/sysbio/syz073

Published: 06 November 2019 Article history ▾



Mexican Highlands: Thousands of UCE Loci Reveal Reticulate Evolution in *Adansonia* (Malvaceae)

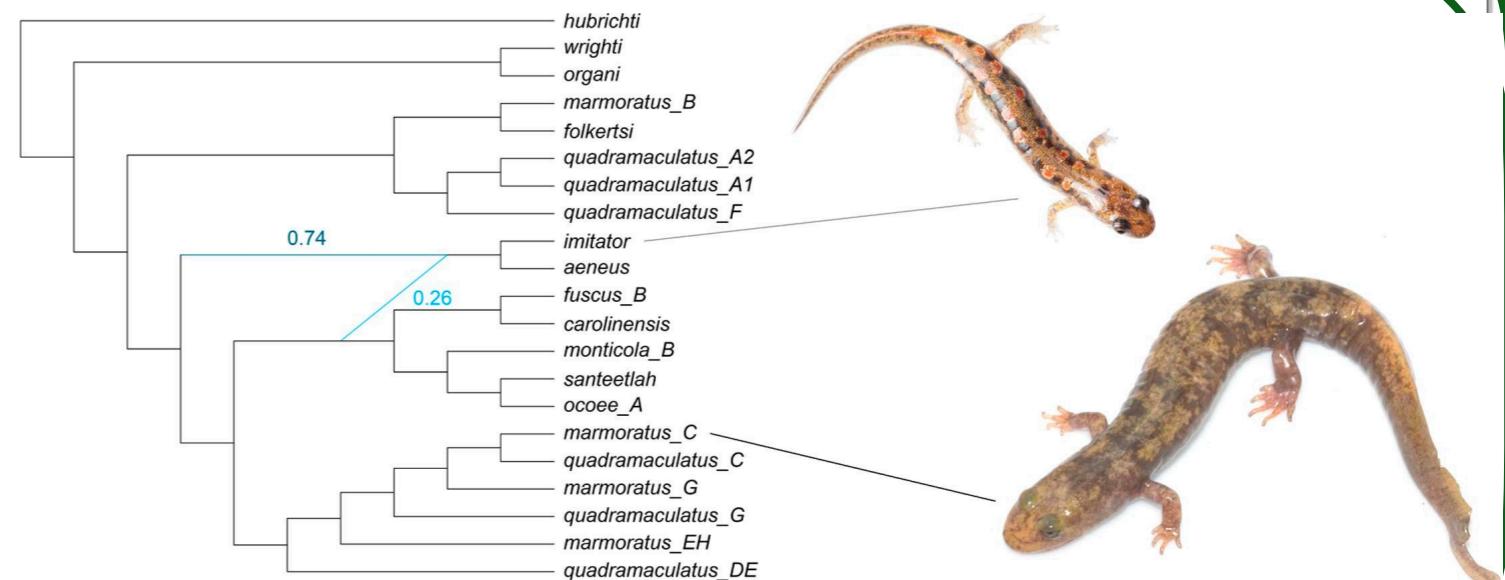
Maximum likelihood network





Phylogenomic data reveal reticulation and incongruence among mitochondrial candidate species in Dusky Salamanders (*Desmognathus*)

R. Alexander Pyron ^{a, b} Kyle A. O'Connell ^{a, b, c}, Emily Moriarty Lemmon ^d, Alan R. Lemmon ^e, David A. Beamer ^f



Daniel J MacGuigan Thomas J N

Systematic Biology, Volume 68, Issue 4, December 2019
<https://doi.org/10.1093/sysbio/syy050>

Published: 03 December 2018 Accepted: 10 October 2019

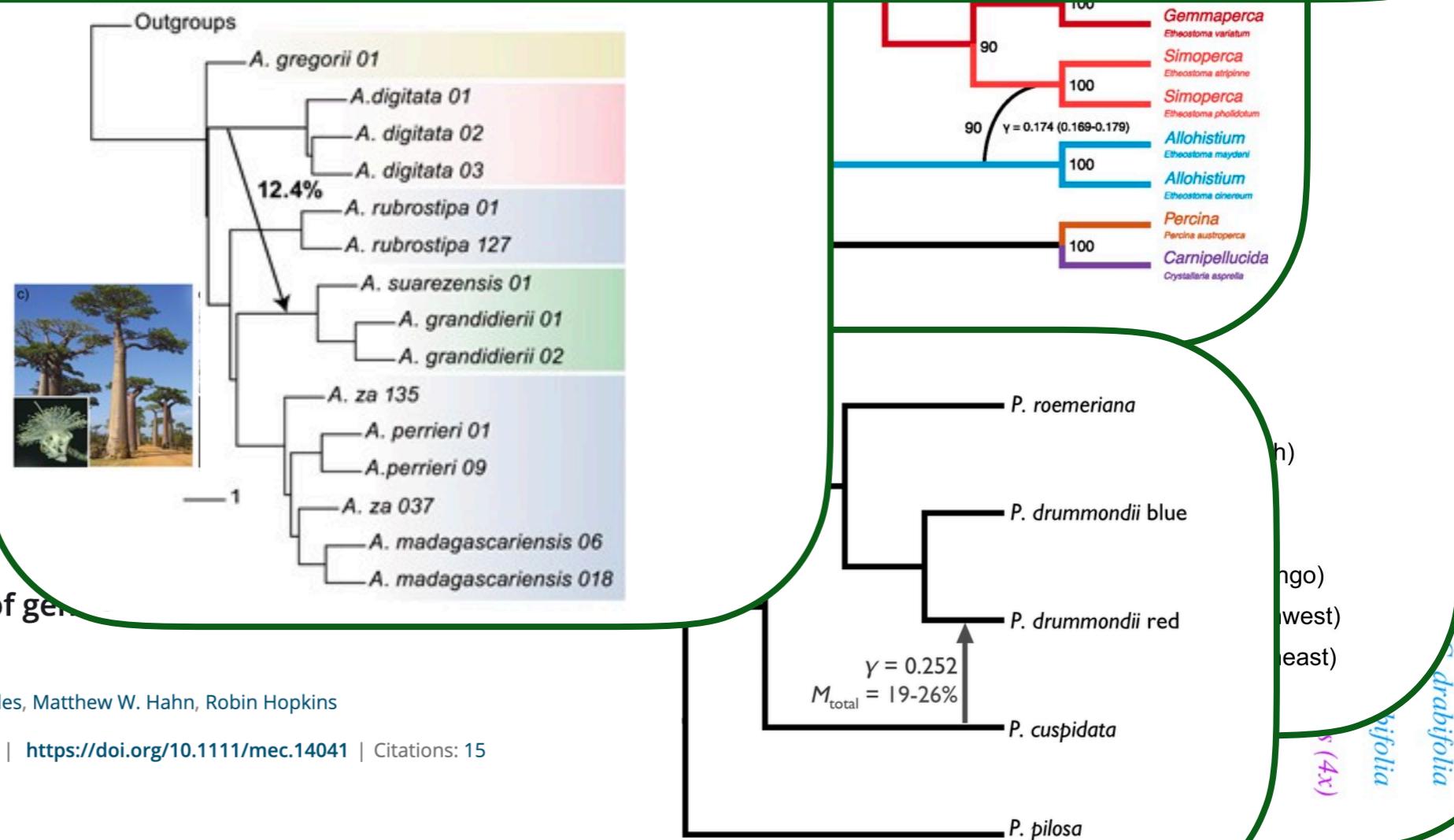


Original Article | Full Access

Genomic evidence of gene flow between *Phlox* species

Federico Roda Fábio K. Mendes, Matthew W. Hahn, Robin Hopkins

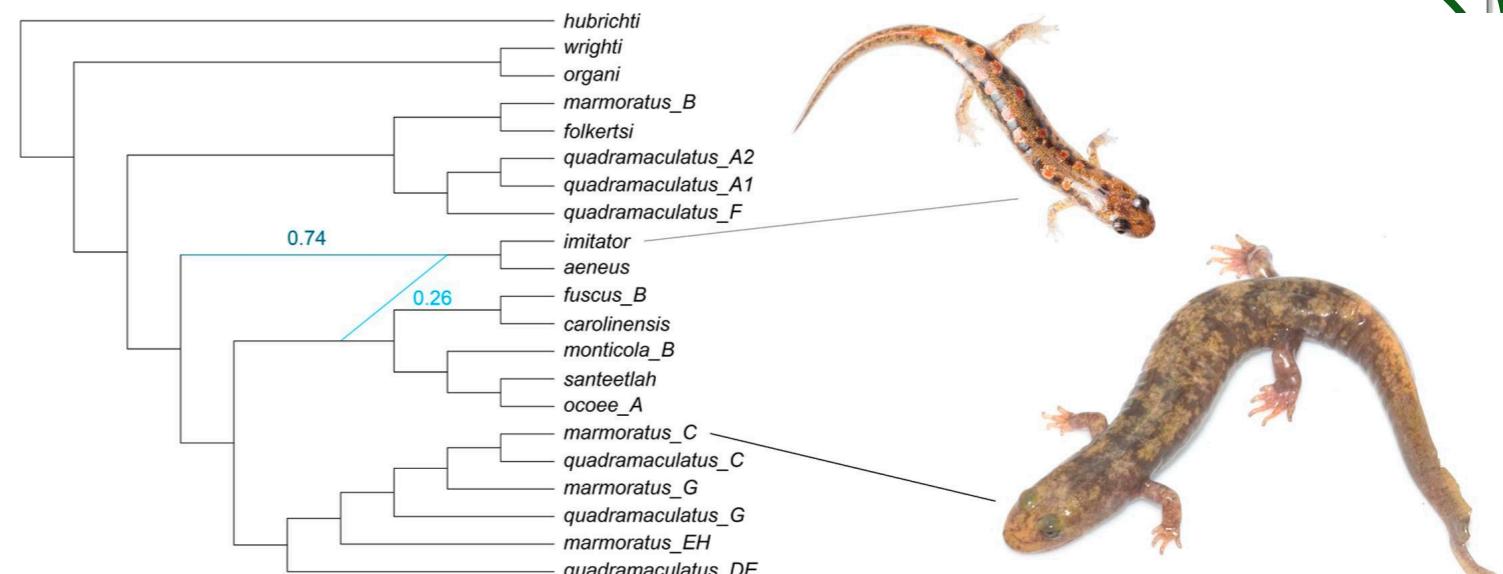
First published: 31 January 2017 | <https://doi.org/10.1111/mec.14041> | Citations: 15





Phylogenomic data reveal reticulation and incongruence among mitochondrial candidate species in Dusky Salamanders (*Desmognathus*)

R. Alexander Pyron ^{a, b} , Kyle A. O'Connell ^{a, b, c}, Emily Moriarty Lemmon ^d, Alan R. Lemmon ^e, David A. Beamer ^f



Daniel J MacGuigan , Thomas J Noe

Systematic Biology, Volume 68, Issue

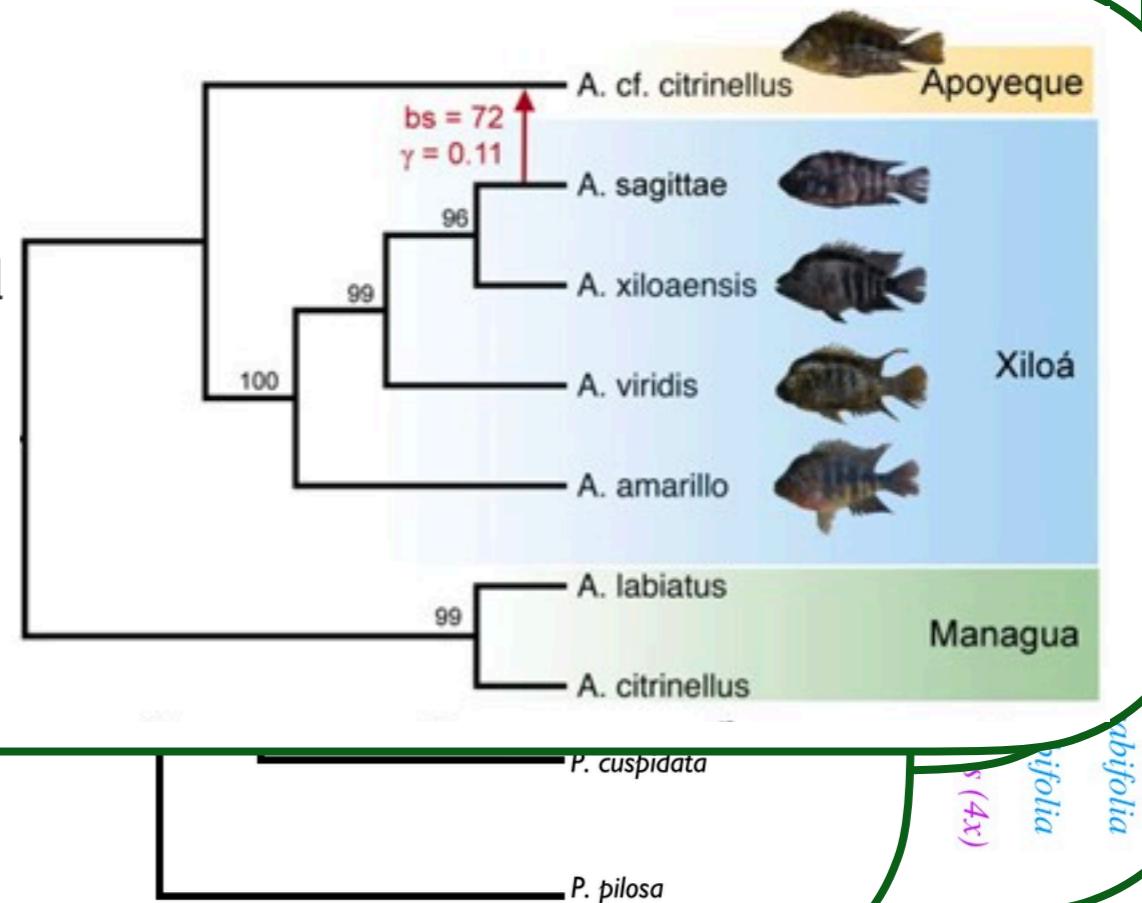
http://

Implementing Large Genomic Single Nucleotide Polymorphism Data Sets in Phylogenetic Network Reconstructions: A Case Study of Particularly Rapid Radiations of Cichlid Fish

Melisa Olave ✉, Axel Meyer ✉

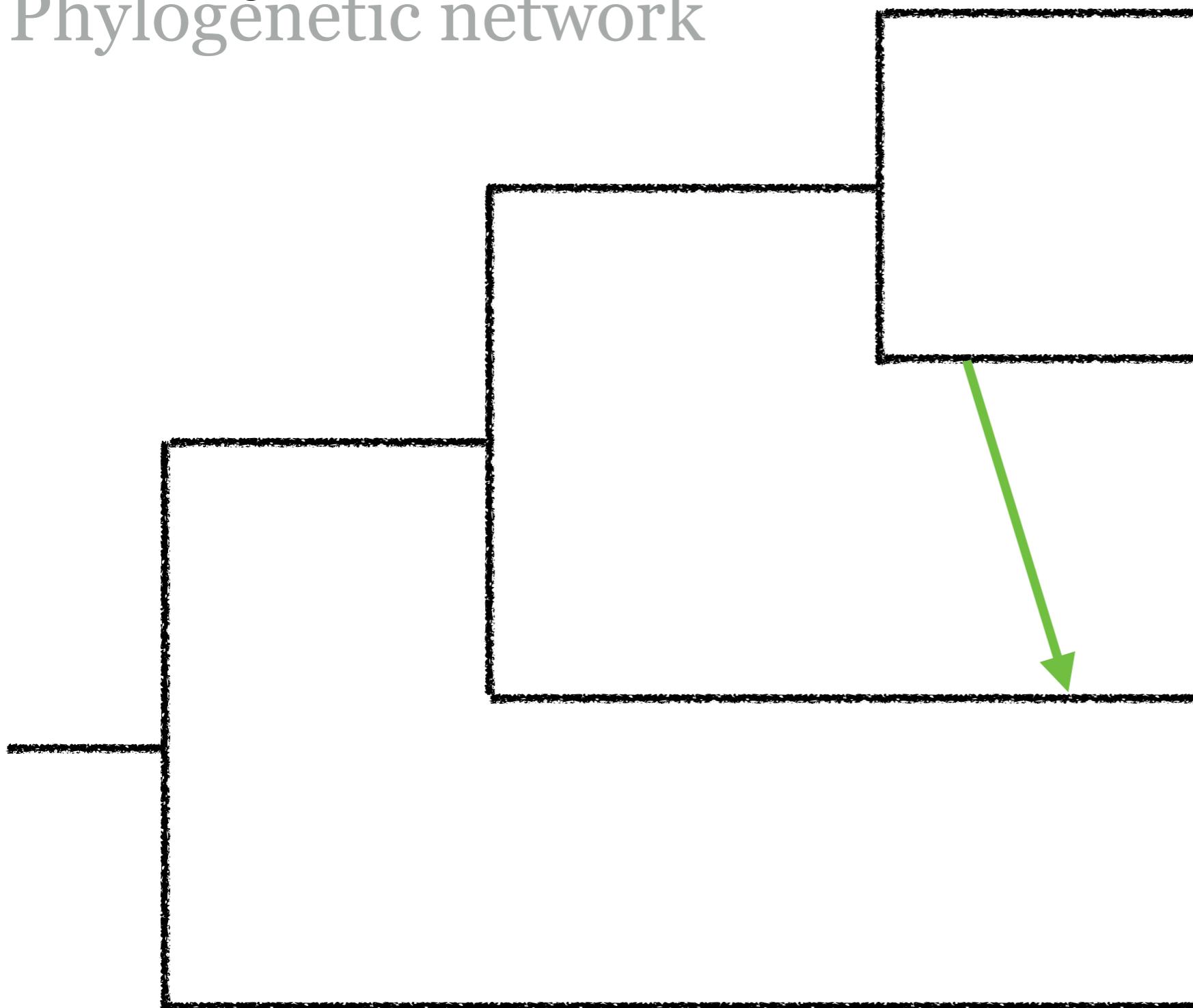
Systematic Biology, Volume 69, Issue 5, September 2020, Pages 848–862,
<https://doi.org/10.1093/sysbio/syaa005>

Published: 03 February 2020 Article history



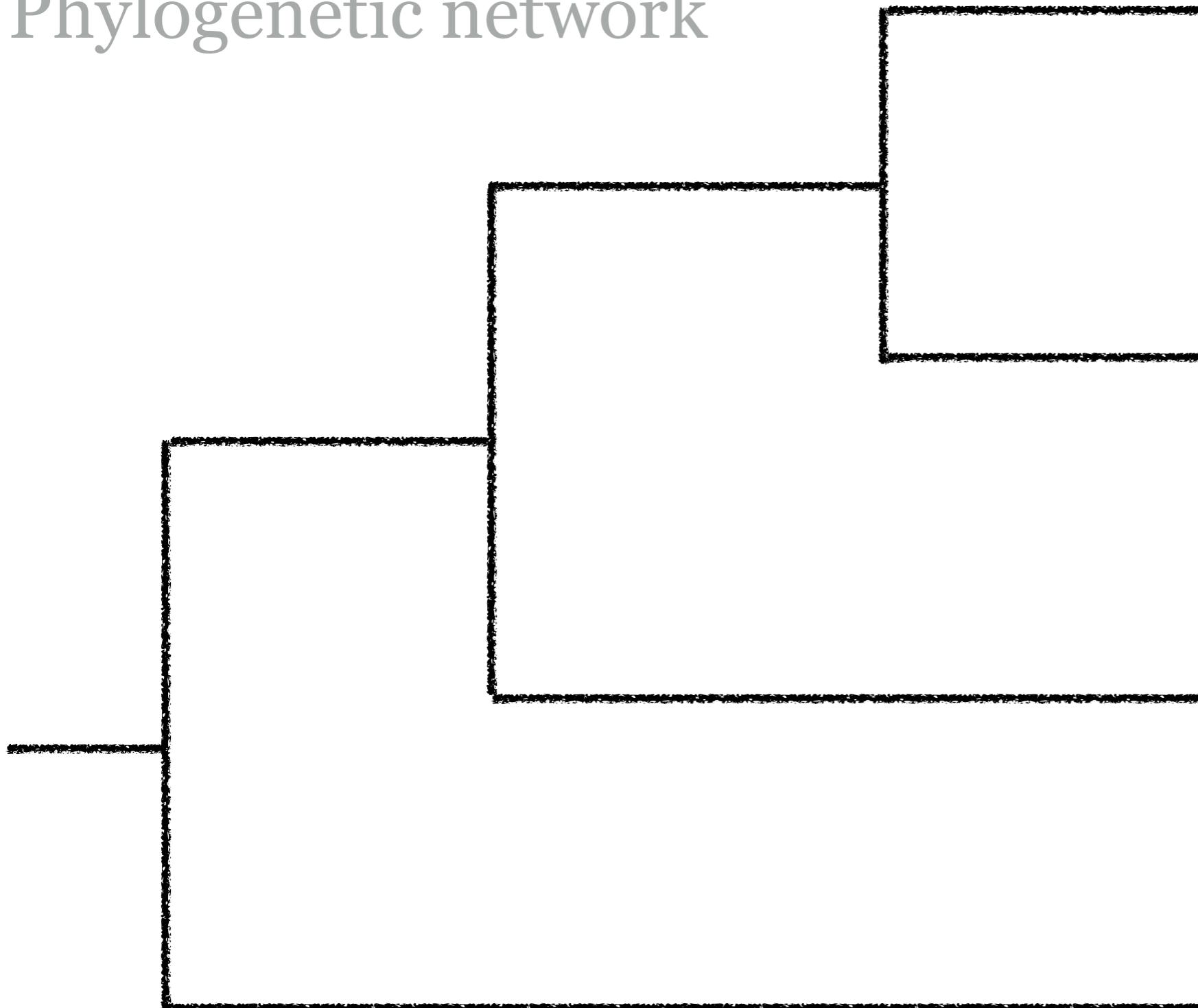
Why?

Phylogenetic network

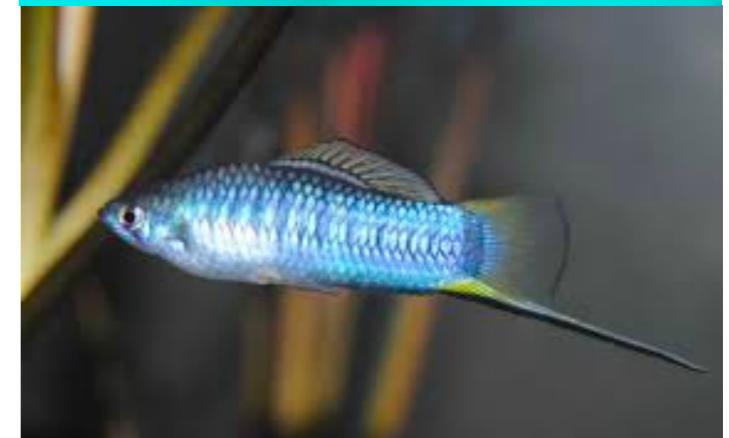


Why?

Phylogenetic network



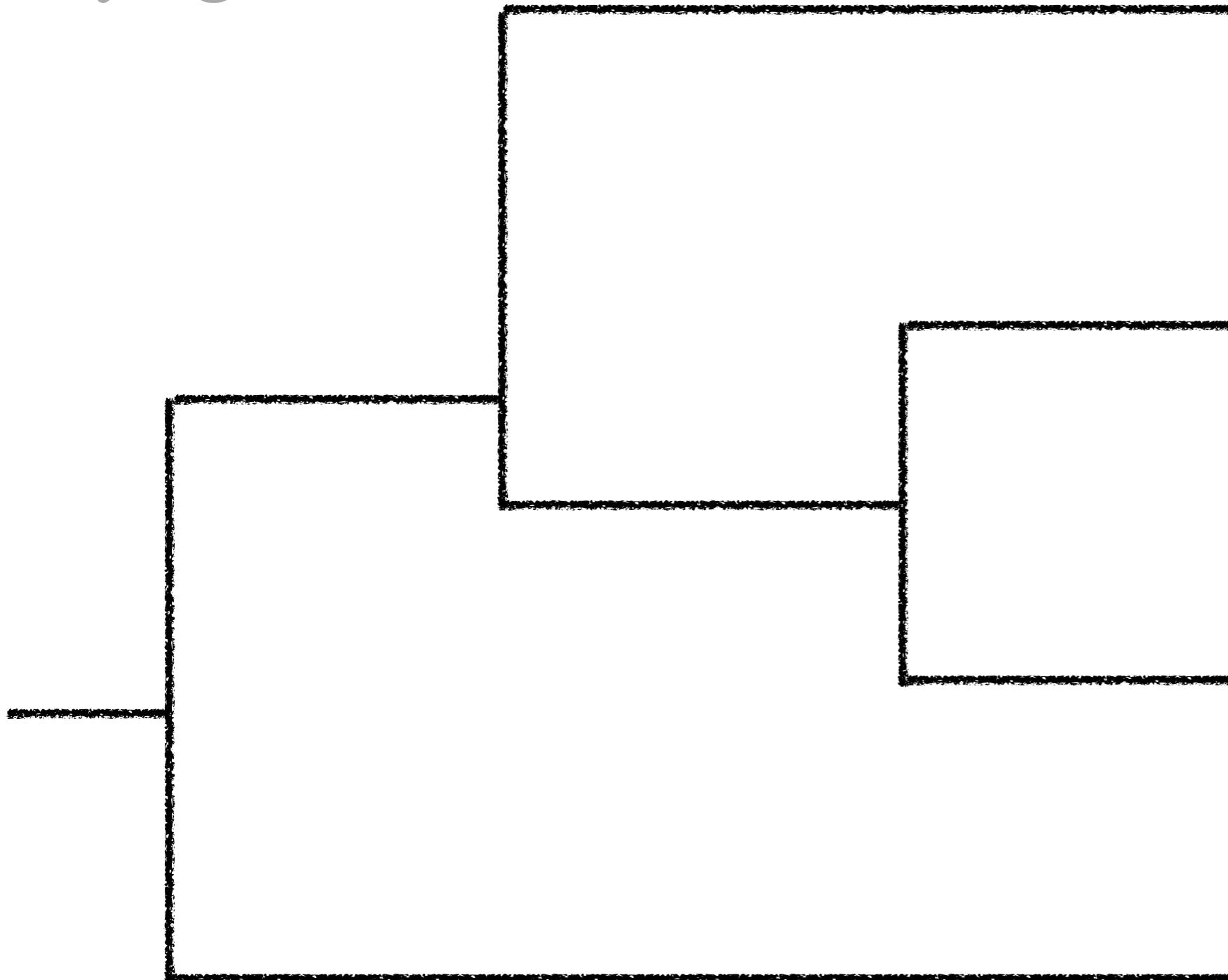
Main tree



Why?

Phylogenetic network

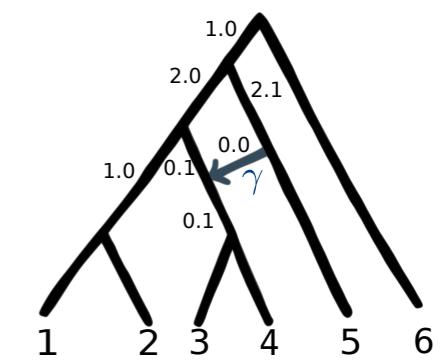
Ignore gene flow
=>Wrong tree!



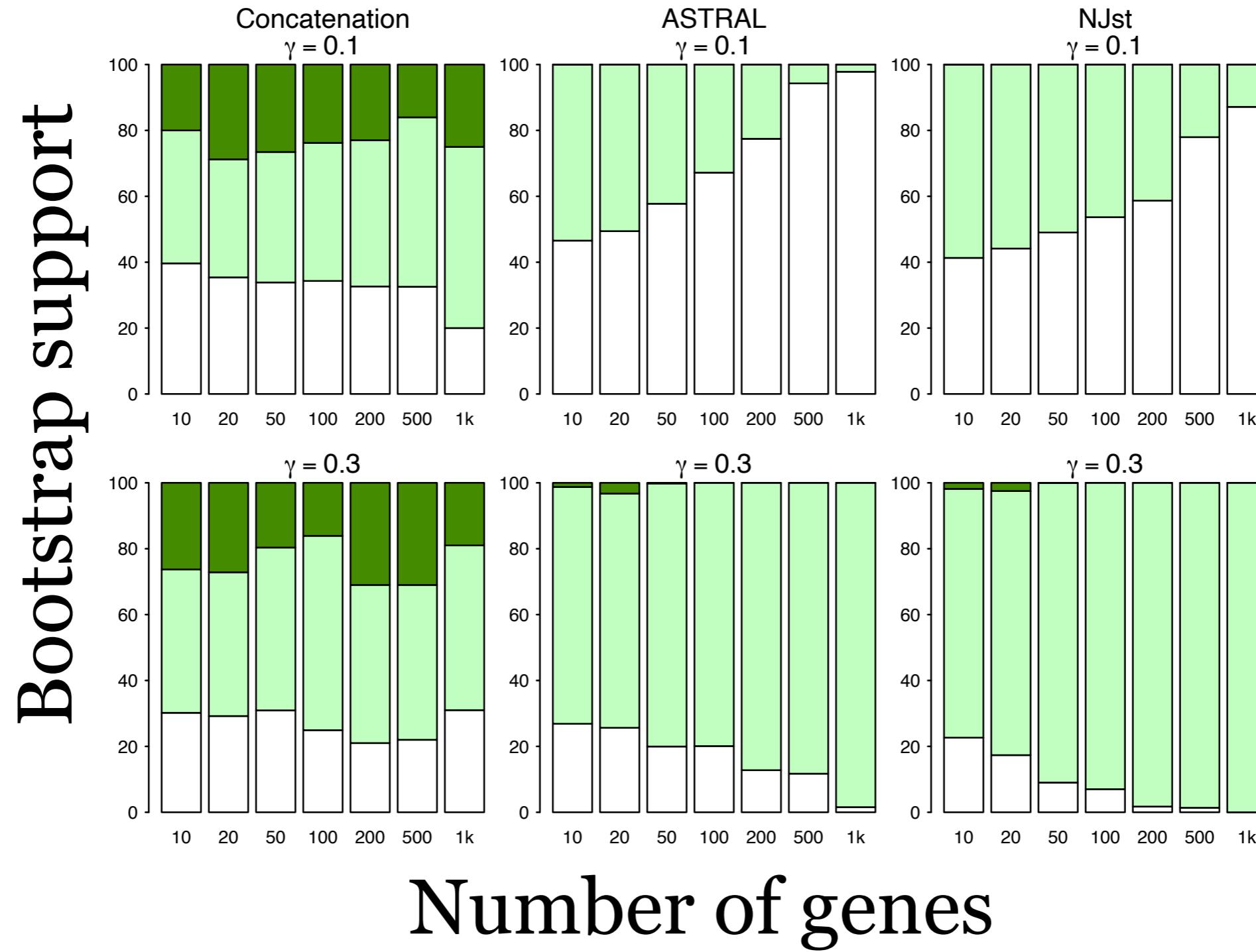
Why?

Phylogenetic network

Coalescent tree methods
not robust to gene flow



White:
true tree

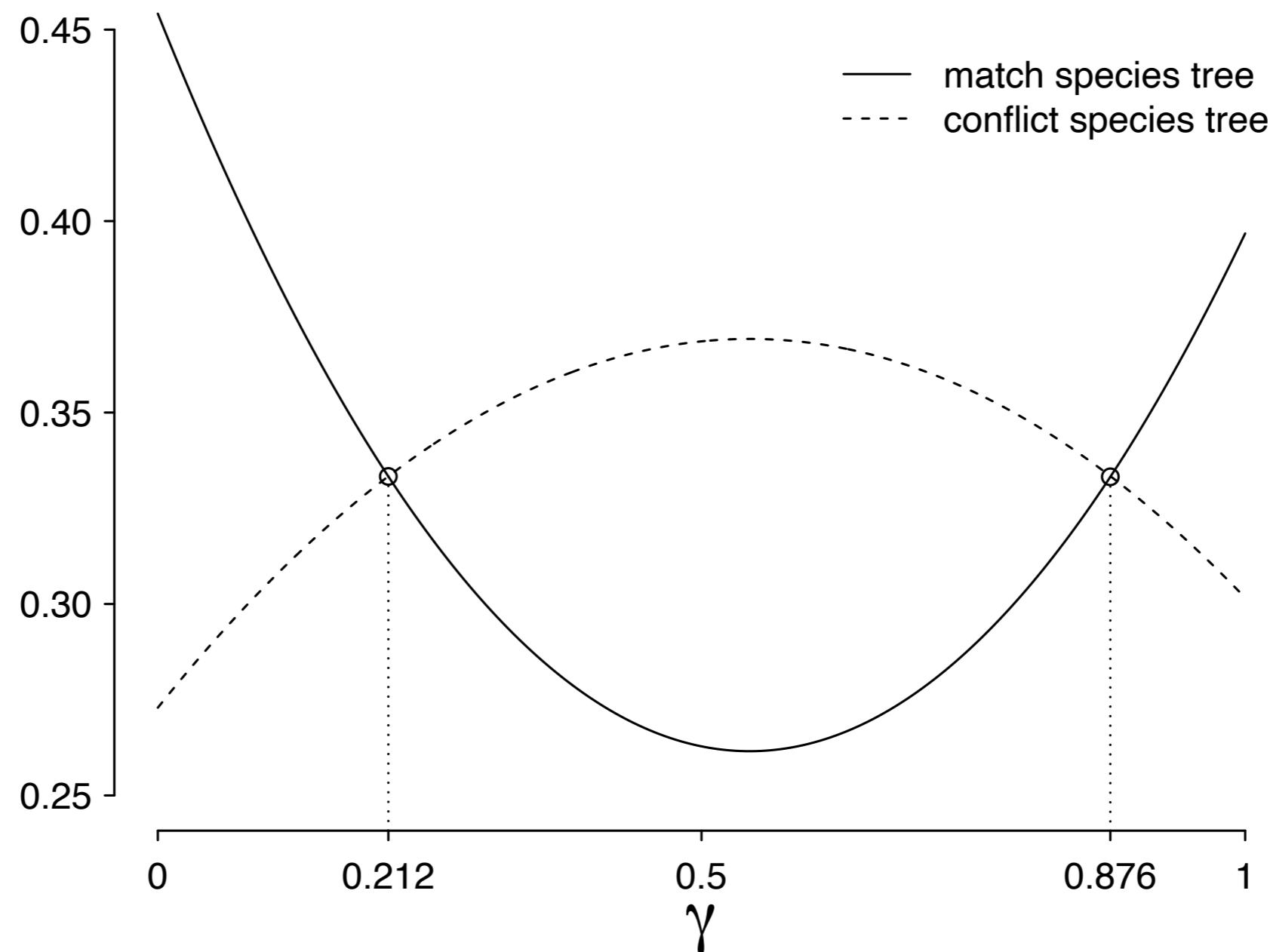


(S.-L., Yang, Ané, 2016, Syst Bio)

ASTRAL (Mirarab et al, 2014)
NJst (Liu&Yu, 2011)

Why? Phylogenetic network

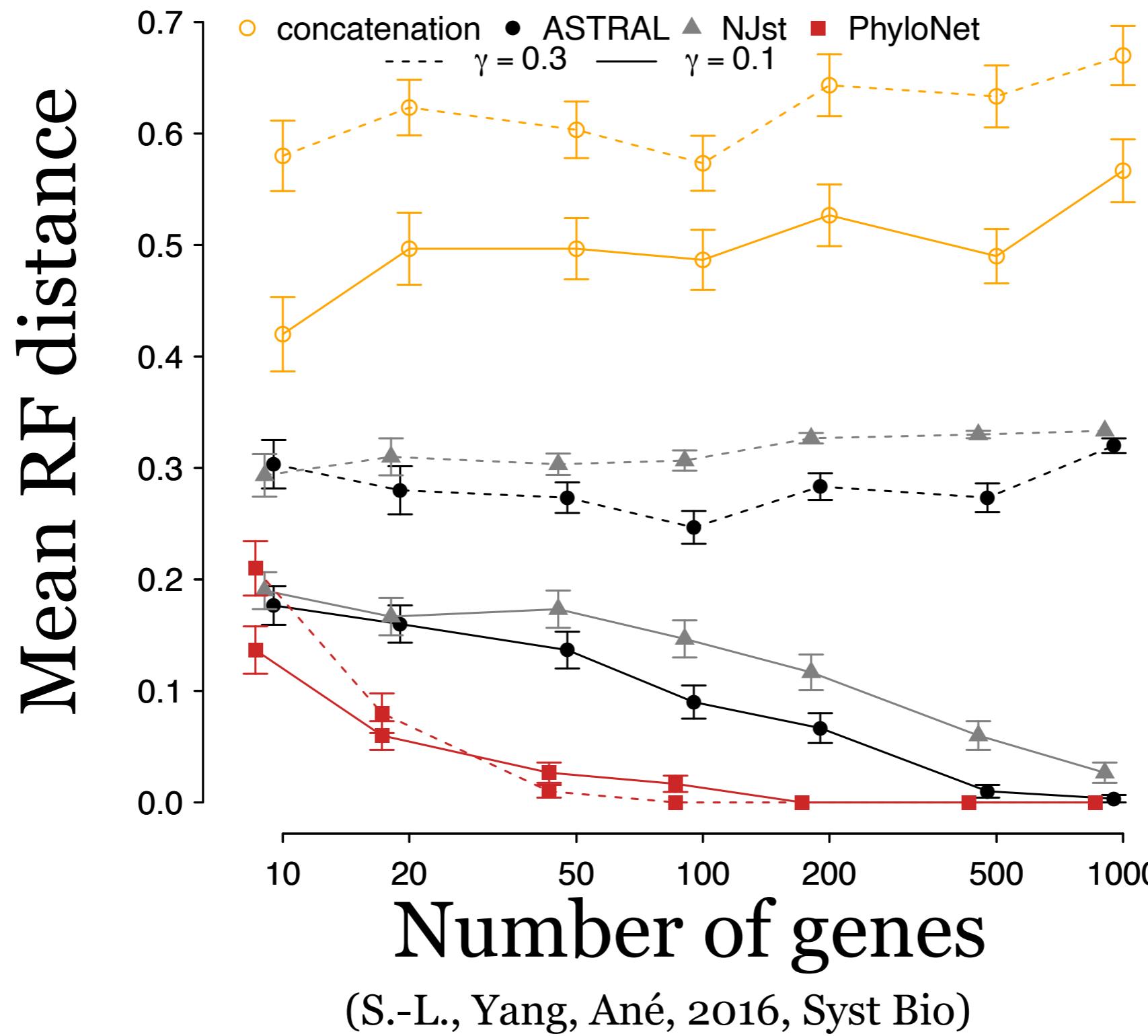
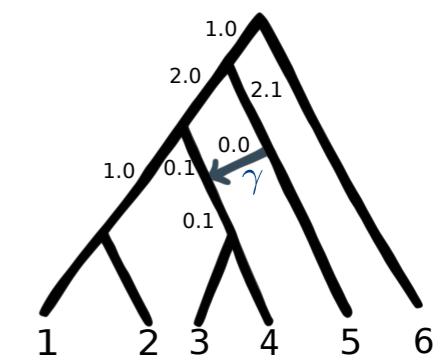
Anomaly zone with
gene flow



Why?

Phylogenetic network

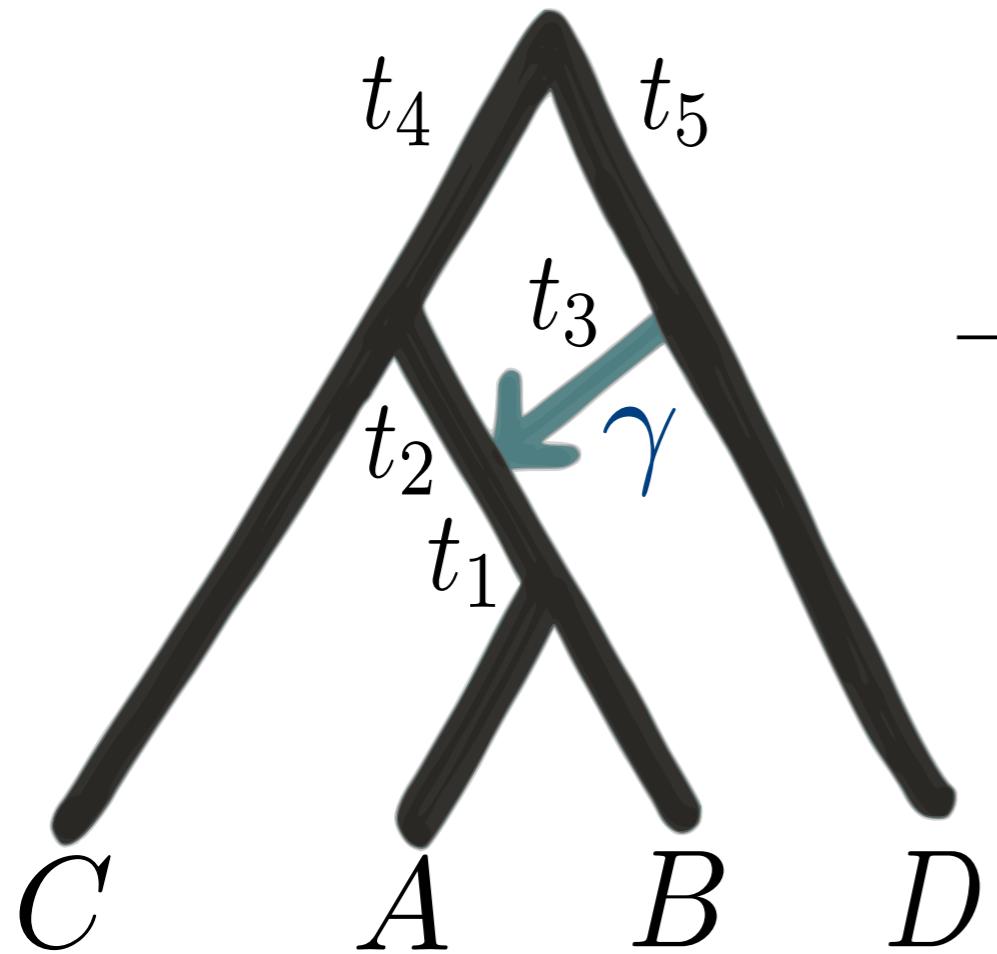
Coalescent tree methods
not robust to gene flow



Why?

Phylogenetic network

Anomalous unrooted gene trees with gene flow



Frequency among gene trees

| Quartet | $\gamma = 0.0$ | $\gamma = 0.1$ | $\gamma = 0.3$ |
|---------|----------------|----------------|----------------|
| $AB CD$ | 0.347 | 0.298 | 0.260 |
| $CA BD$ | 0.327 | 0.351 | 0.370 |
| $CB AD$ | 0.327 | 0.351 | 0.370 |

$$t_1 = t_2 = 0.01, t_3 = t_4 = t_5 = 1$$

- **ILS**: no AUGT on 4 taxa (Degnan, 2013)
- **ILS+HGT**: AUGT on 4 taxa (S.-L., Yang, Ané, 2016, Syst Bio)

So far...

- Networks are good
- Explicit networks are better
- If you ignore gene flow, you can estimate the wrong tree



<https://solislemuslab.github.io/>



mstdn.social/@solislemuslab



crsl4



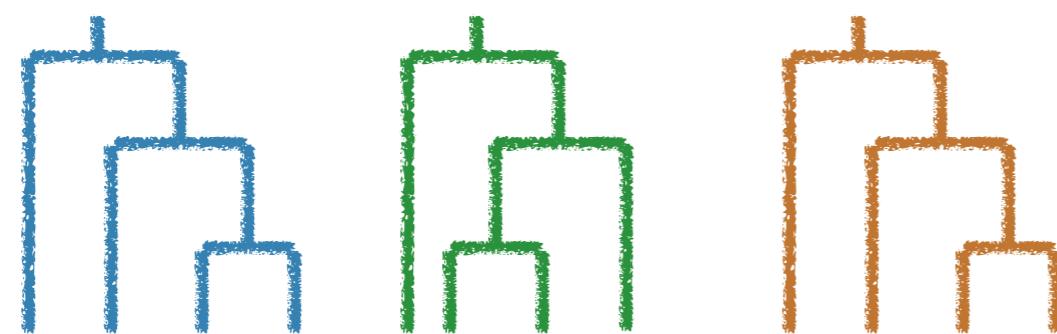
@thestatistician

How?

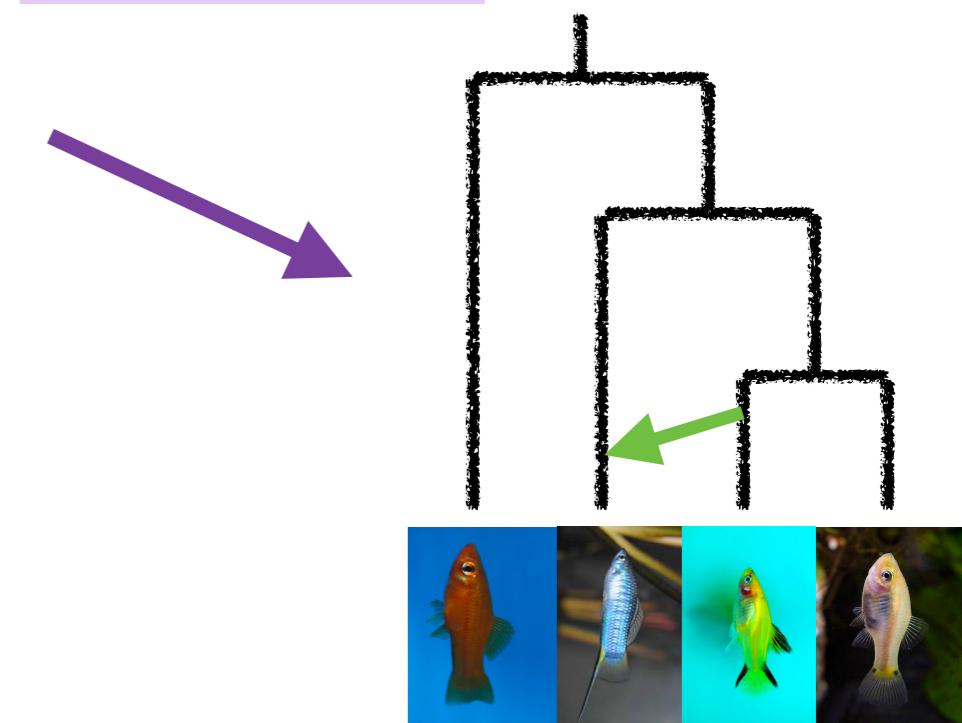
Phylogenetic network



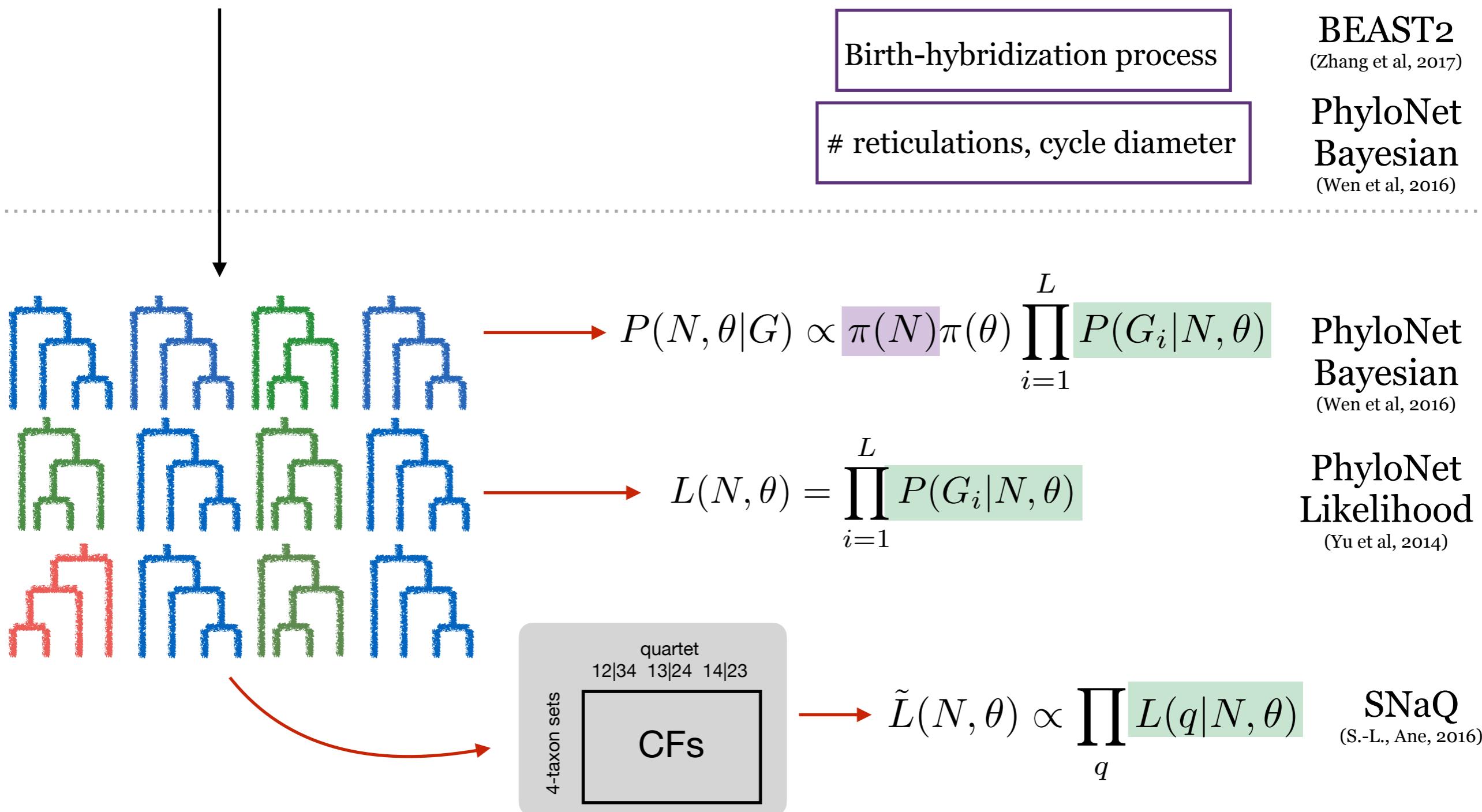
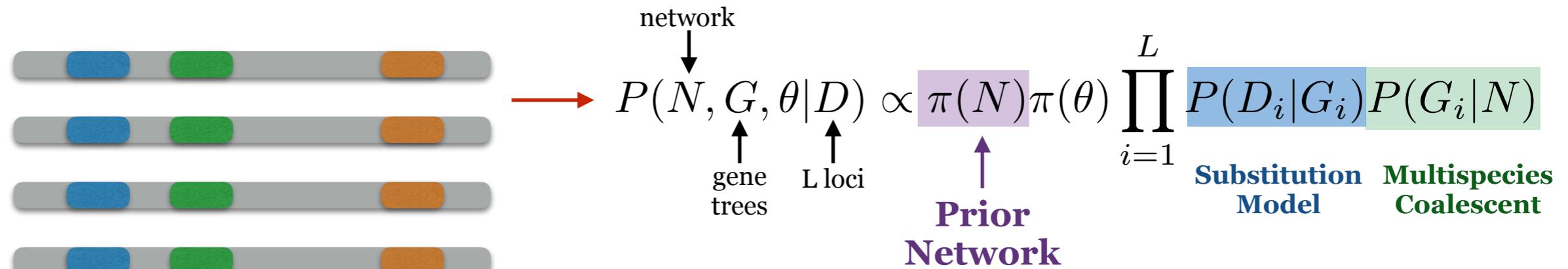
MrBayes
(Huelsenbeck, Ronquist, 2001)
RAxML
(Stamatakis, 2014)
PhyML
(Guindon et al, 2010)
RevBayes
(Hoehna et al, 2016)
IQ-TREE
Nguyen et al. (2015)

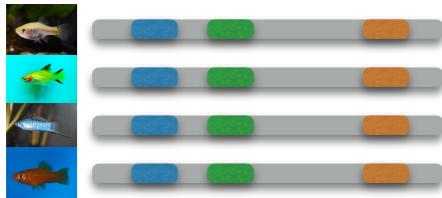


BEAST2
(Zhang et al, 2017)
PhyloNet
(Wen et al, 2016)



SNaQ
(S.-L., Ane, 2016)
PhyloNet
(Yu et al, 2014)

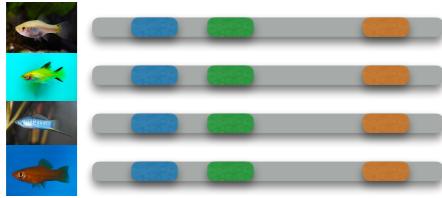




BEAST2
(Zhang et al, 2017)

Birth-hybridization process

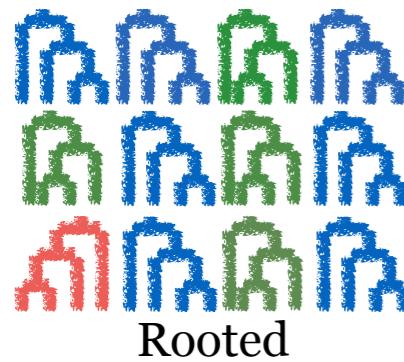
Most accurate,
not scalable



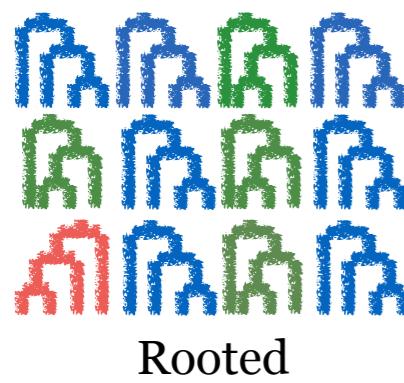
PhyloNet
Bayesian
(Wen et al, 2016)

reticulations,
cycle diameter

MCMC:
Network
moves,
mixing

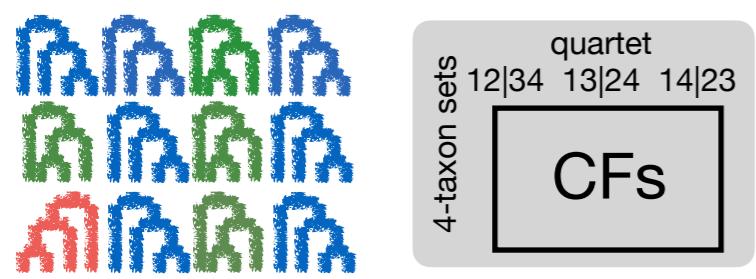


PhyloNet
Bayesian
(Wen et al, 2016)



PhyloNet
Likelihood
(Yu et al, 2014)

**Heuristic
search:**
Network
moves



SNaQ
(S.-L., Ane, 2016)

Level-1
networks

More scalable,
Robust

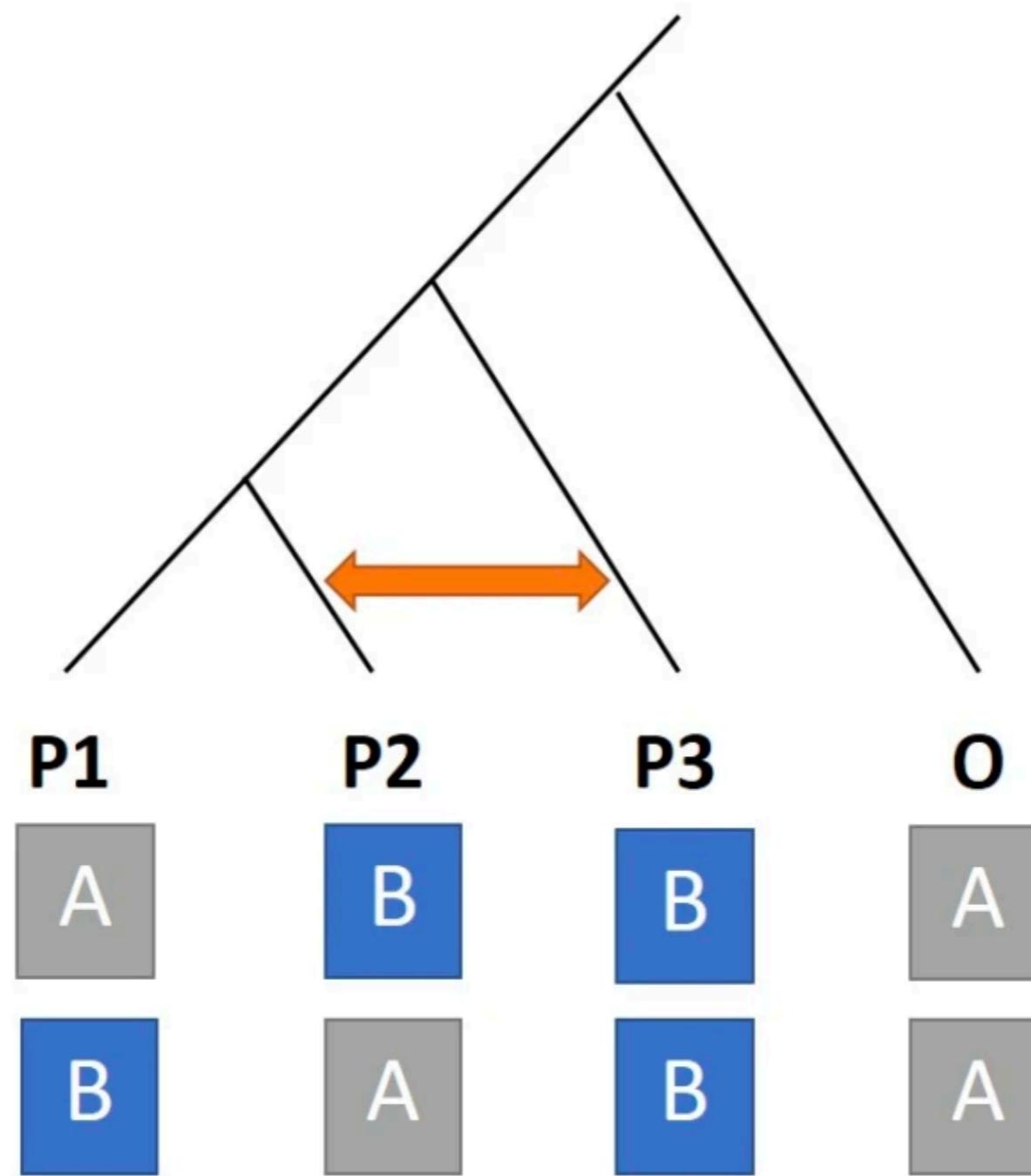
| | | | |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

Hybrid detection methods

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

D statistics





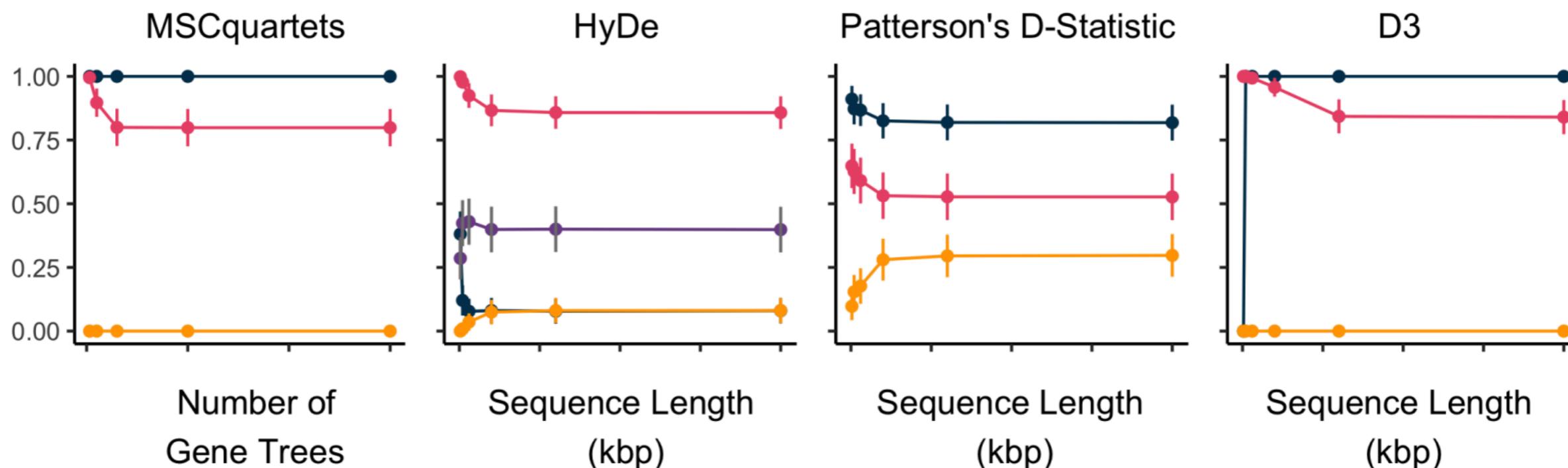
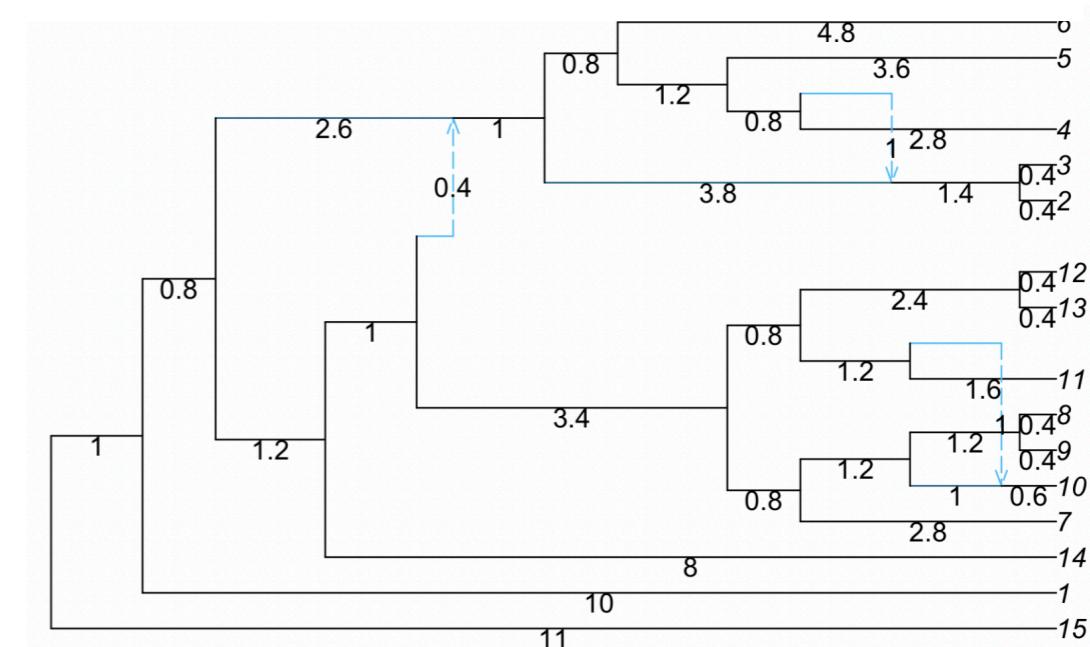
[Submitted on 1 Nov 2022]

Marianne
Bjørner

Detectability of Varied Hybridization Scenarios using Genome-Scale Hybrid Detection Methods

Marianne Bjørner, Erin K. Molloy, Colin N. Dewey, Claudia Solis-Lemus

Ancient gene flow
distorts discovery
rates for hybrid
detection methods



(See also Kong & Kubatko, 2021)

● Precision ● False Positive Rate ● False Negative Rate ● Wrong Hybrid Rate

Hybrid detection methods

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

Estimate
explicit
network
under the
coalescent

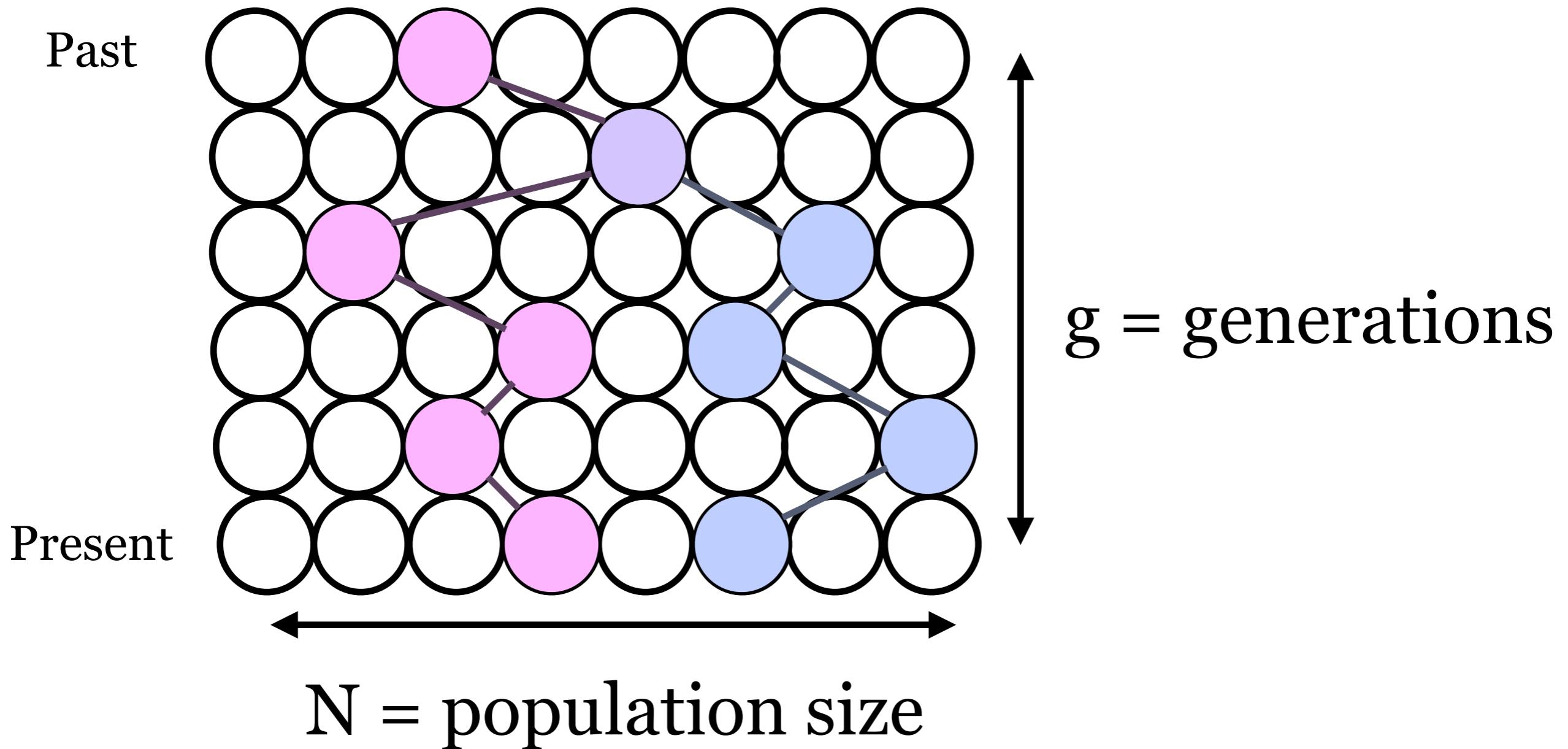


Hybrid detection methods

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

Estimate
explicit
network
under the
coalescent

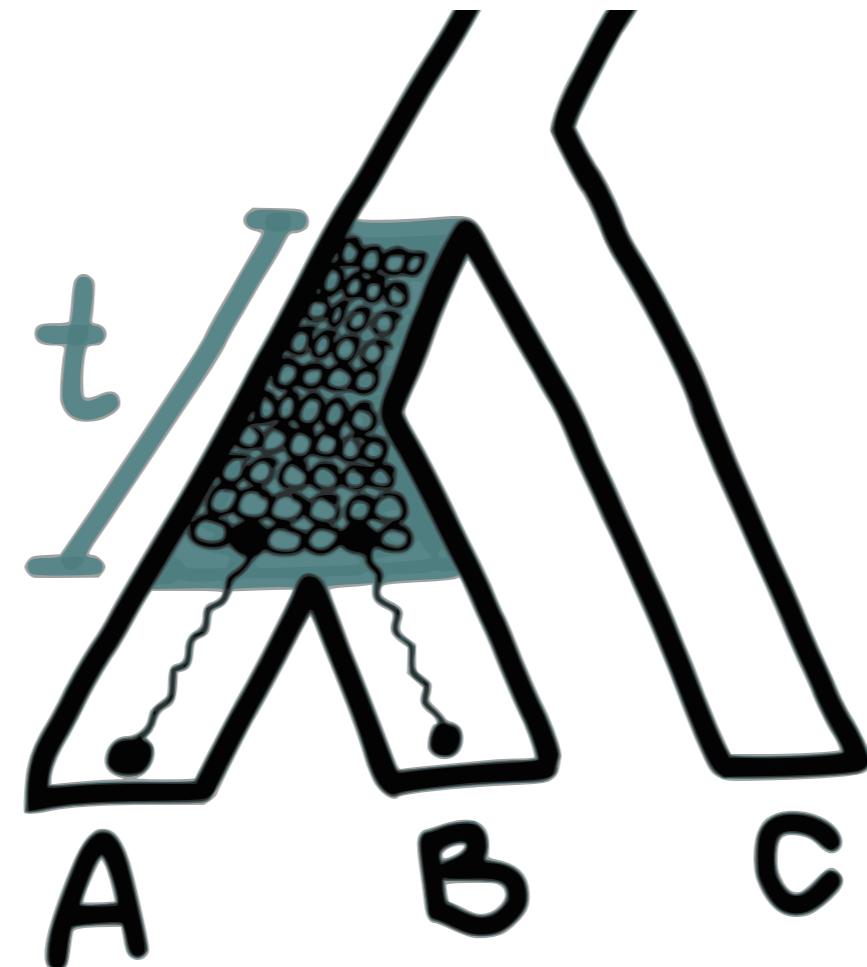
Coalescent model within 1 population



Probability of no coalescence in g generations:

$$\left(1 - \frac{1}{N}\right)^g$$
$$t = g/N \Rightarrow \left(1 - \frac{t}{Nt}\right)^{Nt} \xrightarrow[N \rightarrow \infty]{} e^{-t}$$

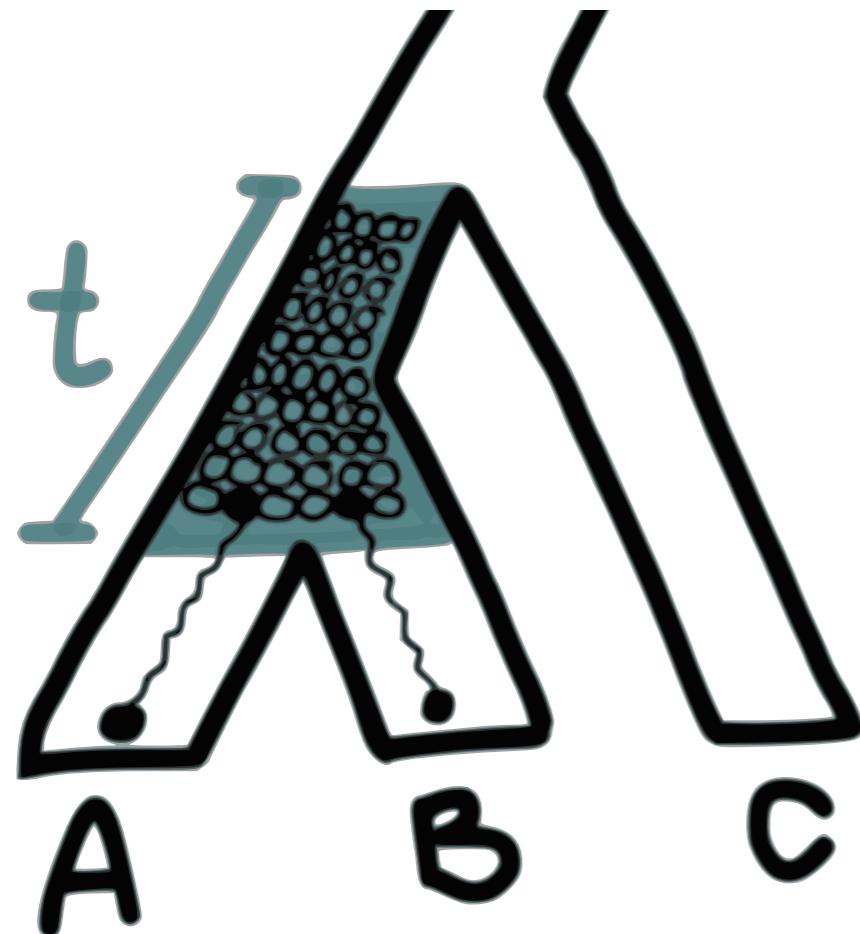
Multispecies coalescent on a tree



$$P(T > t) = e^{-t}$$

$$T = \frac{g}{N} \text{ coalescent units} \sim Exp(1)$$

Multispecies coalescent on a tree

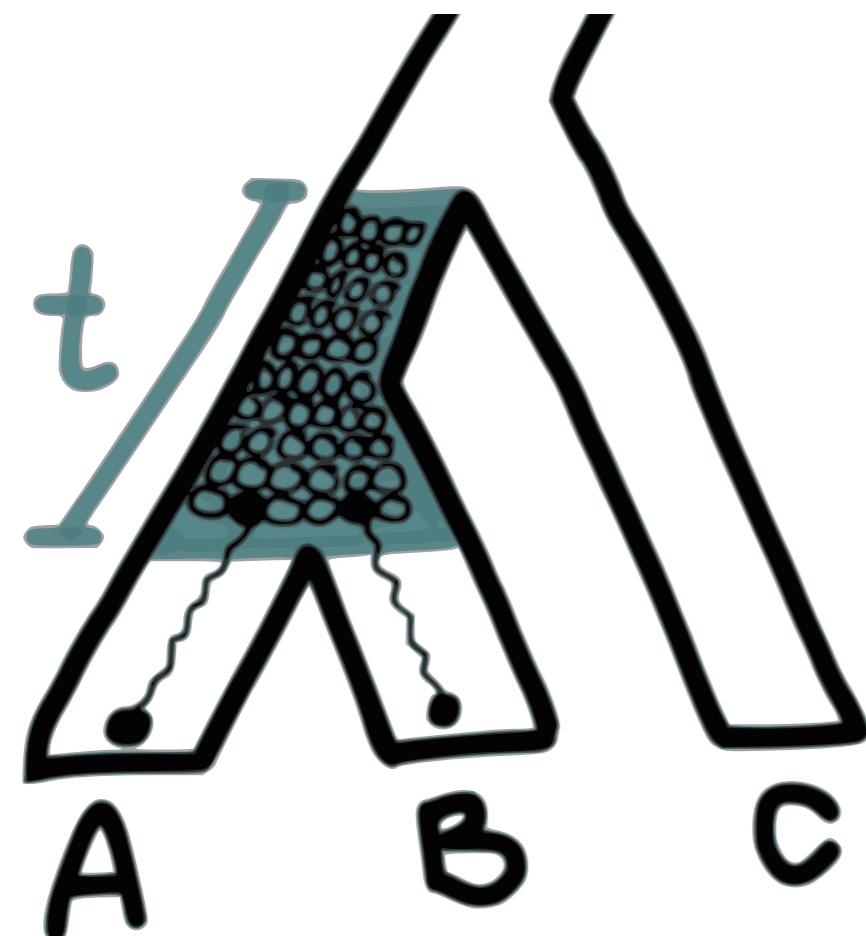


$$P(\text{ } \text{ } \text{ } \text{ } \text{ } \text{ }) =$$

A large black letter P followed by a large black parenthesis containing a phylogenetic tree with three tips labeled A, B, and C. The entire expression is followed by an equals sign.

$$P(T > t) = e^{-t}$$

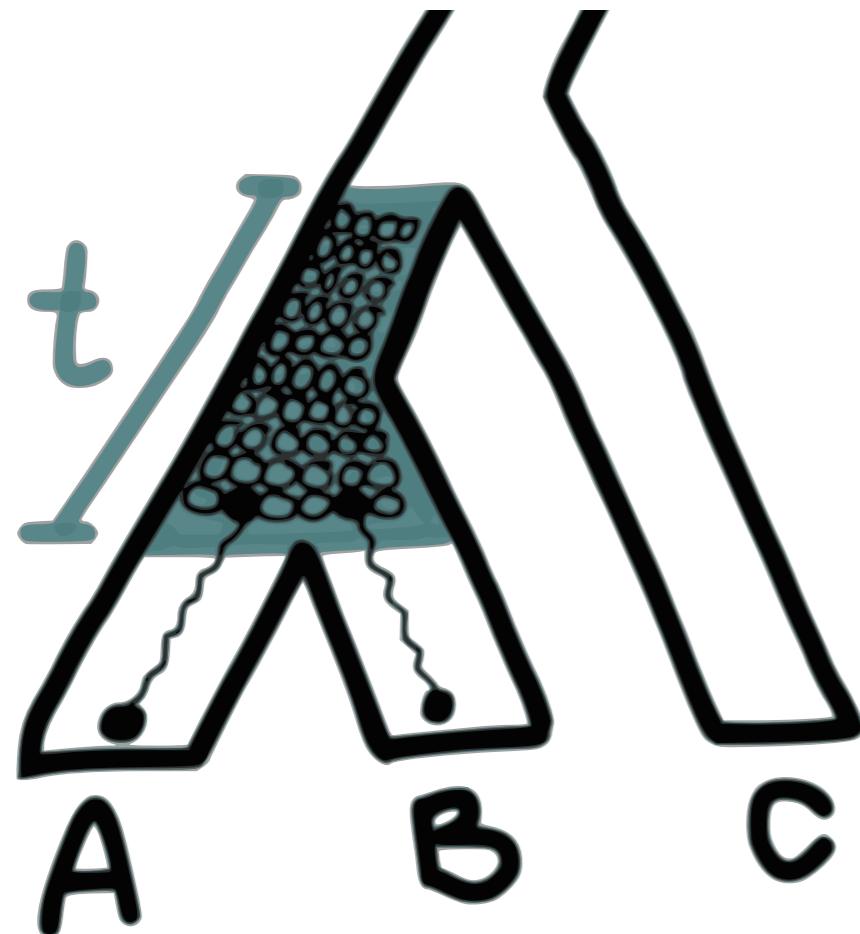
Multispecies coalescent on a tree



$$P(\text{ } \text{ } \text{ } \text{ } \text{ } \text{ }) = \\ 1 - e^{-t}$$

$$P(T > t) = e^{-t}$$

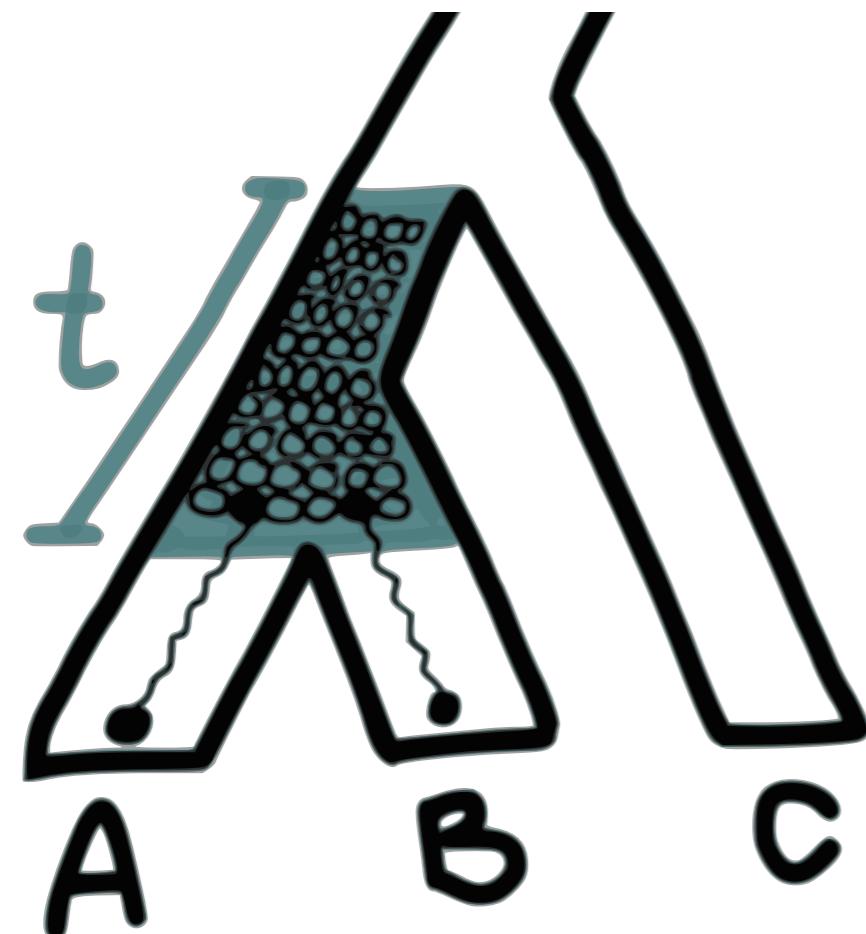
Multispecies coalescent on a tree



$$P(\text{ } \text{ } \text{ } \text{ } \text{ } \text{ }) = \\ 1 - e^{-t} \\ +$$

$$P(T > t) = e^{-t}$$

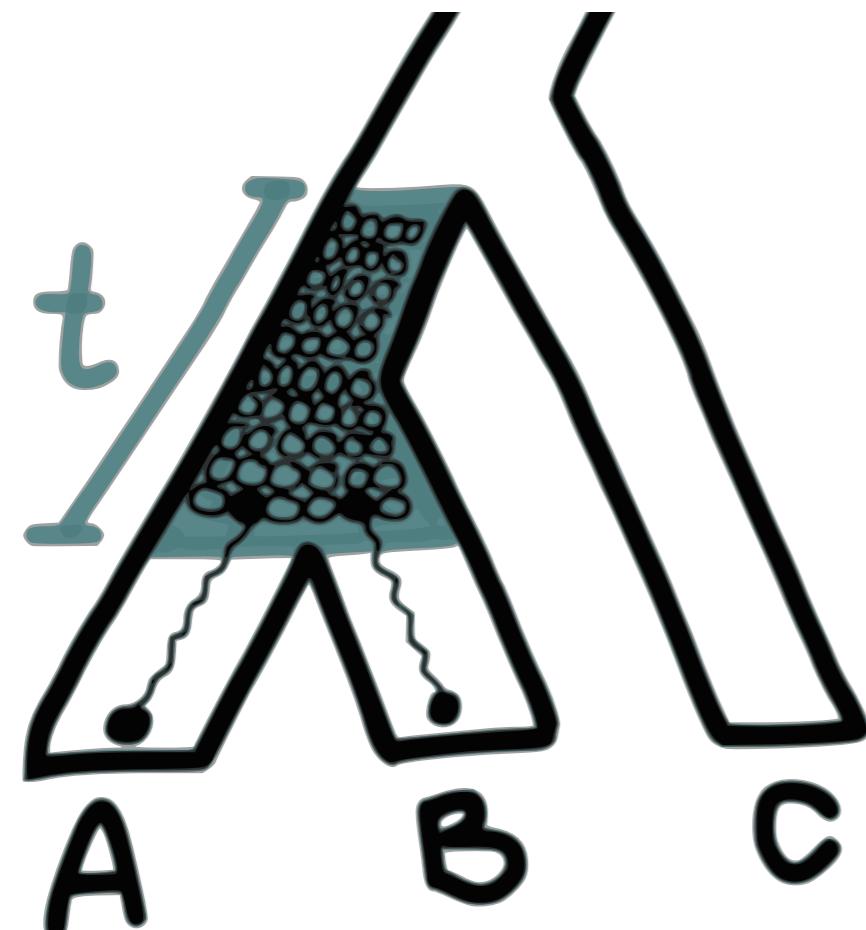
Multispecies coalescent on a tree



$$P(\text{ } \text{ } \text{ } \text{ } \text{ } \text{ }) = \\ 1 - e^{-t} \\ + \\ e^{-t} \times 1/3$$

$$P(T > t) = e^{-t}$$

Multispecies coalescent on a tree



$$P(\text{ } \text{ } \text{ } \text{ } \text{ } \text{ }) =$$

$$1 - e^{-t}$$

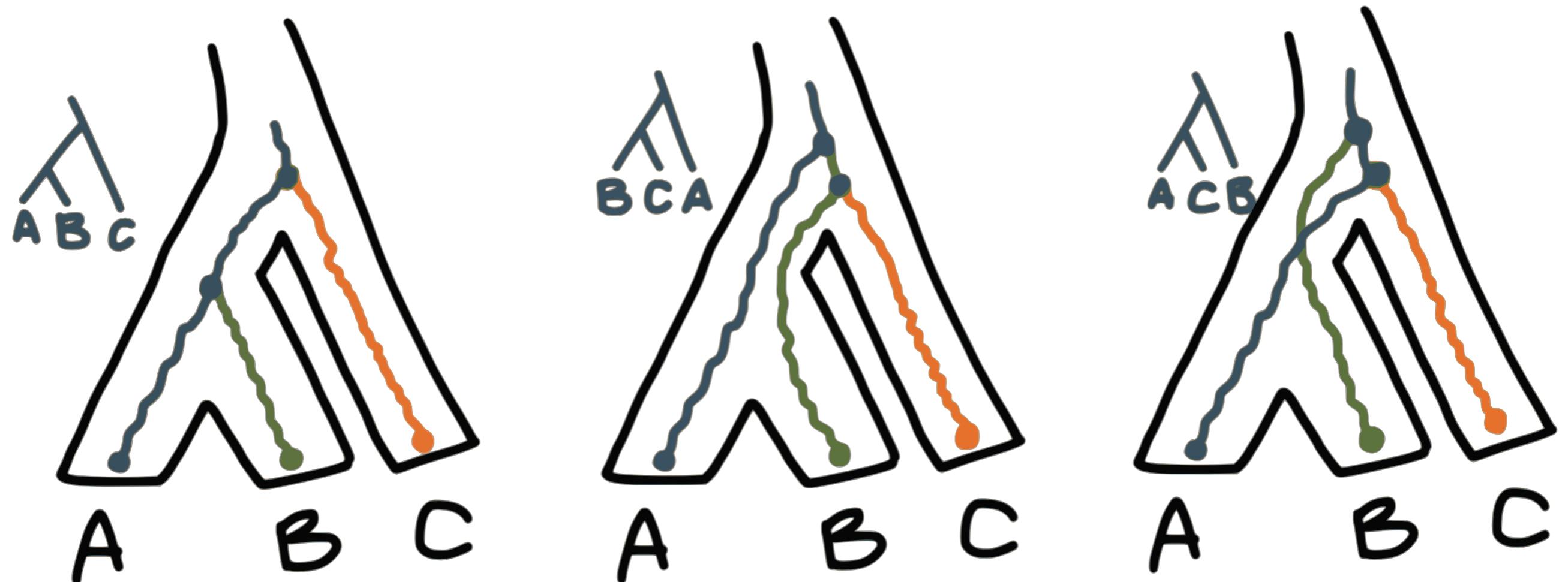
+

$$e^{-t} \times 1/3$$

$$= 1 - \frac{2}{3}e^{-t}$$

$$P(T > t) = e^{-t}$$

Multispecies coalescent on a tree

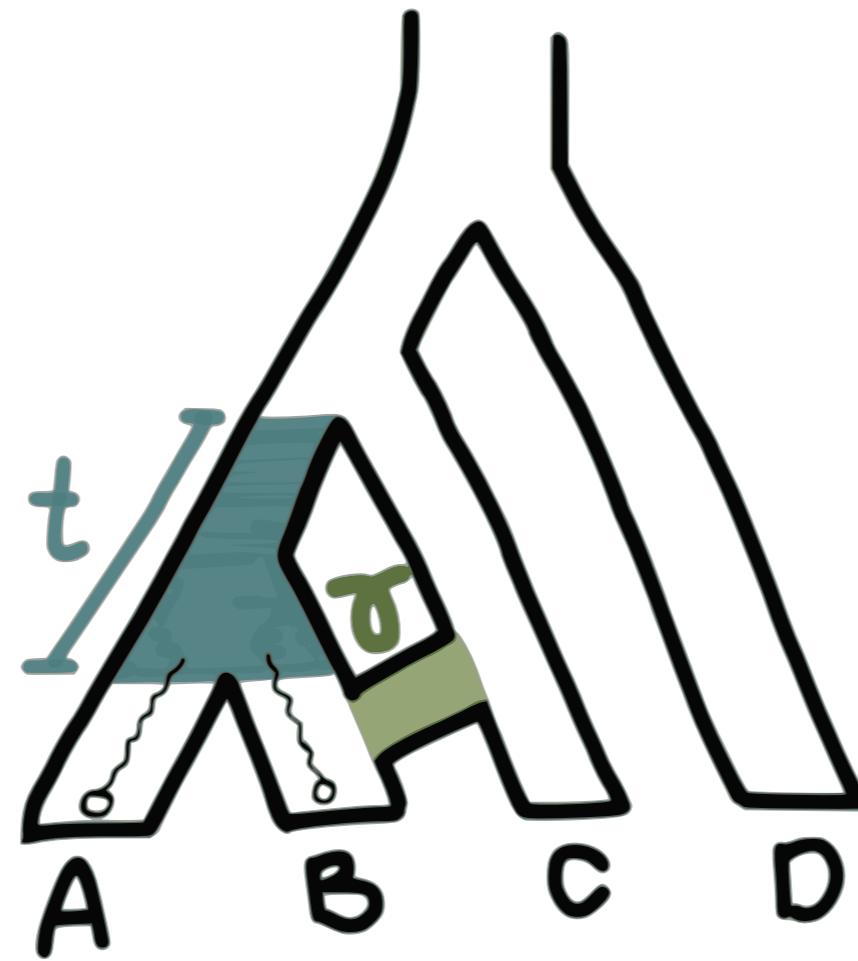


$$1 - \frac{2}{3}e^{-t}$$

$$\frac{1}{3}e^{-t}$$

$$\frac{1}{3}e^{-t}$$

Multispecies coalescent on a network



(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

Multispecies coalescent on a network



(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

Multispecies coalescent on a network



$$(1 - \gamma) \frac{1}{3} e^{-t} + \gamma(1 - \frac{2}{3} e^{-t_2})$$

(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

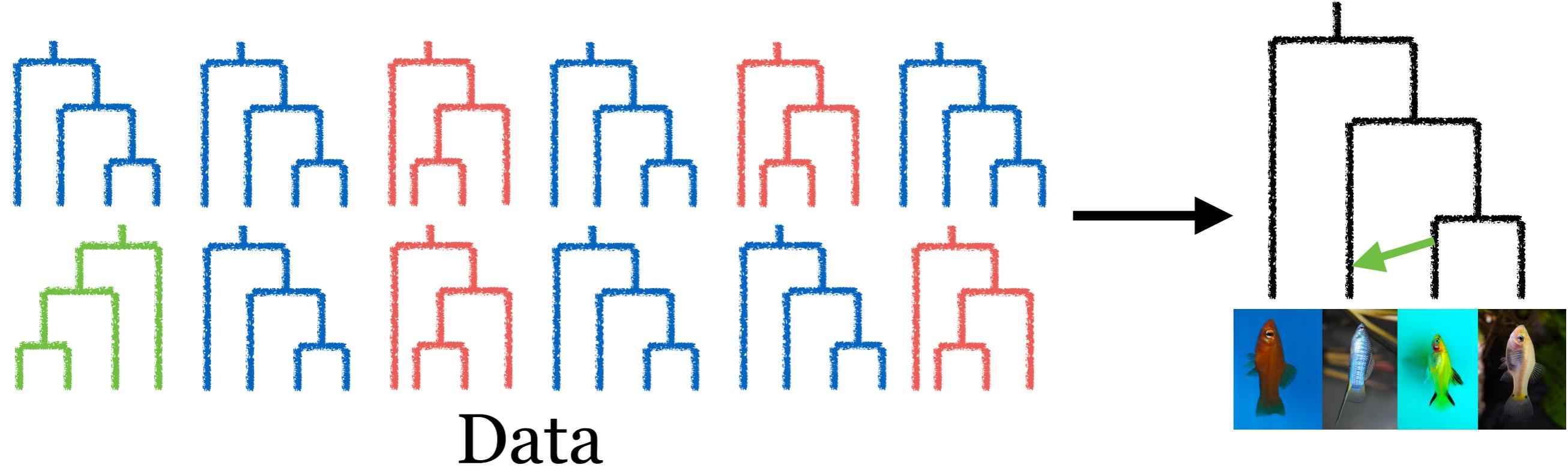
Multispecies coalescent on a network



$$CF_{BC|AD}(t, t_2, \gamma) = (1 - \gamma) \frac{1}{3} e^{-t} + \gamma (1 - \frac{2}{3} e^{-t_2})$$

(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

Maximum pseudolikelihood



$$\tilde{L}(\textit{network}) = \prod L(\textit{quartet})$$

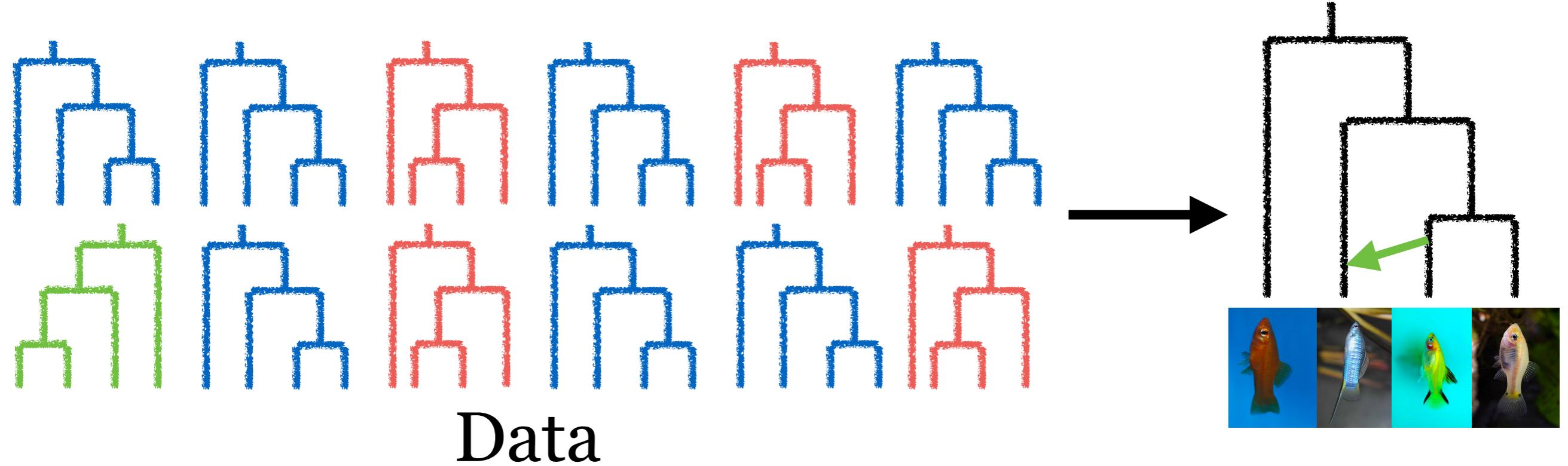
(S-L, Ané, 2016, PLoS Genetics)

www.github.com/CRSL4/PhyloNetworks

snaQ julia

Quartet-based inference

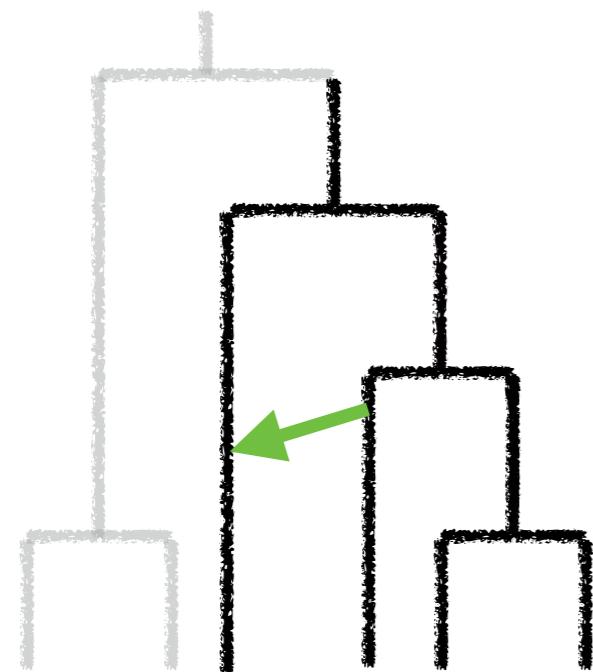
Maximum pseudolikelihood



$$\tilde{L}(\textit{network}) = \prod L(\textit{quartet})$$

(S-L, Ané, 2016, PLoS Genetics)

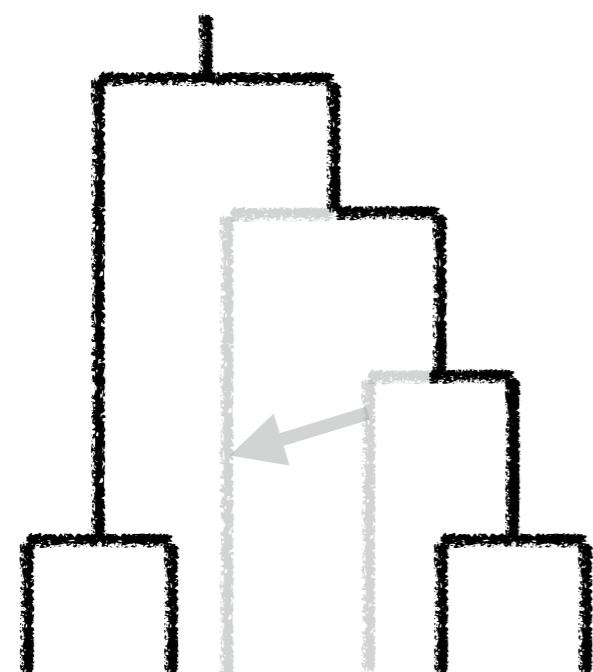
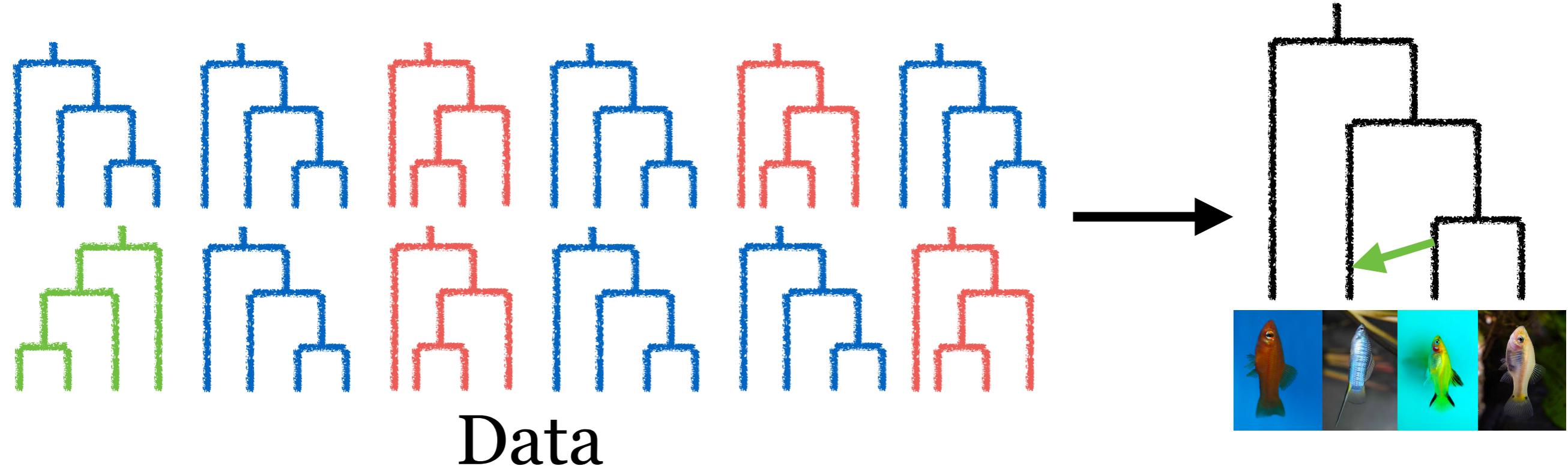
www.github.com/CRSL4/PhyloNetworks



Quartet-based inference

snaQ julia

Maximum pseudolikelihood



Quartet-based inference

$$\tilde{L}(\textit{network}) = \prod L(\textit{quartet})$$

(S-L, Ané, 2016, PLoS Genetics)

www.github.com/CRSL4/PhyloNetworks

snaQ julia

Maximum pseudolikelihood

Unrooted gene trees

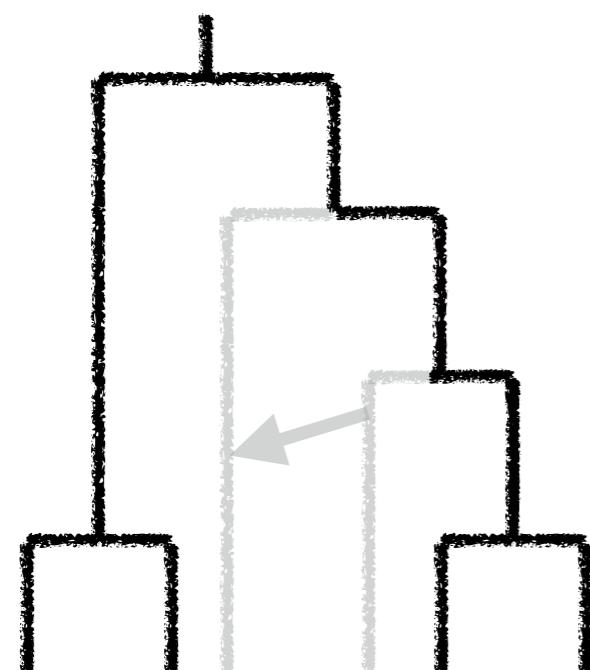
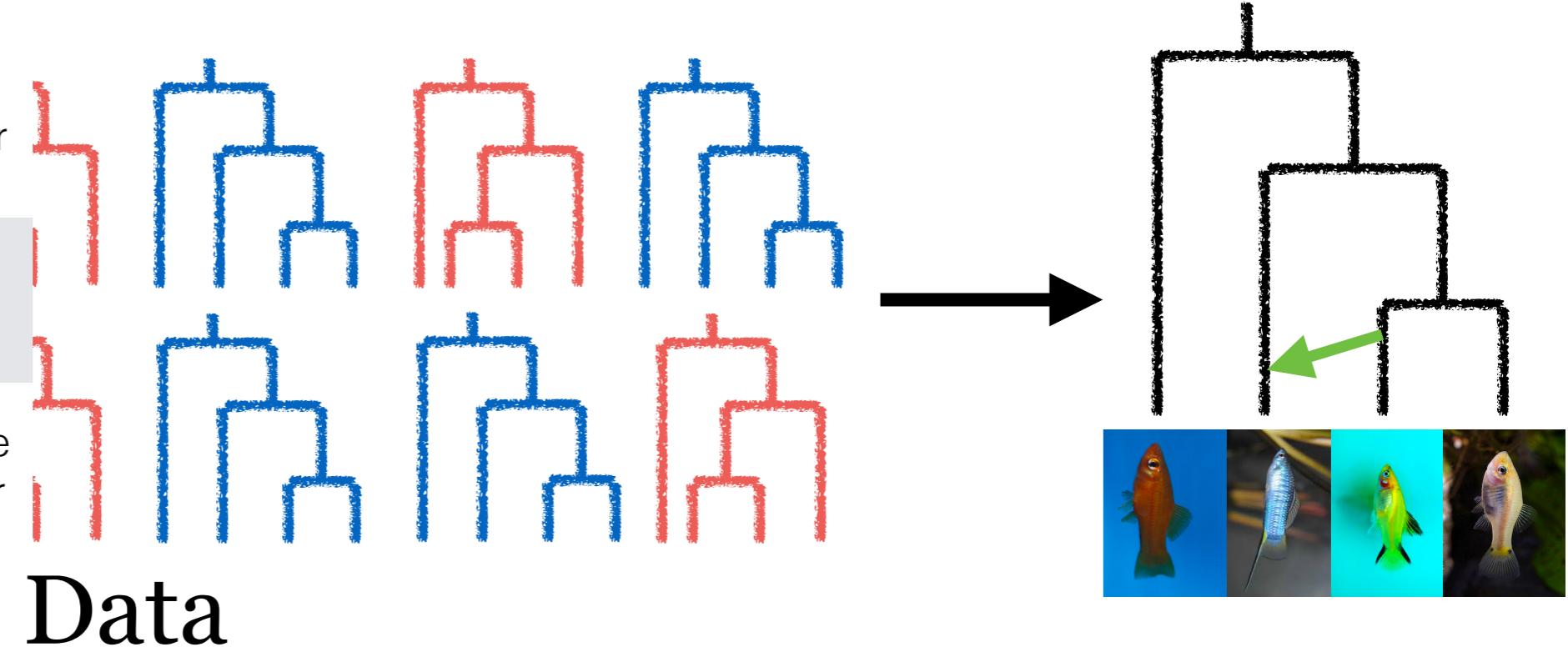
No branch lengths

Concordance factors

No rooting error

No molecular clock assumption

Account for tree estimation error



Quartet-based inference

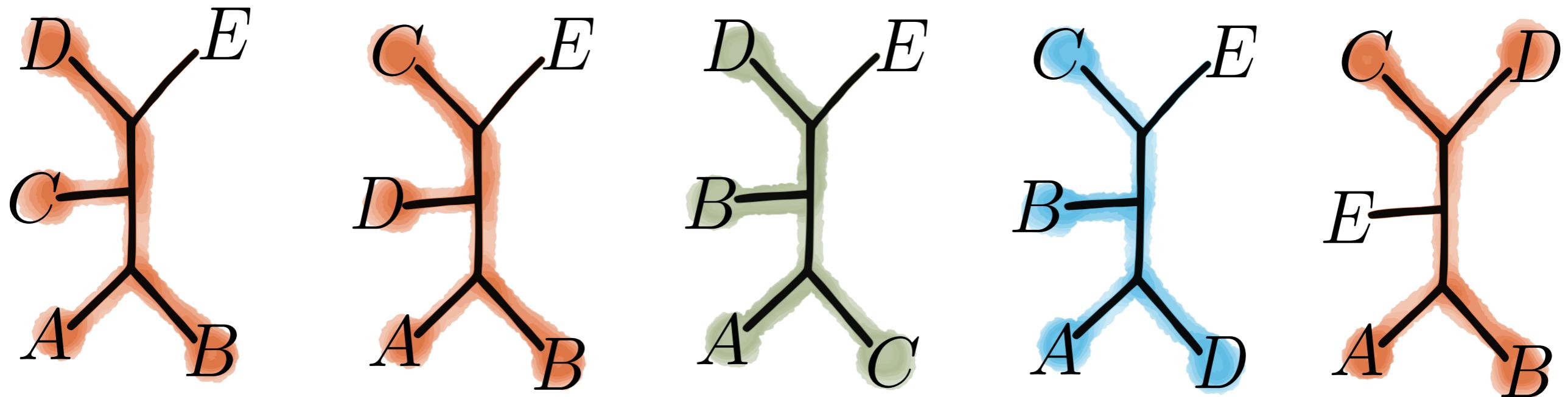
$$\tilde{L}(\textit{network}) = \prod L(\textit{quartet})$$

(S-L, Ané, 2016, PLoS Genetics)

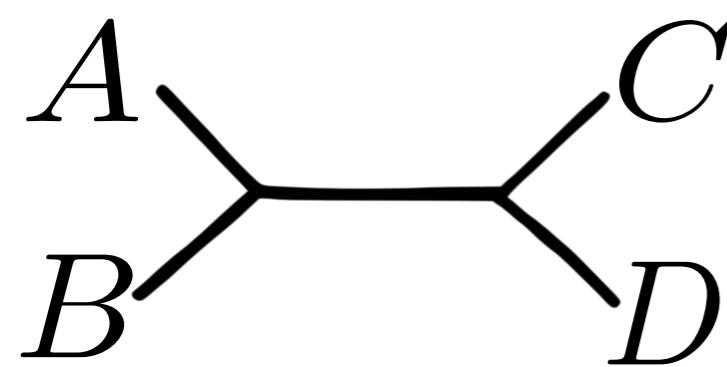
www.github.com/CRSL4/PhyloNetworks

snaQ julia

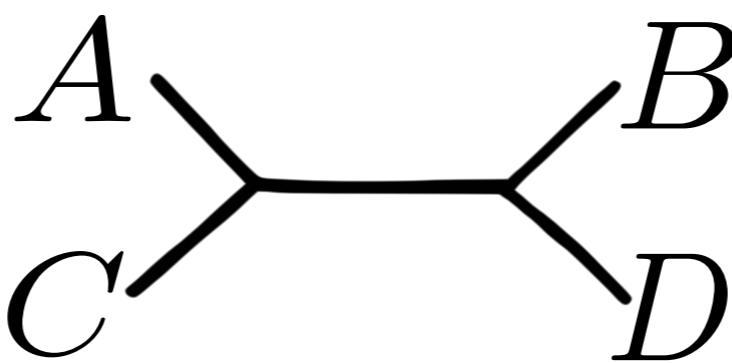
Quartet-based inference



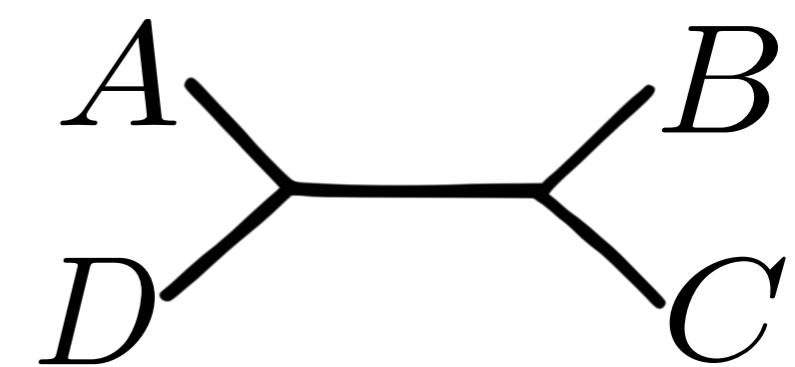
Concordance factors (CF):
% of genes having the quartet in their tree



3/5



1/5



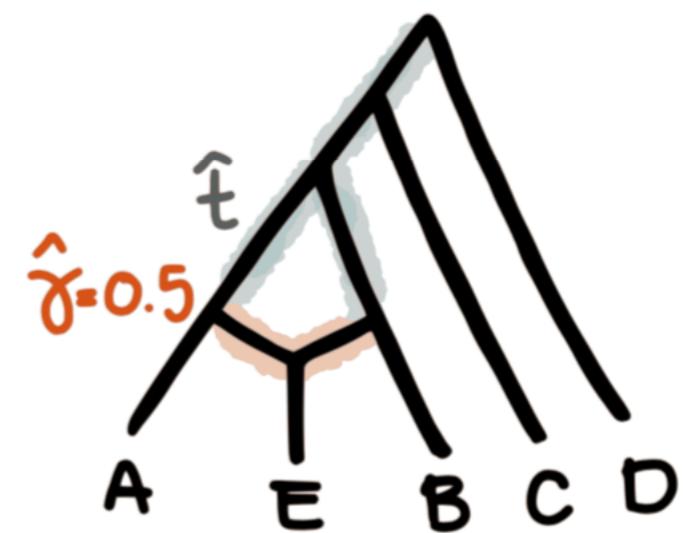
1/5

Quartet-based inference

Observed **quartet** CFs:

| 4 taxon set | CF_1 | CF_2 | CF_3 |
|-------------|--------|--------|--------|
| A B C D | .80 | .10 | .10 |
| A B C E | .40 | .40 | .20 |
| A B D E | .40 | .40 | .20 |
| A C D E | .84 | .08 | .08 |
| B C D E | .82 | .10 | .08 |

inferred network:

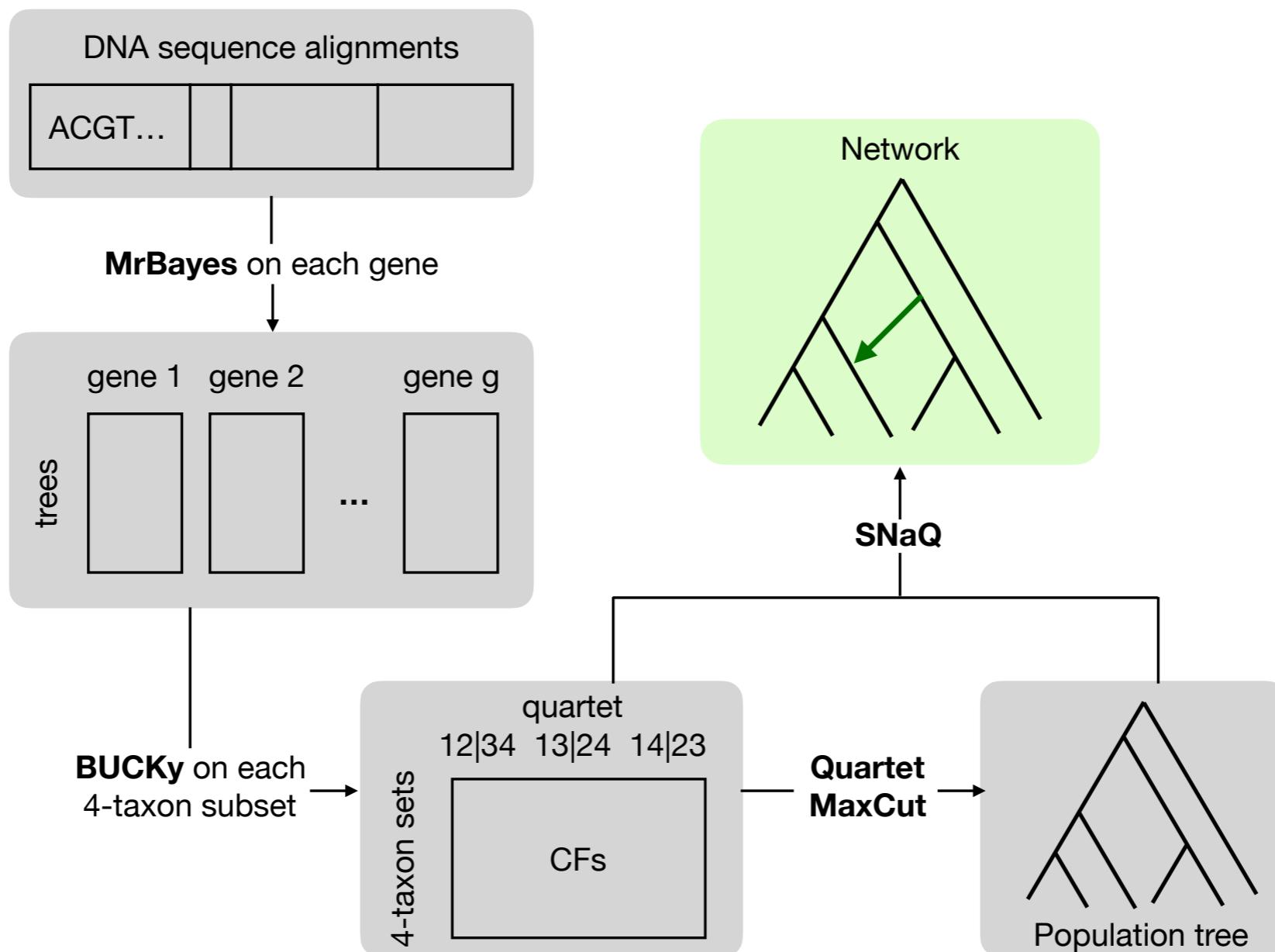


Maximum Pseudo-Likelihood:

$$\log \tilde{L} = \sum_{q \in Q(N)} CF_{in,1} \log(CF_{net,1}) + CF_{in,2} \log(CF_{net,2}) + CF_{in,3} \log(CF_{net,3})$$

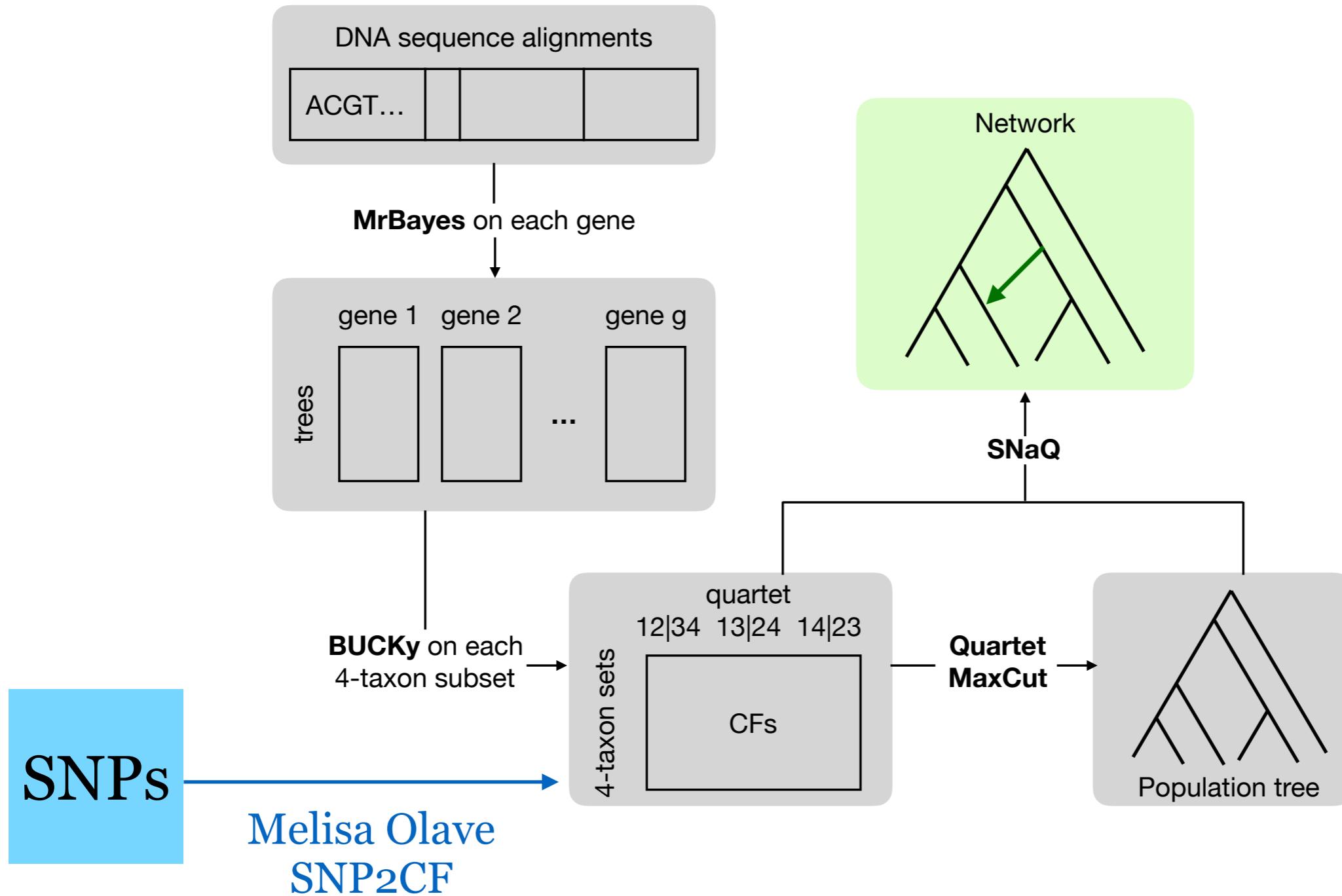
How?

Phylogenetic network



How?

Phylogenetic network



Hybrid detection methods

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

Estimate
explicit
network
under the
coalescent

Hybrid
detection

+

Network
estimation

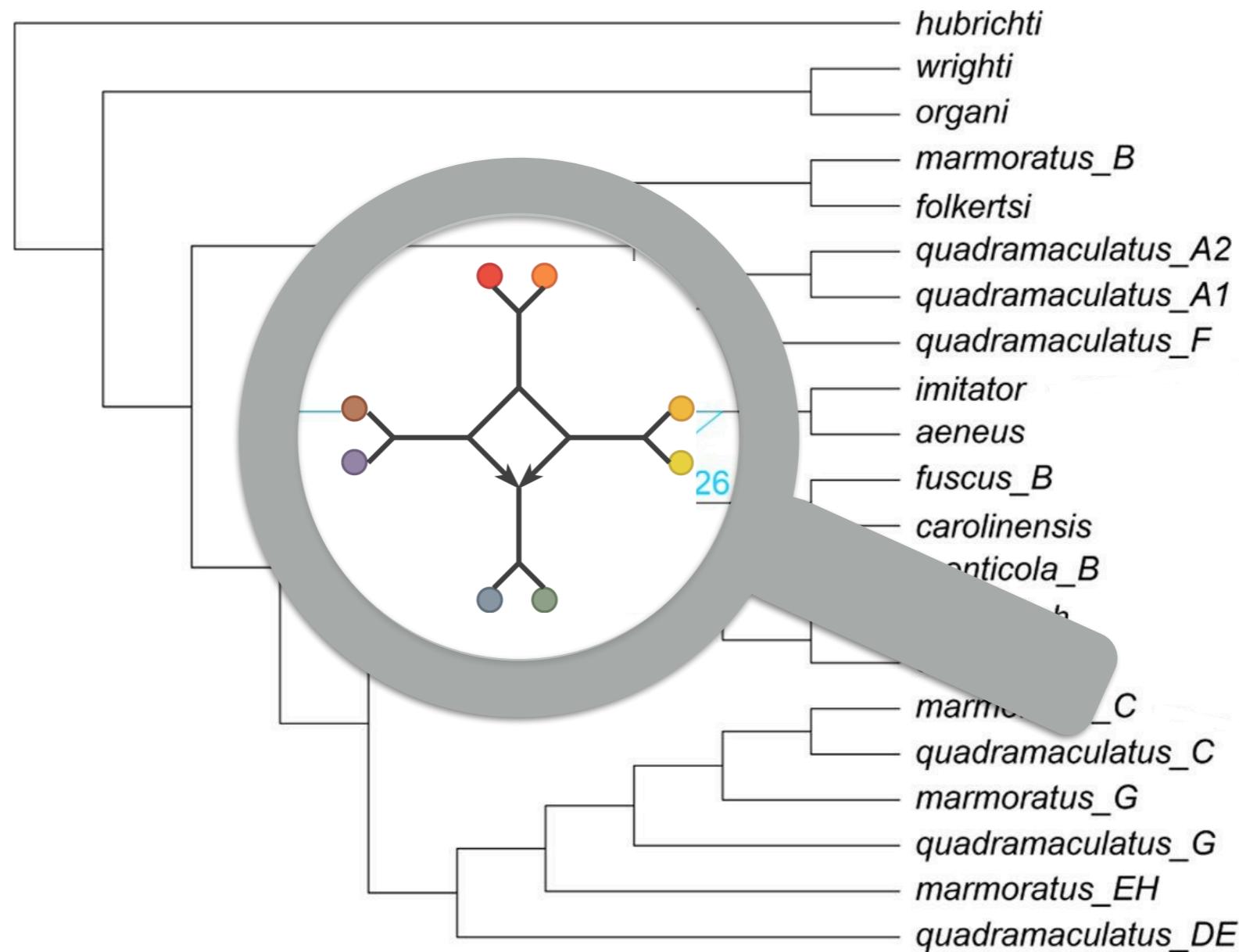
Bella
Wu

Ultrafast learning of 4-node hybridization cycles in phylogenetic networks using algebraic invariants

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes

**Phylo
Diamond.jl**

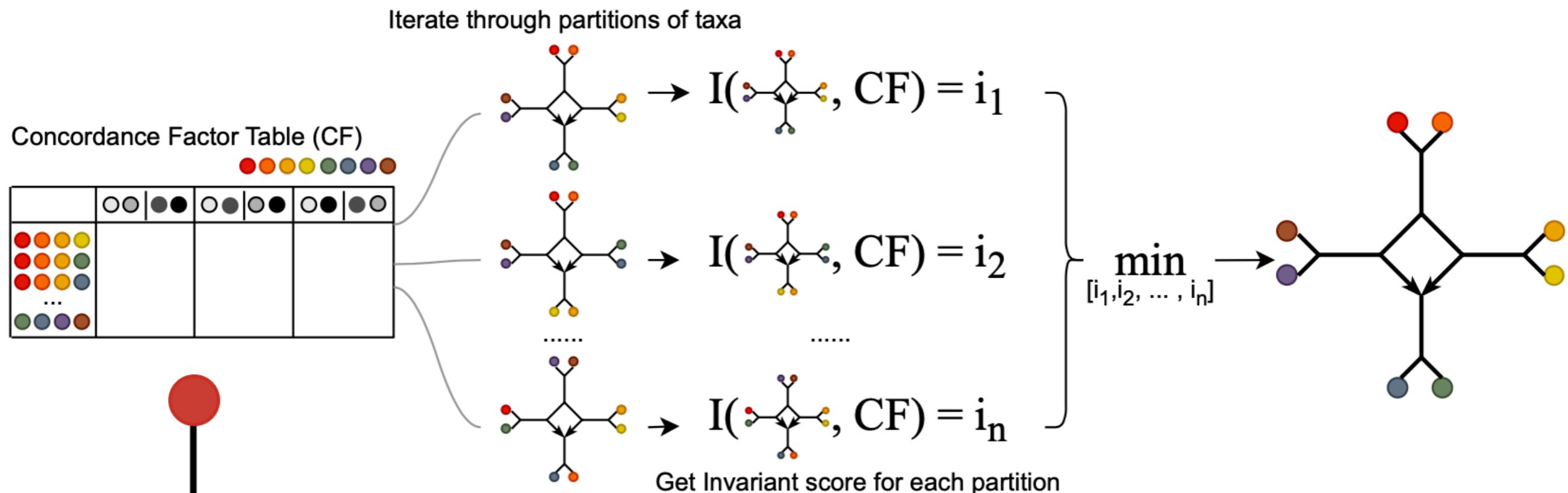


Bella
Wu

Ultrafast learning of 4-node hybridization cycles in phylogenetic networks using algebraic invariants

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes



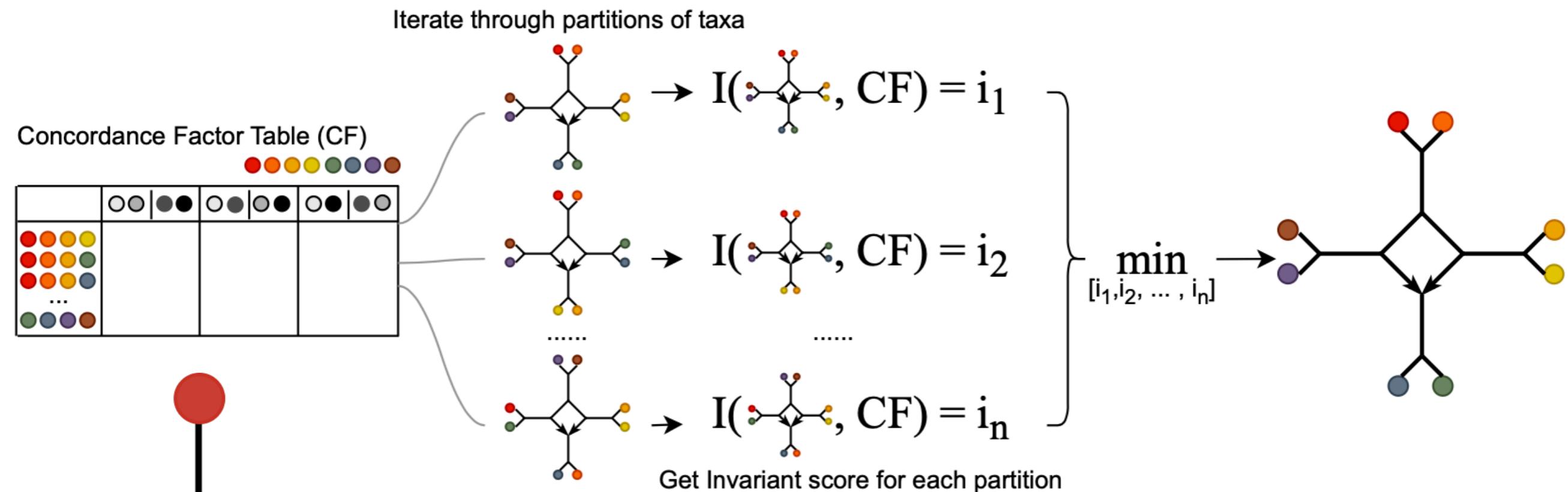
Phylo
Diamond.jl

Bella
Wu

Ultrafast learning of 4-node hybridization cycles in phylogenetic networks using algebraic invariants

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes



Phylo
Diamond.jl

Note: it cannot resolve relationships among taxa in each of the four clades

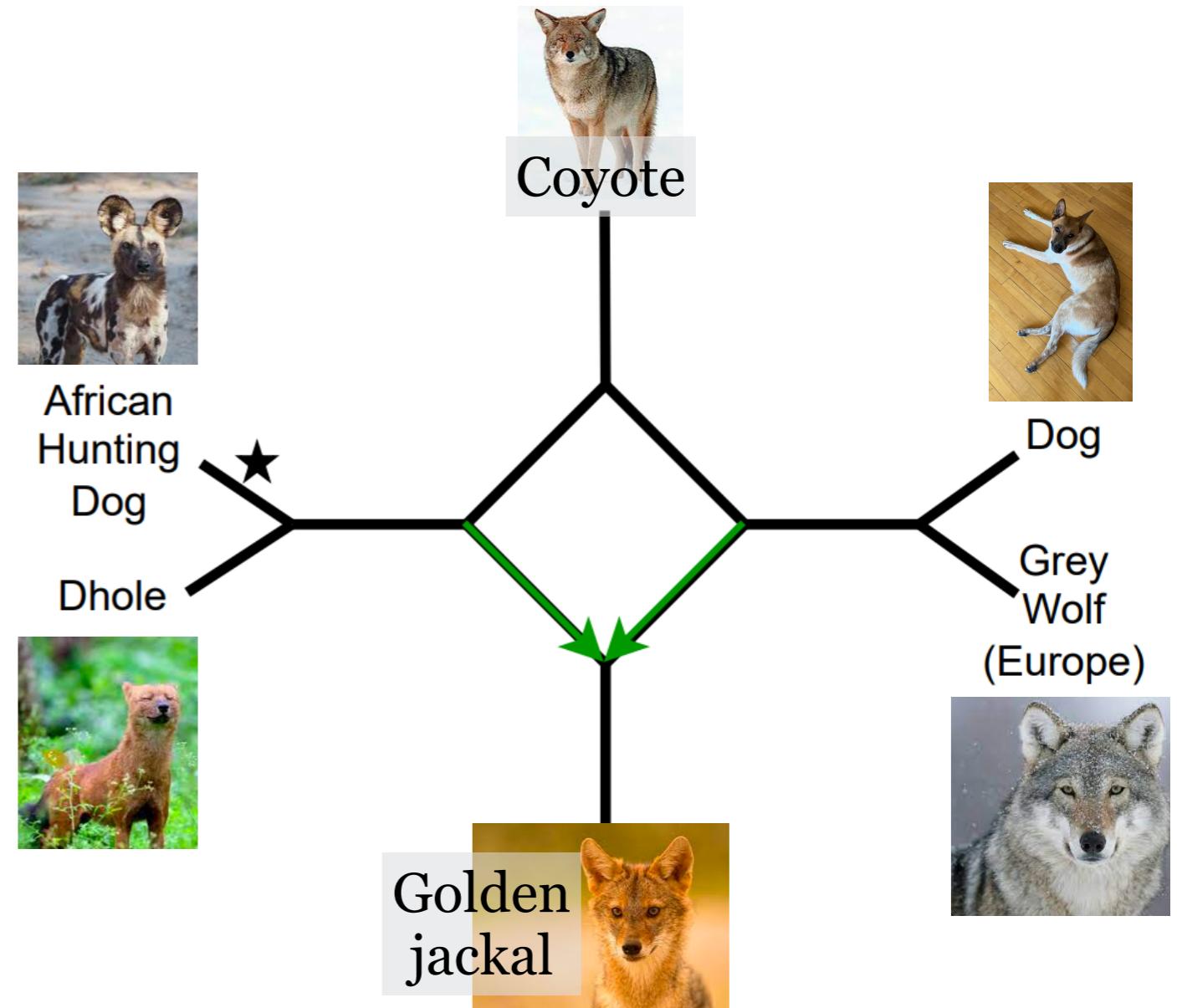
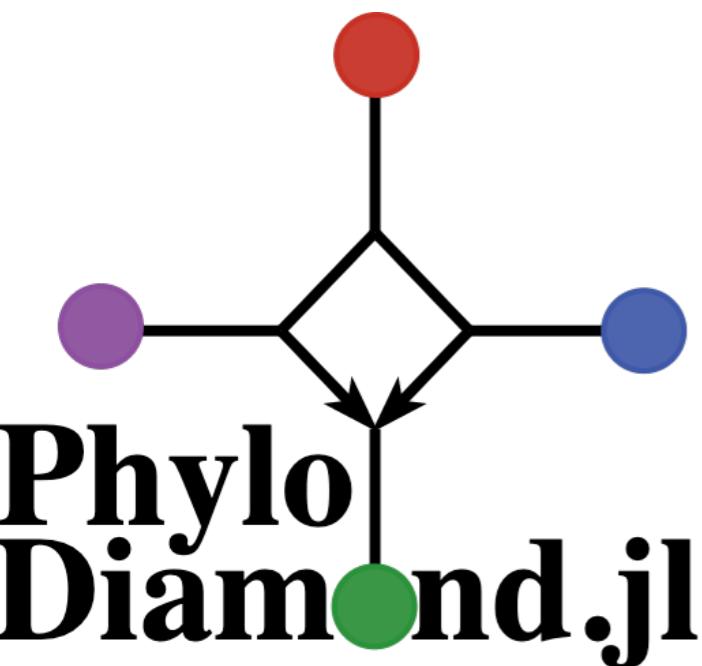
Bella
Wu

Ultrafast learning of 4-node hybridization cycles in phylogenetic networks using algebraic invariants

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes

| Method | Time (seconds) |
|--------------------------------------|----------------|
| Phylogenetic invariants (our method) | 6.78 |
| SNaQ | 140.58 |
| PhyloNet ML | 2723.99 |
| PhyloNet MPL | 281.25 |



(Data from Gopalakrishnan et al, 2018)

Hybrid detection methods

| STEM-hy | gene trees rooted, BL | likelihood | hybridization b/w sister lineages |
|------------------------------|------------------------------|-----------------------|--------------------------------------|
| PhyloNet InferNetwork_ML | gene trees rooted | likelihood | |
| PhyloNet InferNetwork_MPL | gene trees rooted | triplet likelihood | |
| PhyloNetworks SNaQ | gene trees or quartet CFs | quartet likelihood | level-1 network |
| PhyloNet MCMC_GT | gene trees rooted | Bayesian | compound prior |
| PhyloNet MCMC_SEQ | alignments | Bayesian | compound prior no rate variation |
| BEAST2 SpeciesNetwork | alignments | Bayesian | birth-hyb prior |
| PhyloNet MLE_BiMarkers | biallelic sites | likelihood | compound prior |
| PhyloNet MCMC_BiMarkers | biallelic sites | Bayesian | compound prior |
| HyDe | sites | invariants | 4 taxa, 1 hyb. |
| D-statistics | sites | freq. patterns | 4 taxa, 1 hyb. |
| MSCQuartets | gene trees or quartet CF | quartet likelihood | 4 taxa, 1 hyb. |
| NANUQ | gene trees or quartet CF | quartet likelihood | split (level-1) network |
| RF-Net | gene trees | min RF distance | No ILS |
| PhyNEST.jl | sites | quartet likelihood | level-1 network |
| PhyloDiamond.jl | gene trees or quartet CF | quartet likelihood | level-1 4-cycle network |

Estimate
explicit
network
under the
coalescent

Hybrid
detection

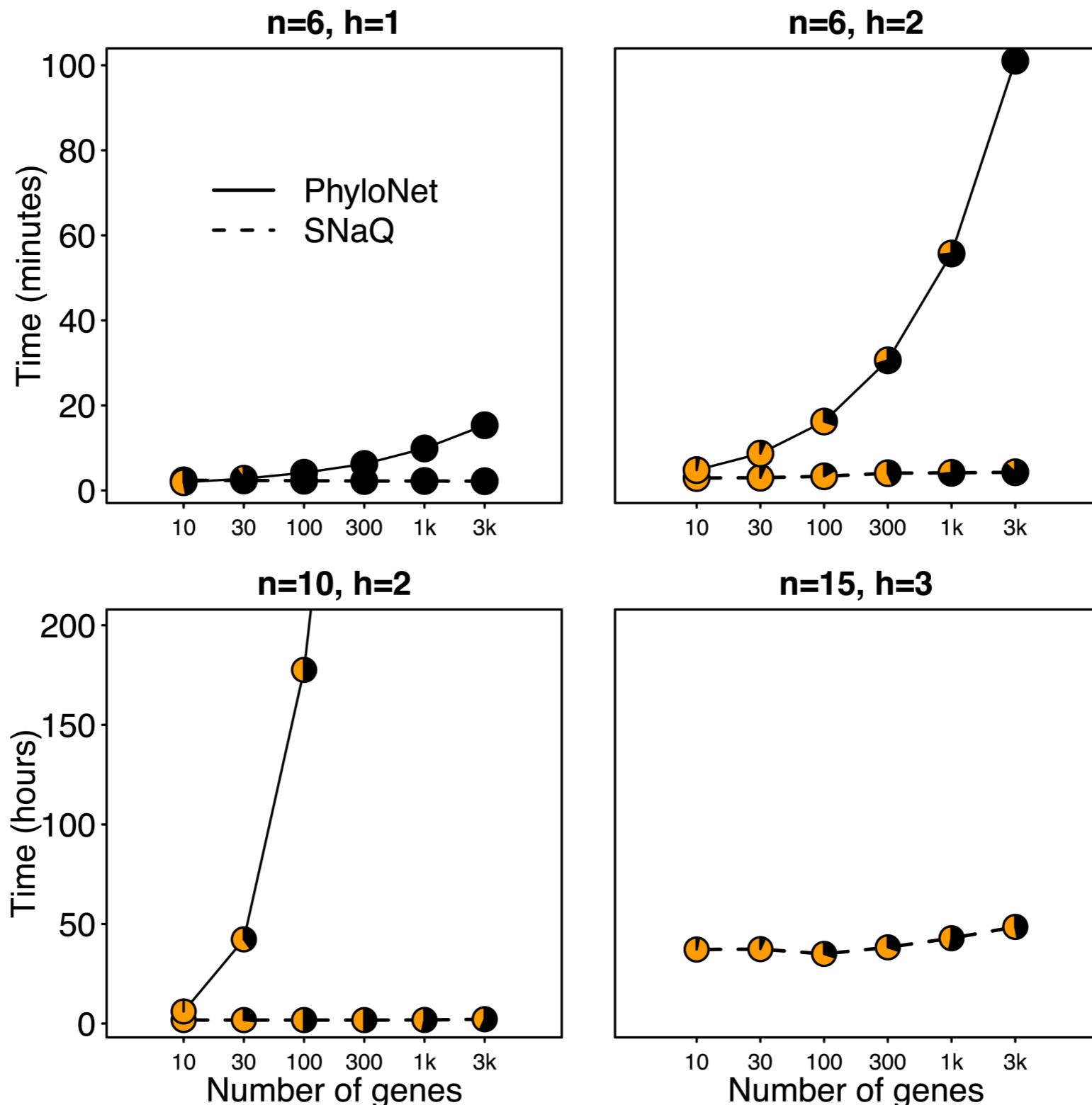
+

Network
estimation

Network challenges

- Scalability
- Identifiability
- Network comparison

Scalability

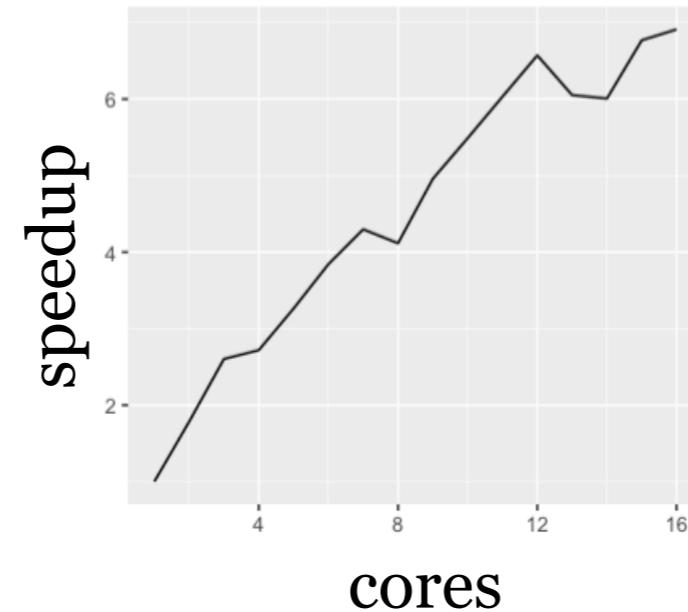
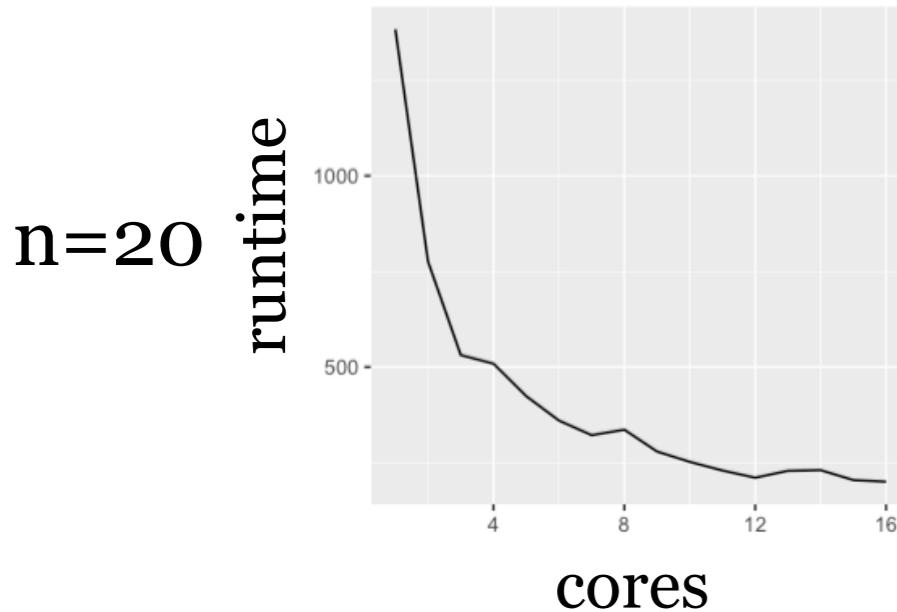
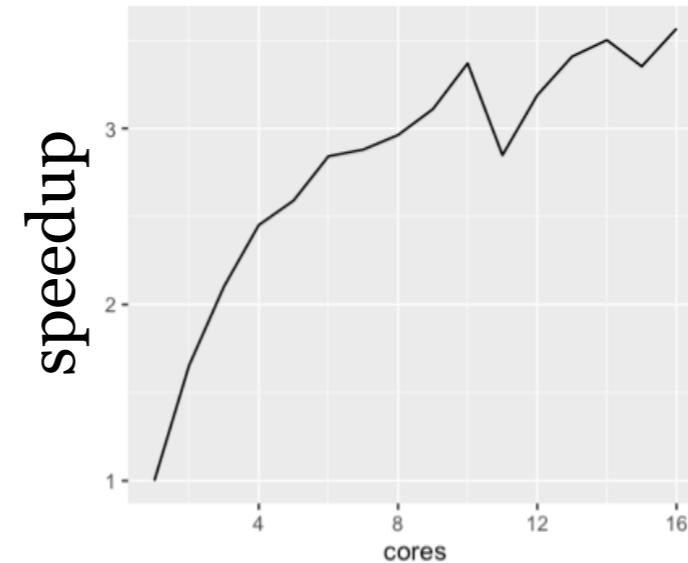
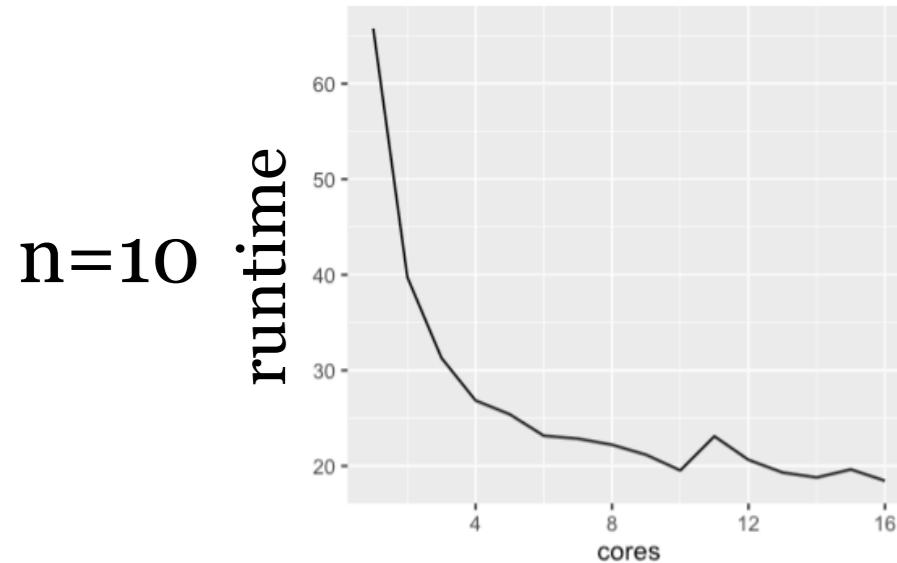


(Solís-Lemus, Ané, 2016, PLoS Genetics)

SNaQ is faster, but is it fast?



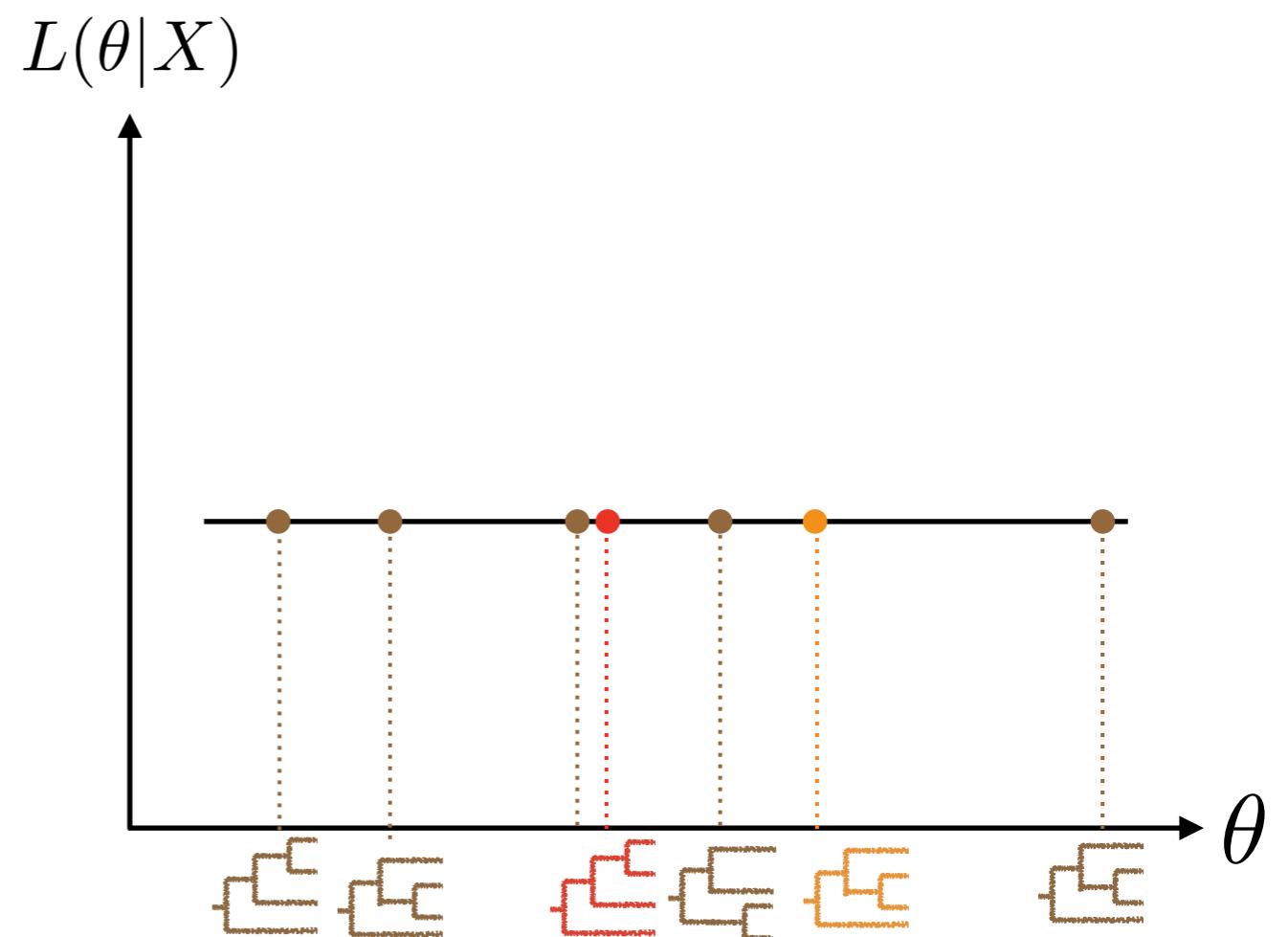
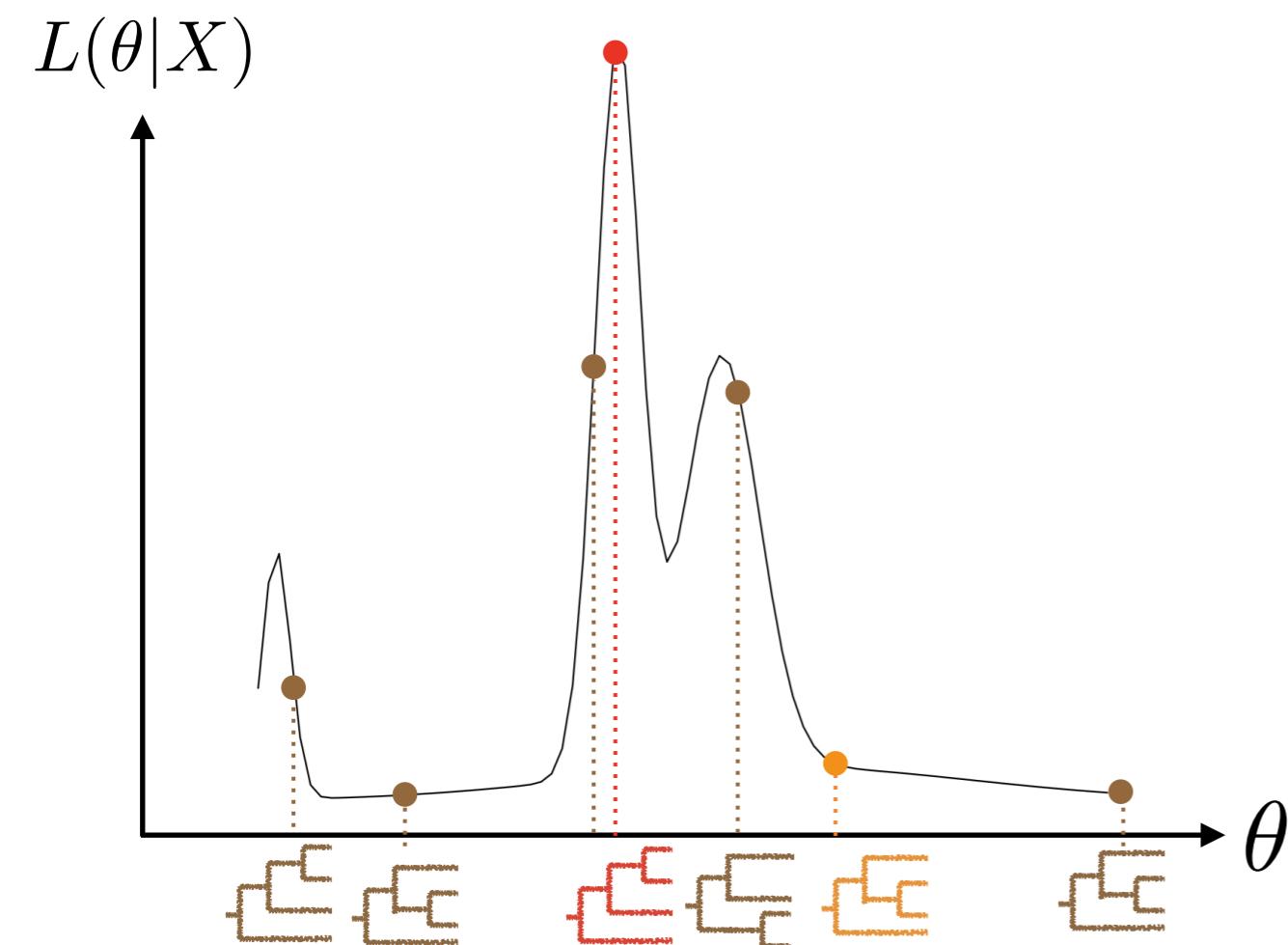
Tyler Chafin



- Parallel quartet likelihood computation
- Weights to quartets: deviations from expected CFs
- Sampling of quartets without biasing inference

**SNaQ 2.0
coming up soon!**

Identifiability

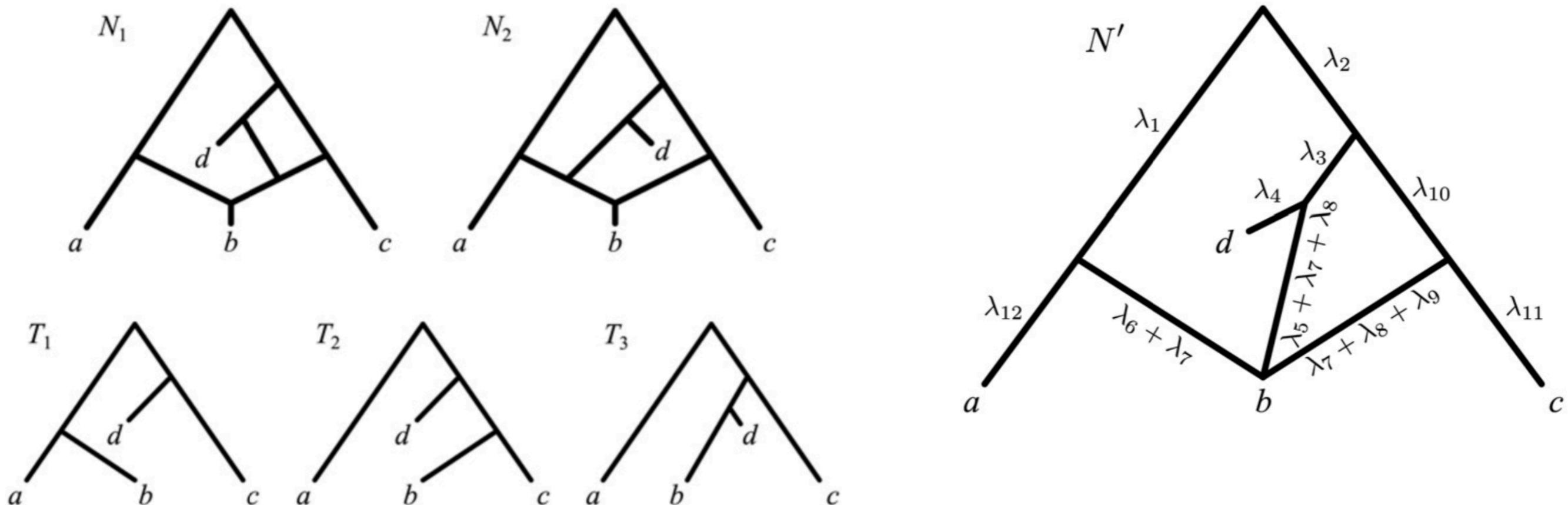


RESEARCH ARTICLE

Reconstructible Phylogenetic Networks: Do Not Distinguish the Indistinguishable

Fabio Pardi^{1,3*}, Celine Scornavacca^{2,3}

1 Laboratoire d’Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM, UMR 5506) CNRS, Université de Montpellier, France, **2** Institut des Sciences de l’Evolution de Montpellier (ISE-M, UMR 5554) CNRS, IRD, Université de Montpellier, France, **3** Institut de Biologie Computationnelle, Montpellier, France

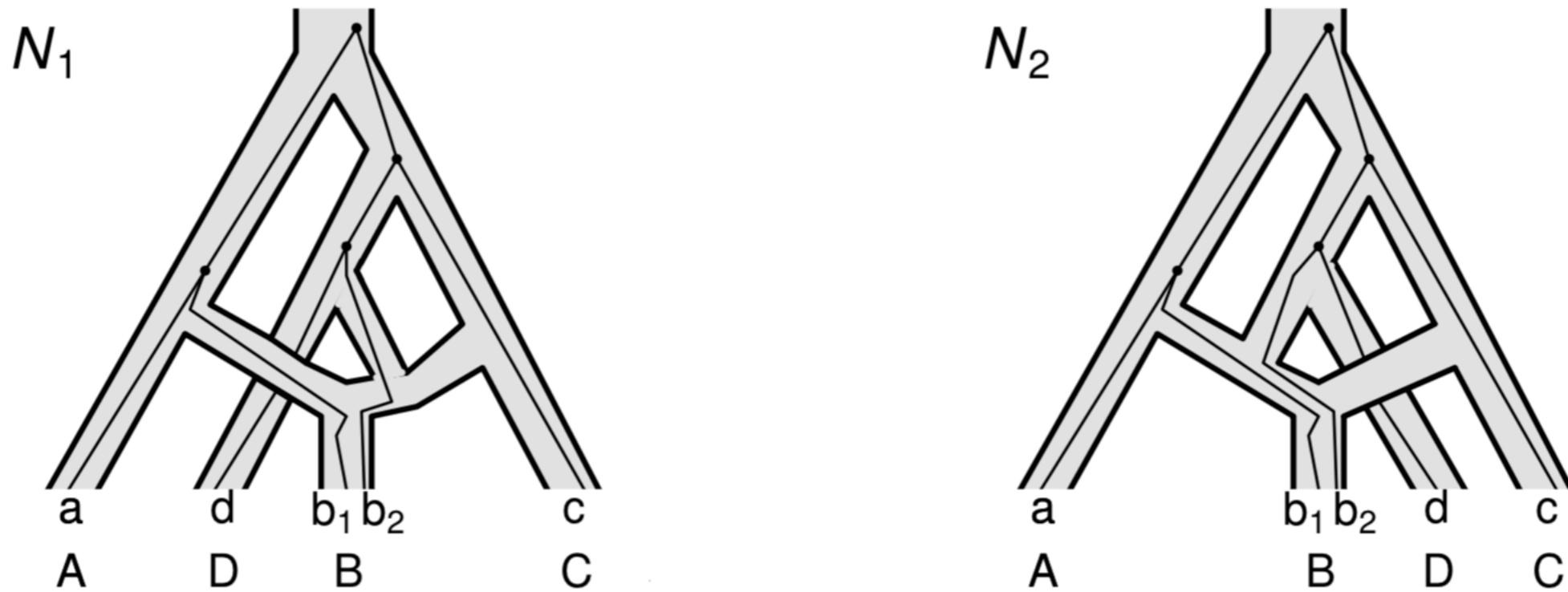


Undistinguishable with the
“displayed trees” criterion

Solution: Canonical
network (“unzipped”)

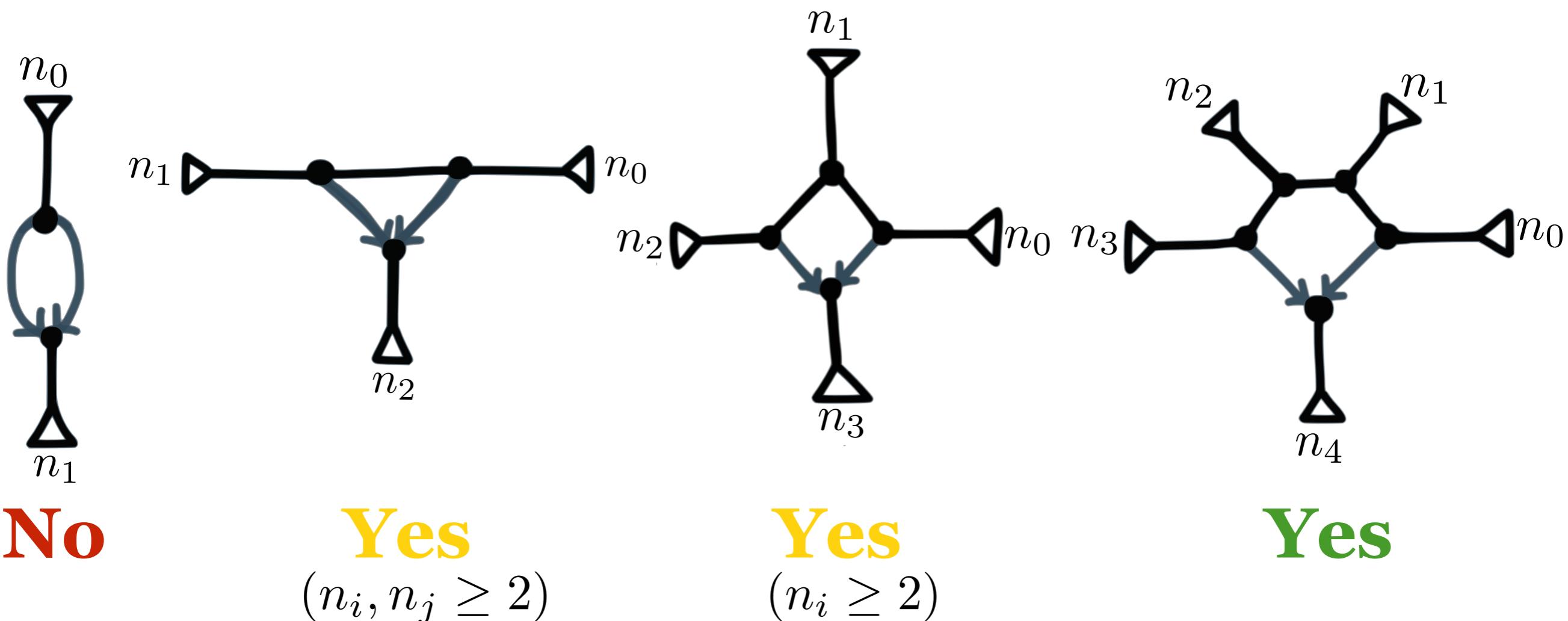
Displayed Trees Do Not Determine Distinguishability Under the Network Multispecies Coalescent

Sha Zhu¹, James H. Degnan²



Distinguishable under the MSC

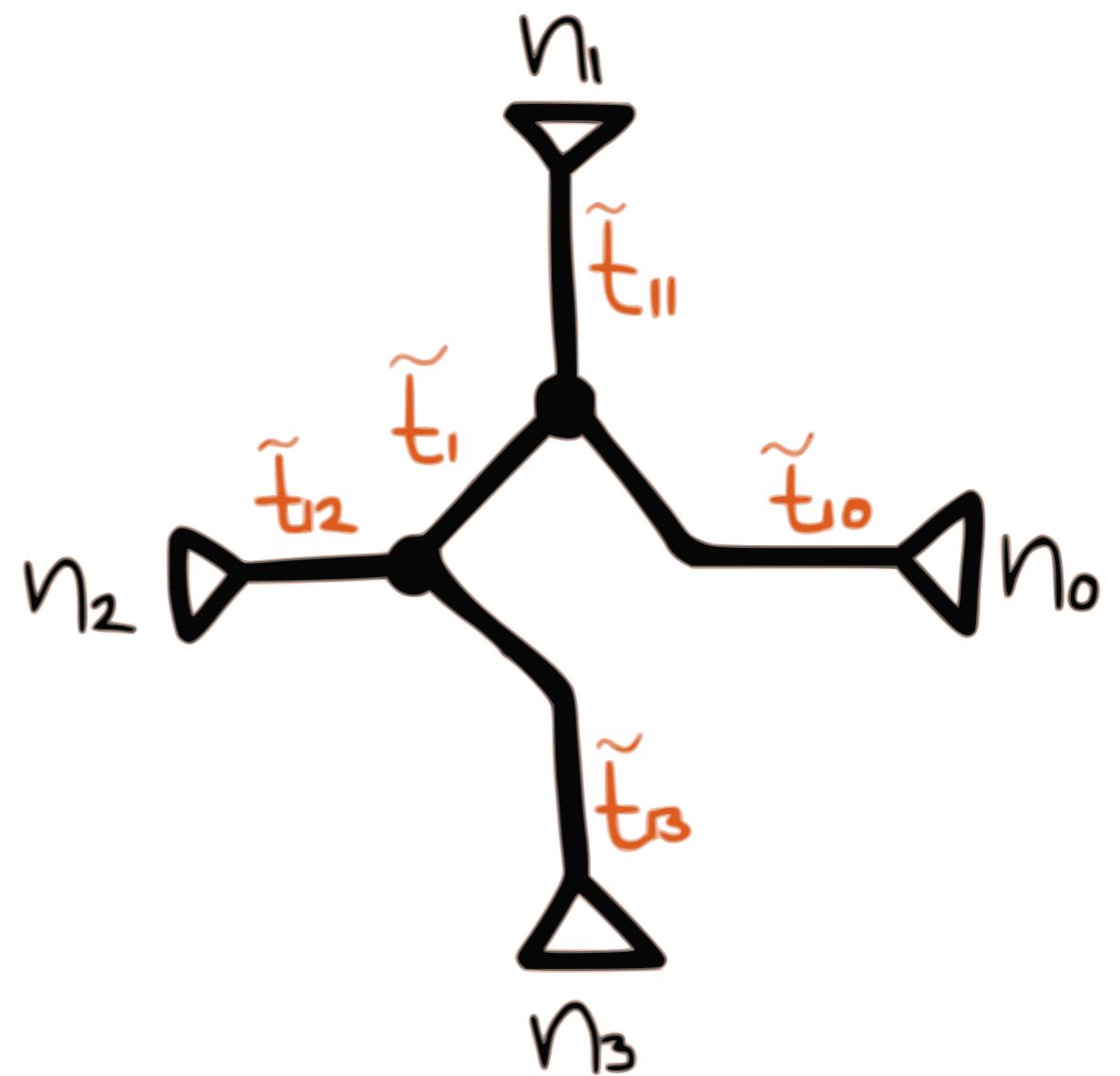
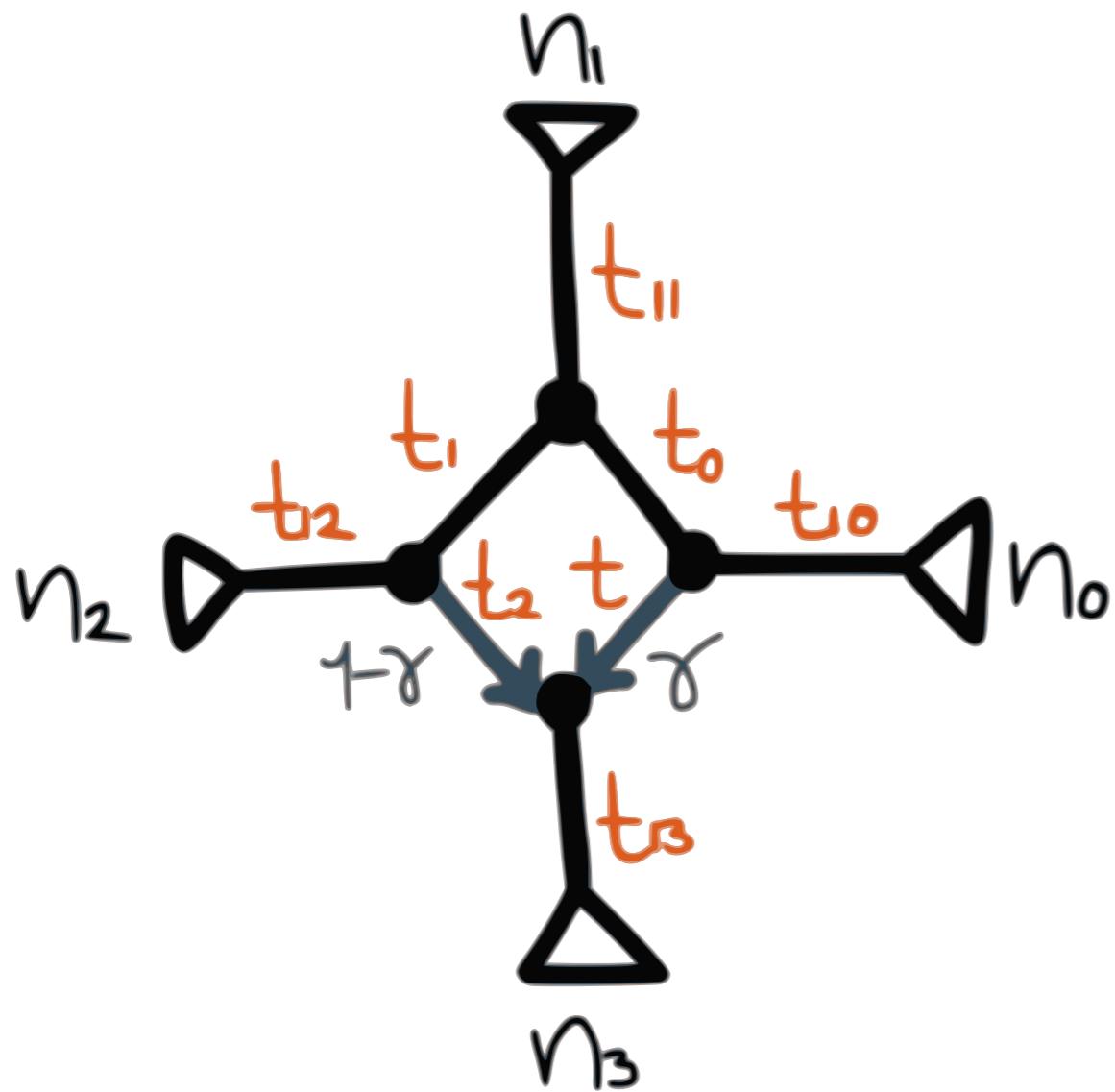
RESEARCH ARTICLE

Inferring Phylogenetic Networks with
Maximum Pseudolikelihood under
Incomplete Lineage SortingClaudia Solís-Lemus^{1*}, Cécile Ané^{1,2}Can we detect the
presence of
hybridization in level-1
networks?

Generic Identifiability

 $t_i \in (0, \infty), \gamma \in (0, 1)$

Idea of proof of identifiability: hybridization



System of equations

$\{\text{CF}_{\text{network}}\}$

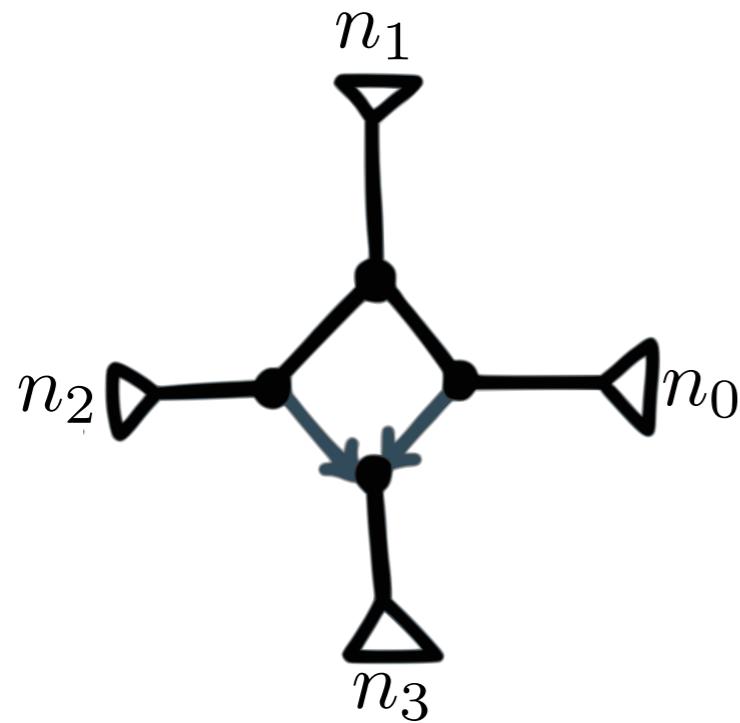
(Solís-Lemus & Ané, 2016;
Solís-Lemus et al, 2020)

System of equations

$\{\text{CF}_{\text{tree}}\}$

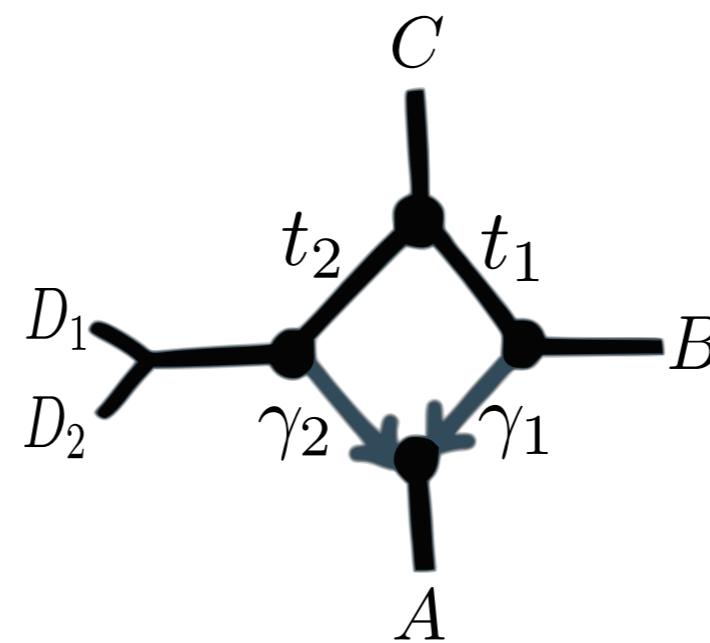
Can we detect the presence of hybridization in level-1 networks?

In theory

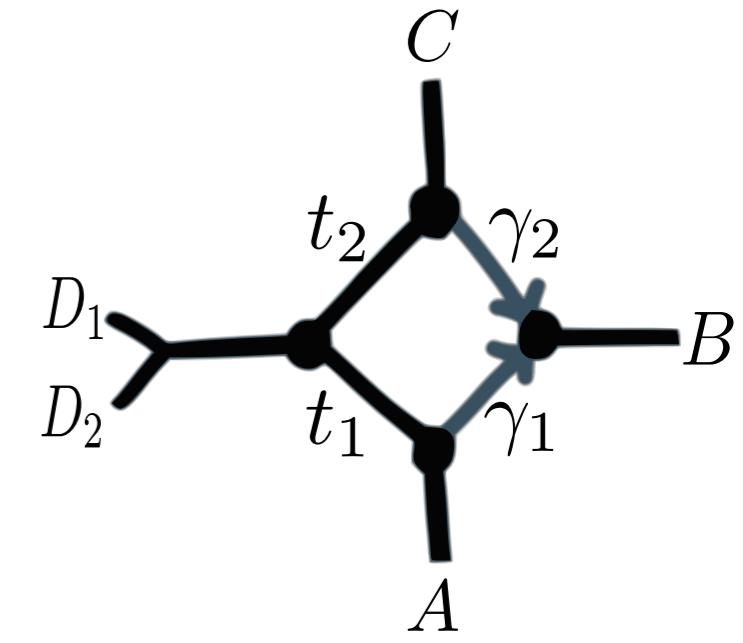


Yes
 $(n_i \geq 2)$

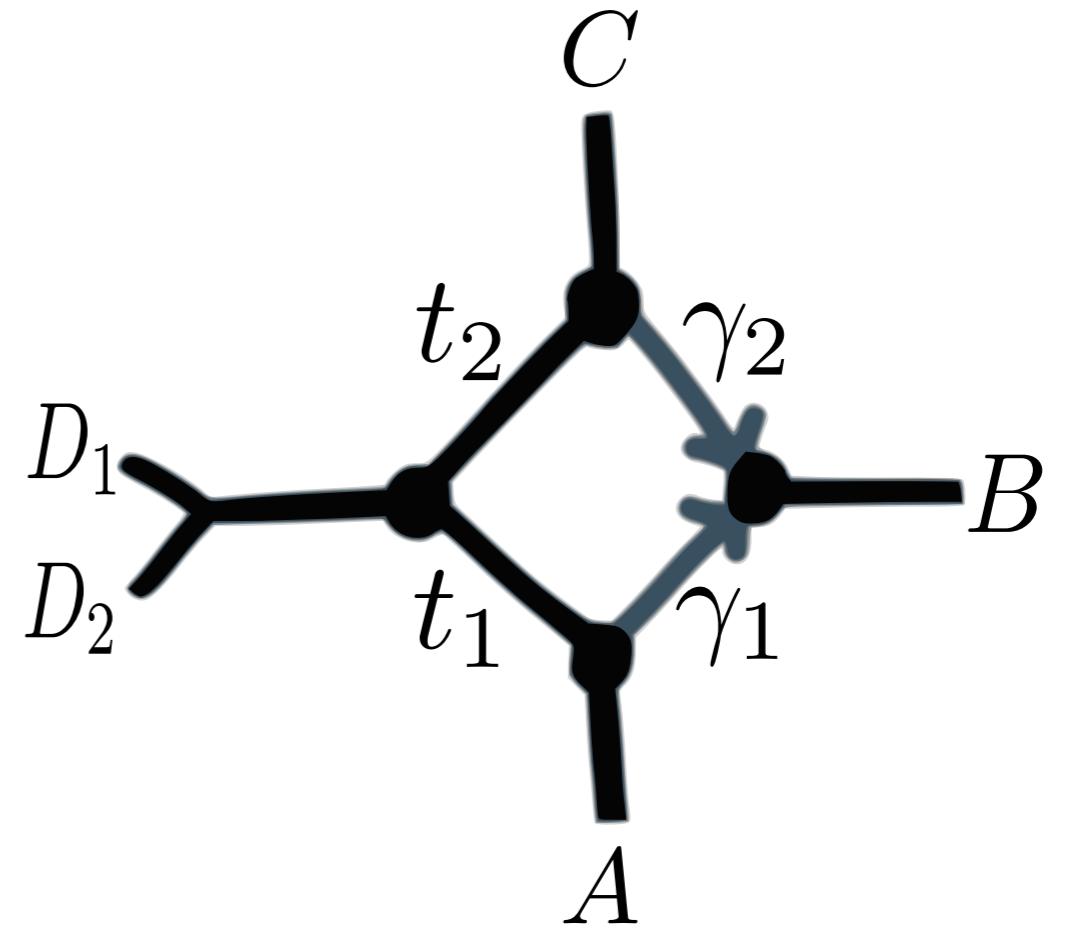
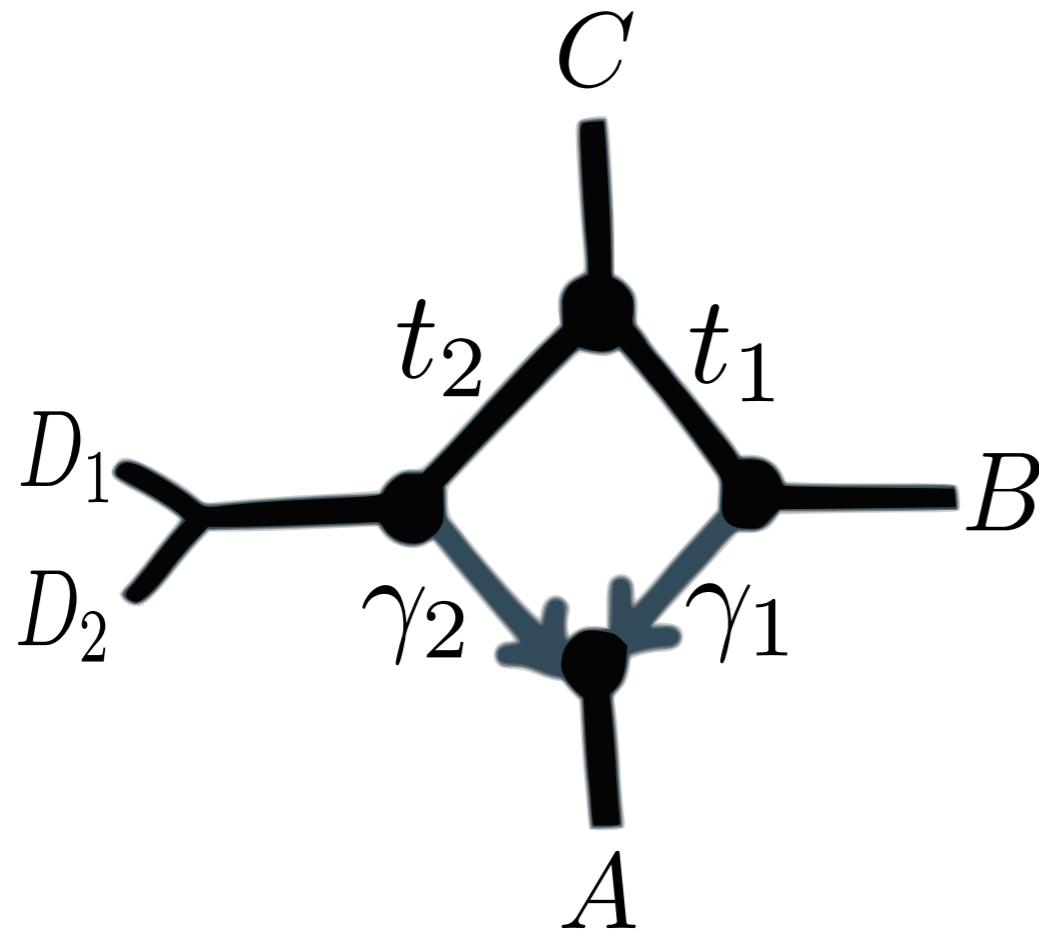
In practice



Sometimes



In practice: flat pseudolikelihood



(S.-L., Ané, 2016, PLoS Genetics)

snaQ Diamond identifiability

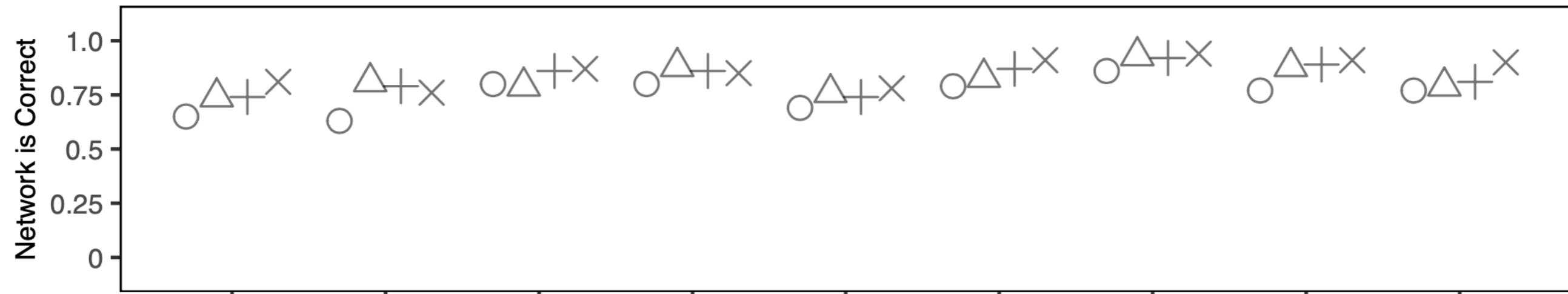


George Tiley

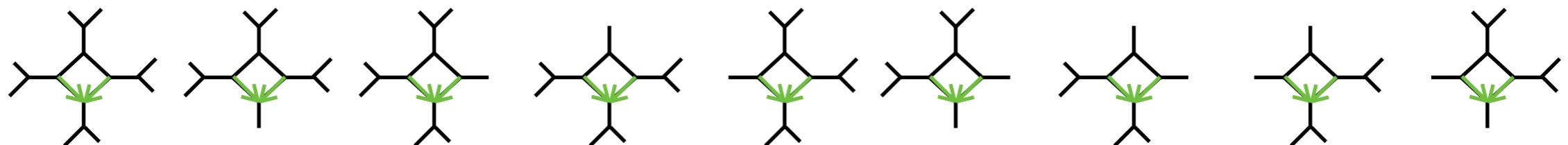
nTrees ○ 100 △ 500 + 1000 × 5000



snaQ

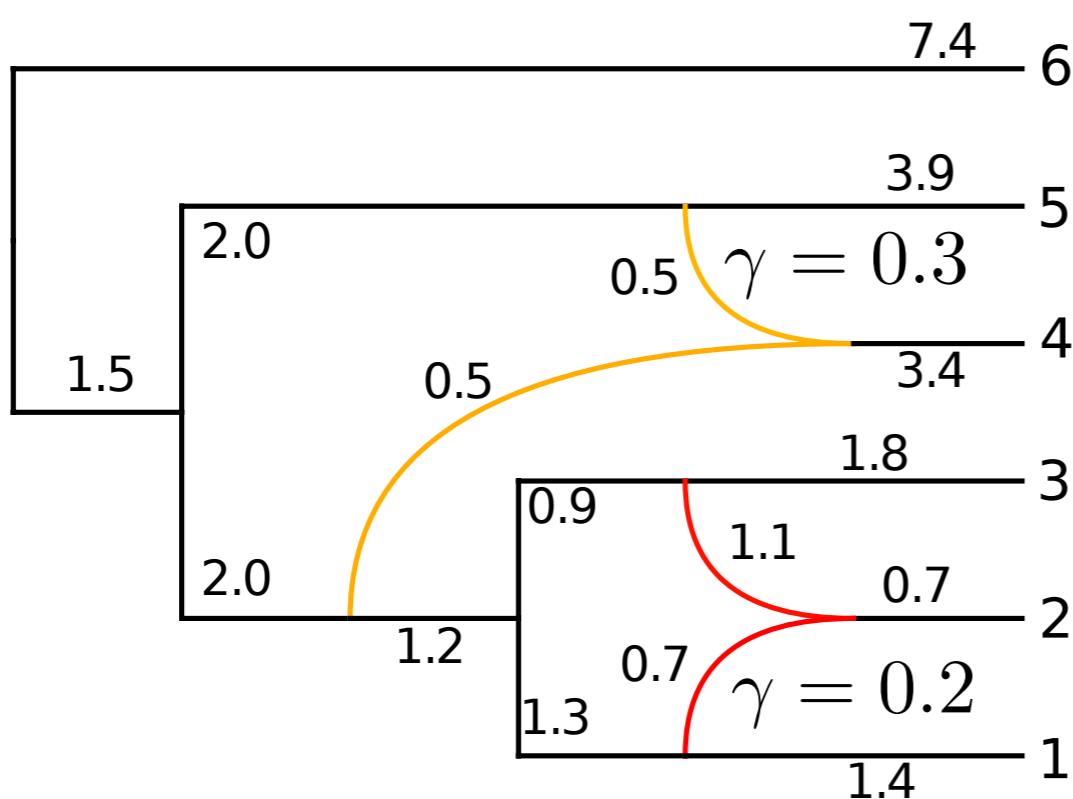


PhyloNet

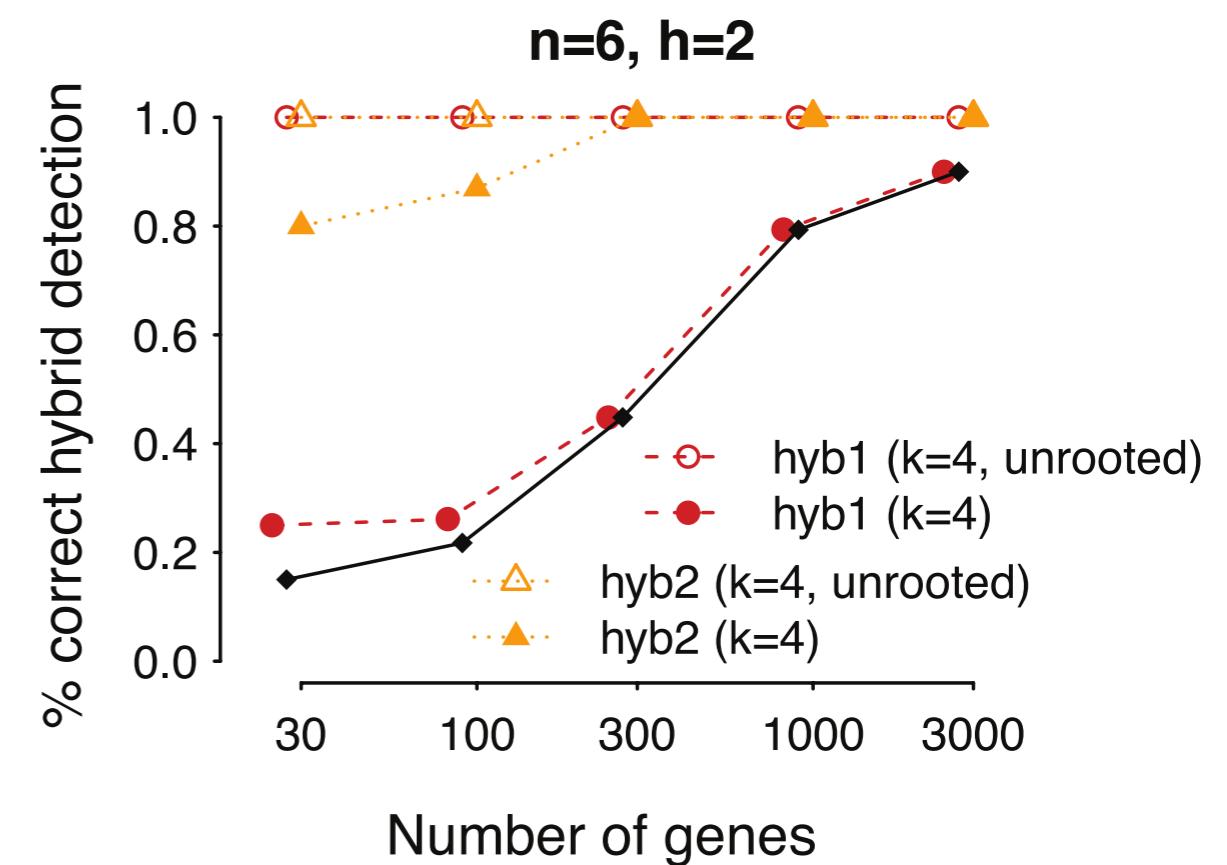


Identifiability matters: SNaQ performance

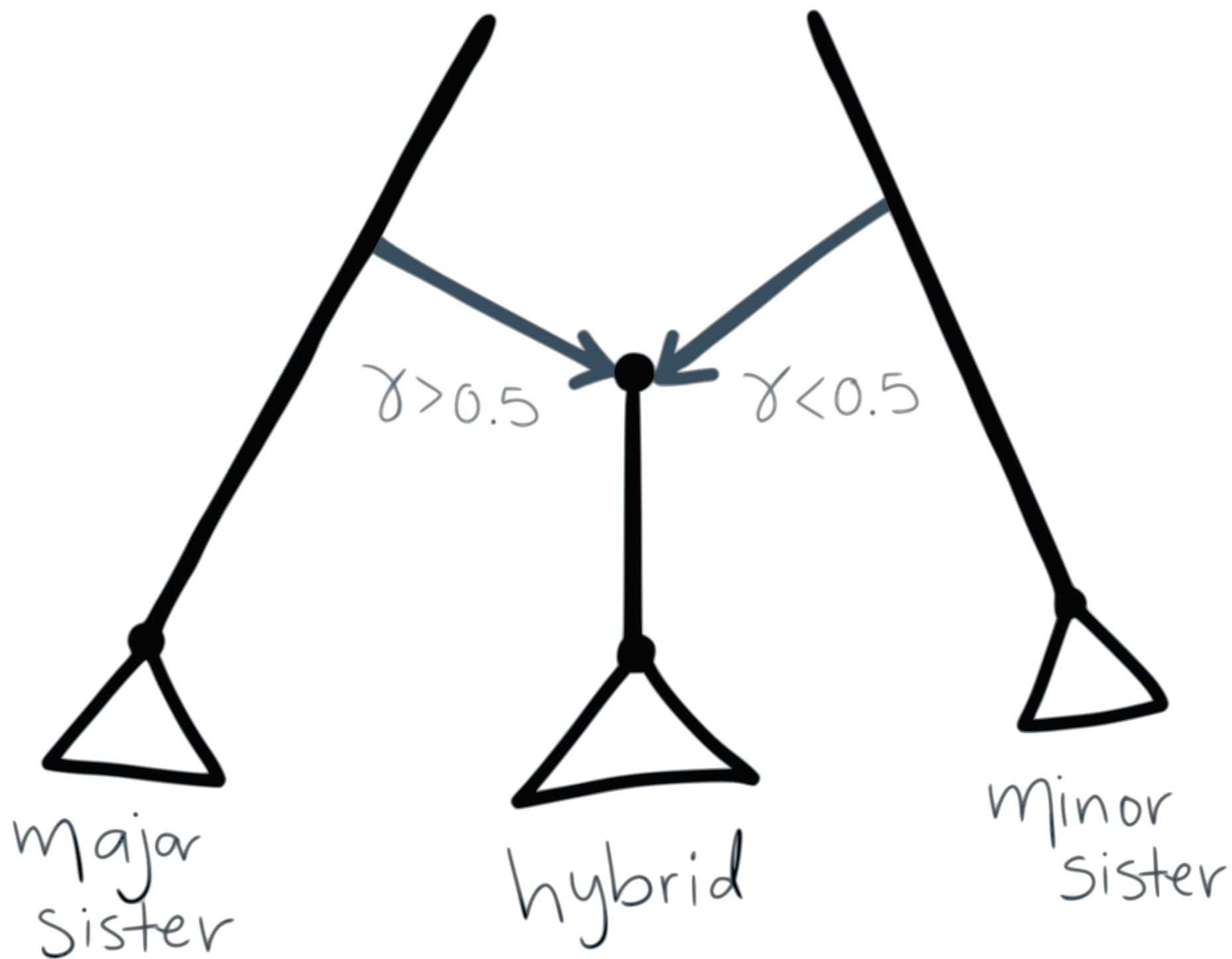
Good diamond



Bad diamond

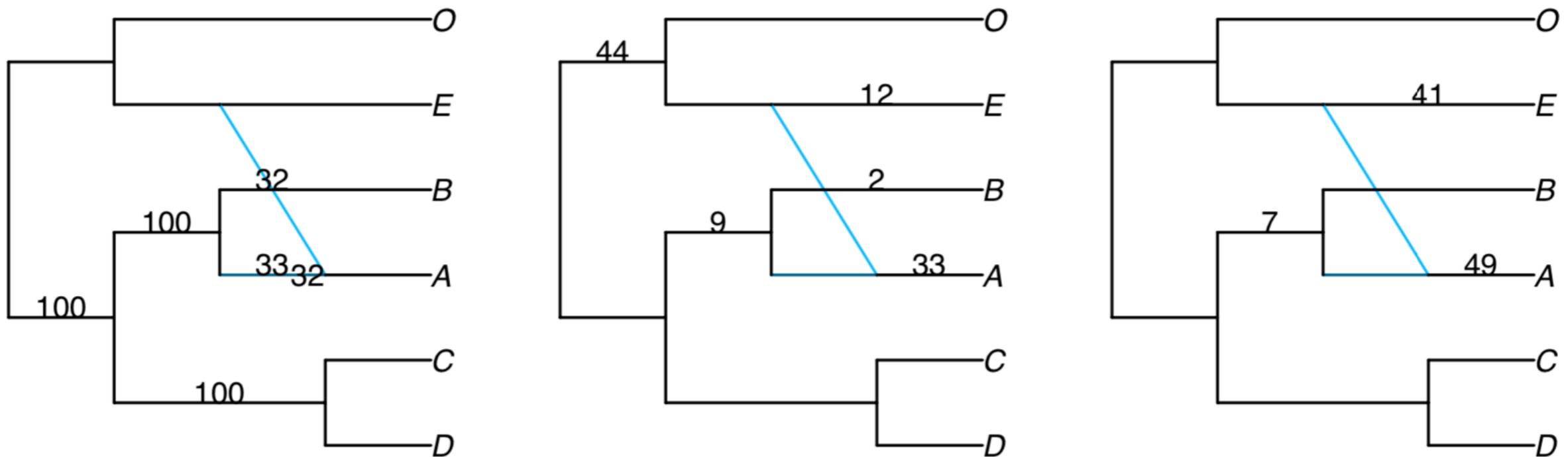


Network Comparison



(S.-L. et al, 2017, MBE)

Network summary

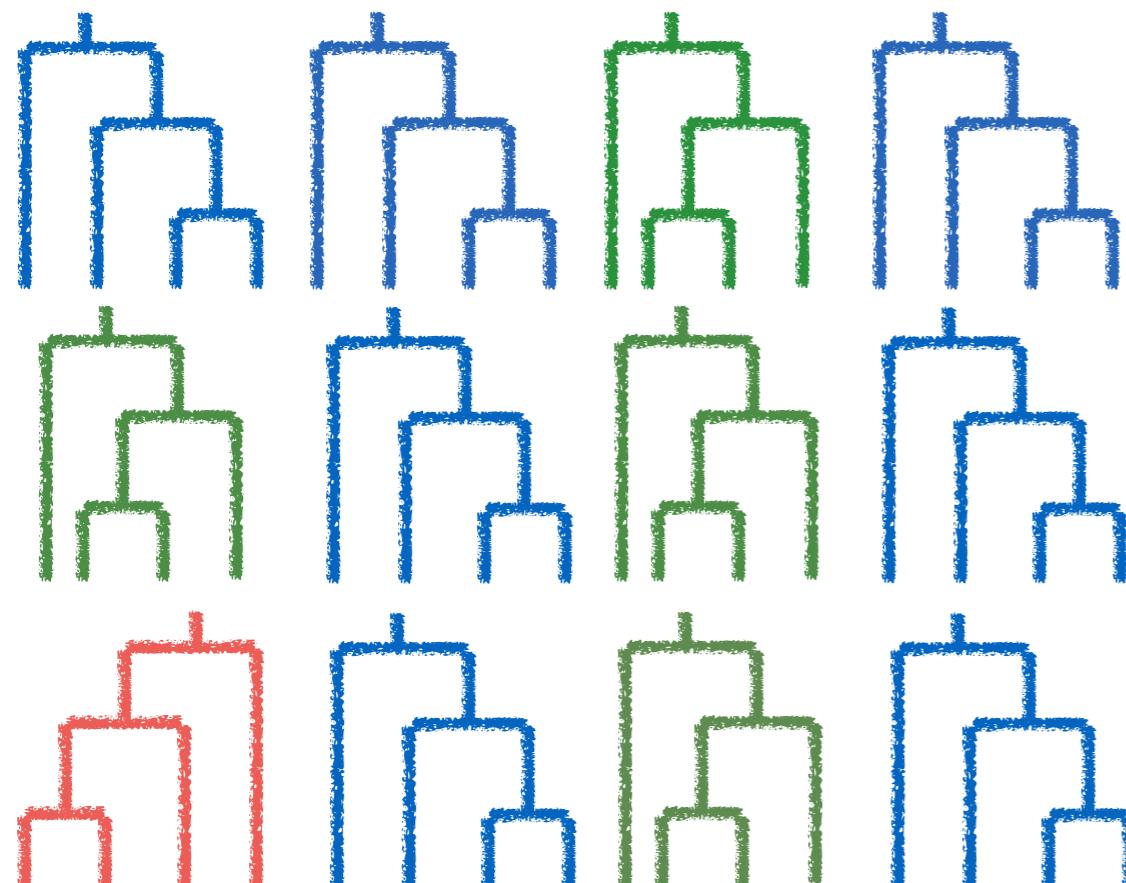
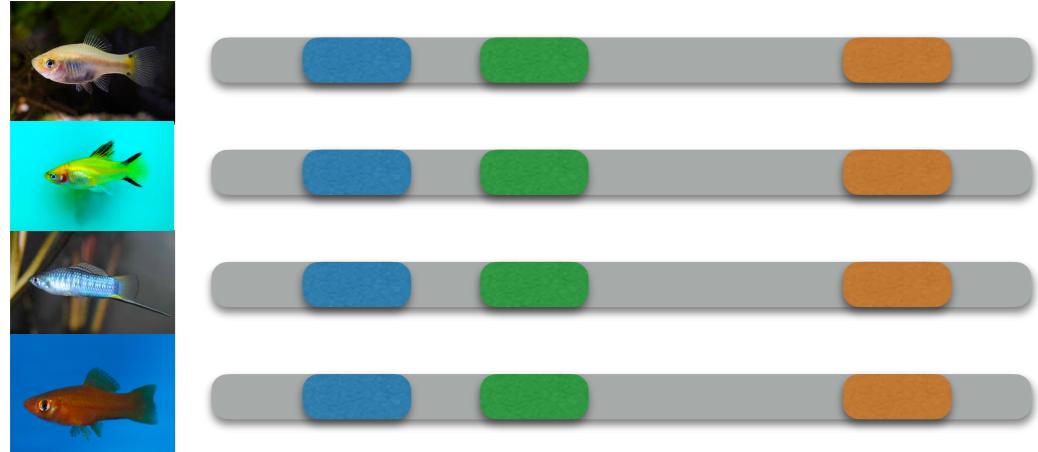


Hybrid
clades

Minor
sister
clades

When?

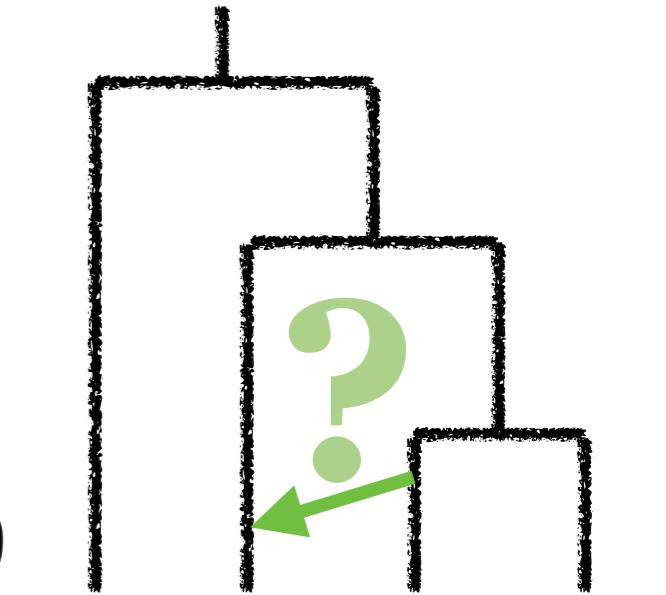
Phylogenetic network



Data

Goodness-of-fit test
Hypothesis test:
Is a tree a good fit?

TICR
→
GitHub



<https://github.com/nstenz/TICR>
(Stenz et al, 2015, Syst Bio)

Practical advice

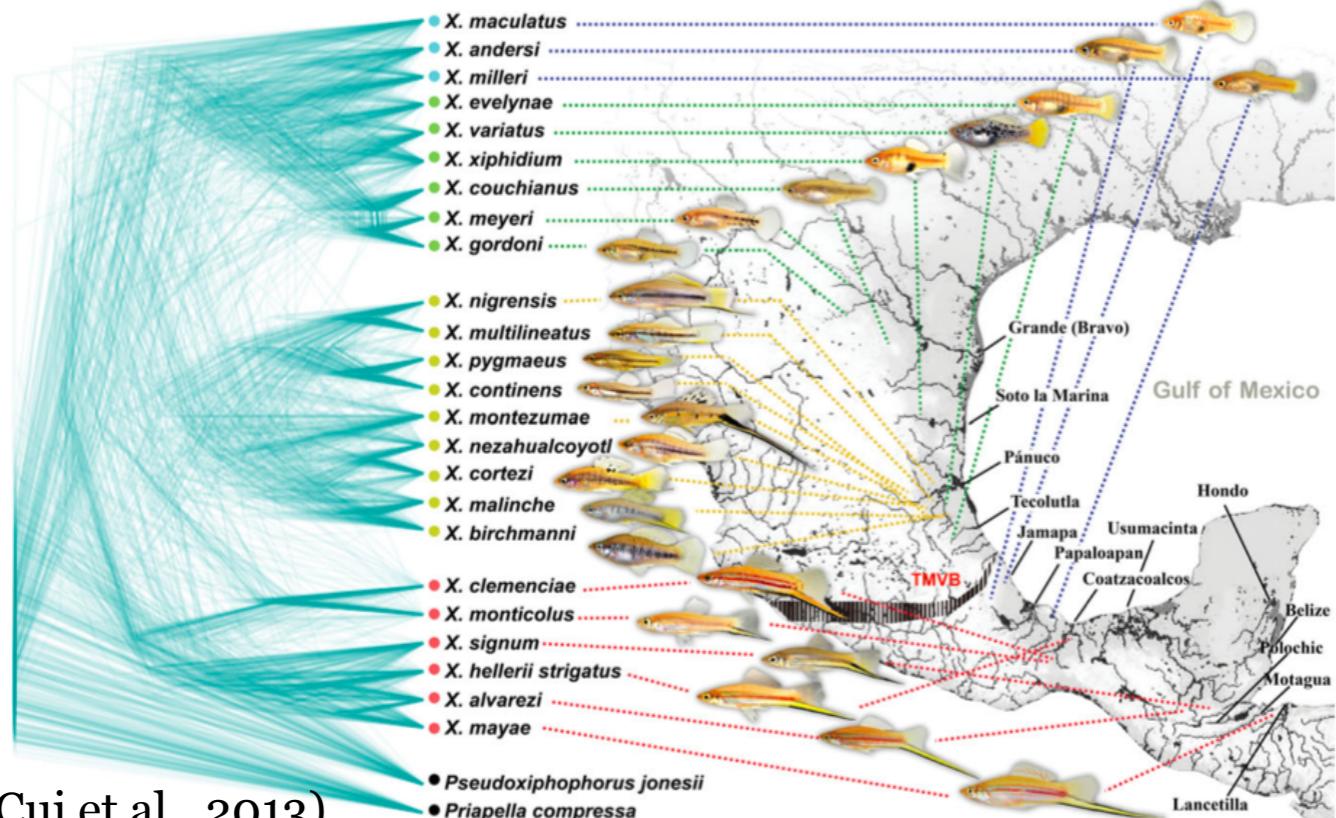
- Do multiple runs
- Do bootstrap
- Check the .networks output file (especially if hybridization conflicts with outgroup)
 - Choice of outgroup important!
- Interpretation of ghost lineages
- What is the quality of my input data (gene trees/CFs)?
- Run SNaQ sequentially: $h=0, h=1, h=2, \dots$

Practical advice

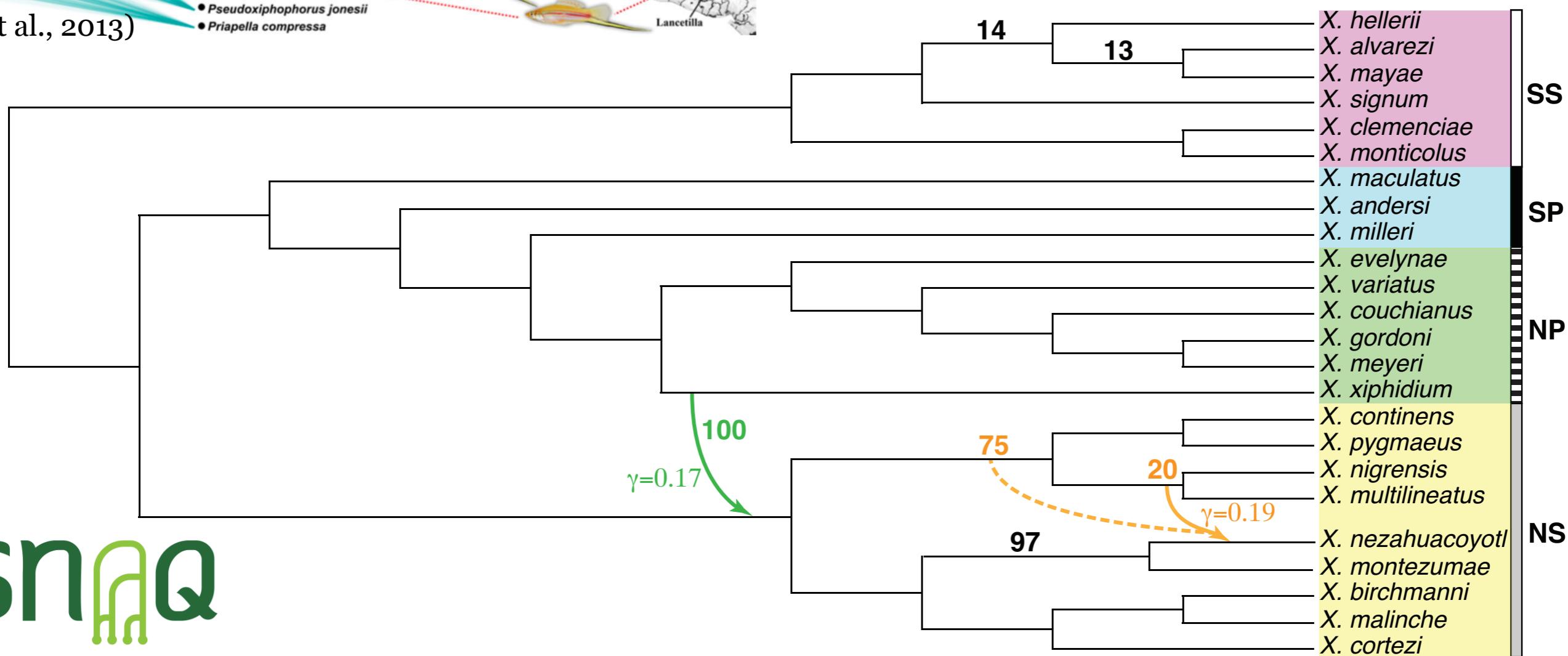
- Do multiple runs
- Do bootstrap
- Check the .networks output file (especially if hybridization conflicts with outgroup)
 - Choice of outgroup important!
- Interpretation of ghost lineages
- What is the quality of my input data (gene trees/CFs)?
- Run SNaQ sequentially: $h=0, h=1, h=2, \dots$

When to stop?
(Cai and Ané, 2020)

1183 genes, 24 swordtails and platyfish



Xiphophorus fish data

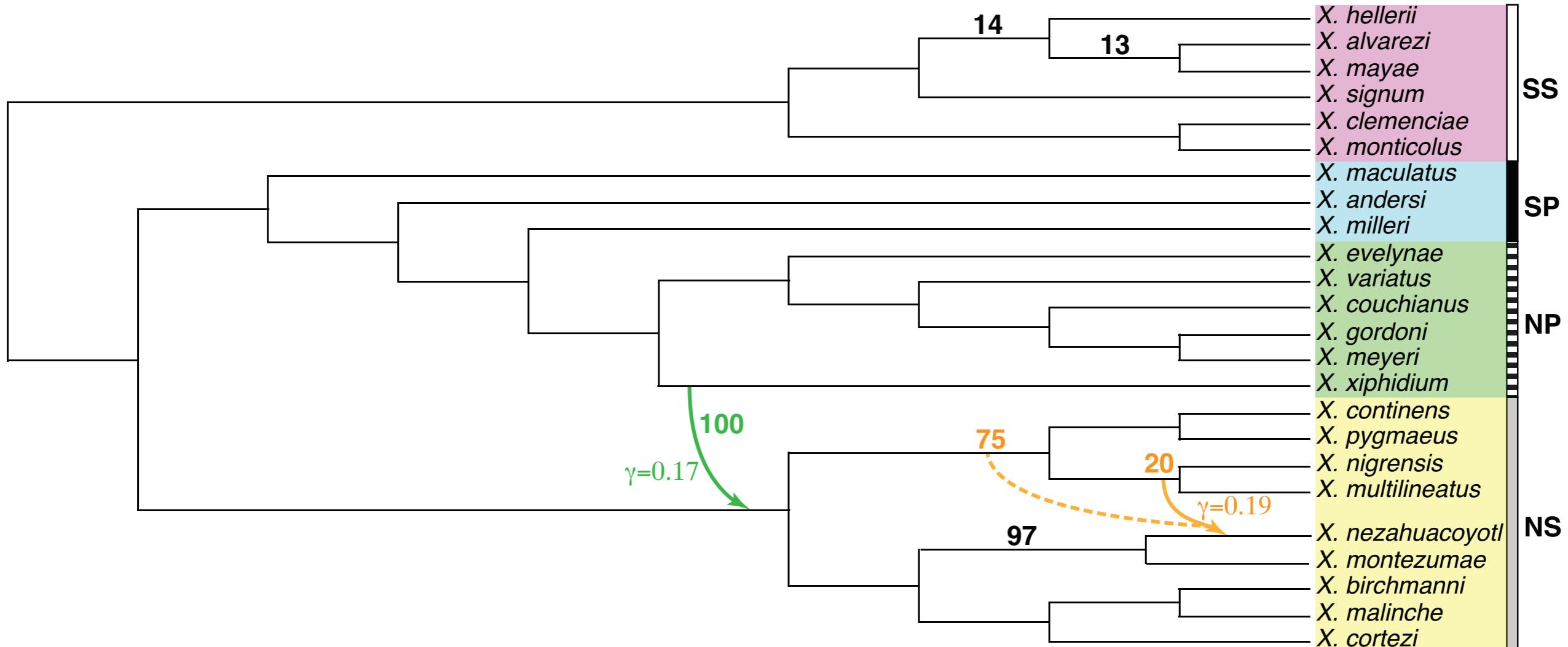


snaQ

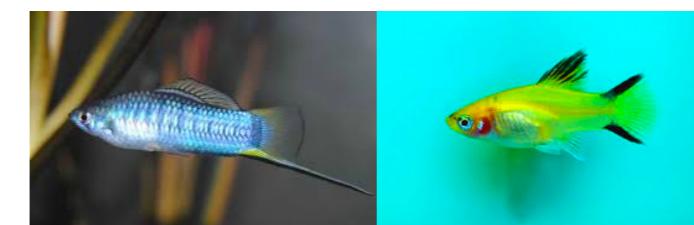
(Solís-Lemus, Ané, 2016, PLoS Genetics)

Part II

I have the network, now what?

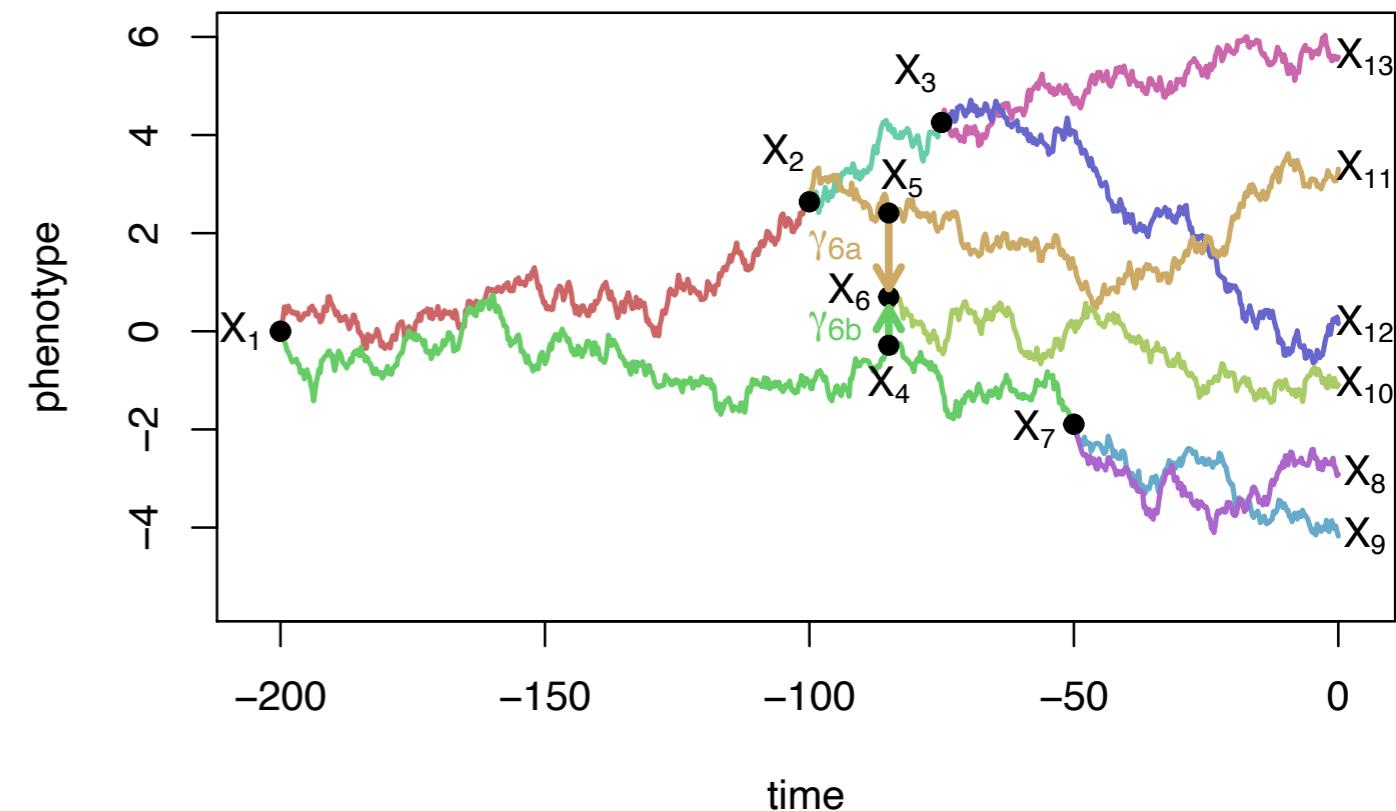
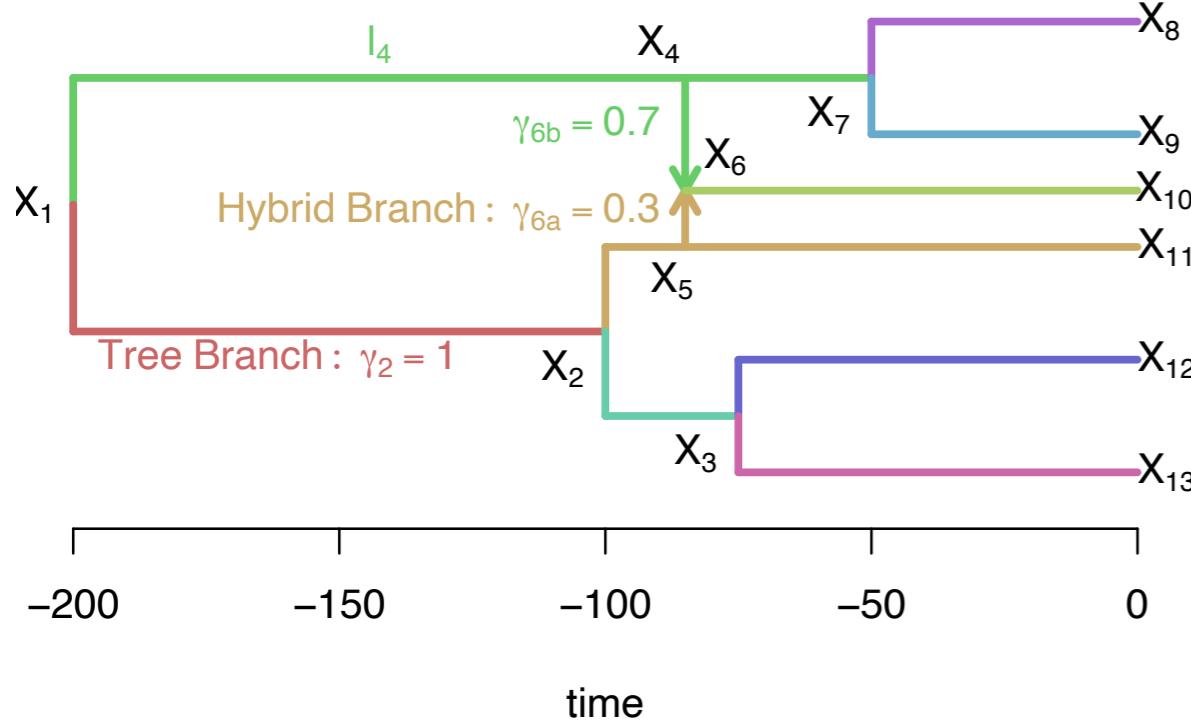


- Sword index
- Female preference



(Cui et al., 2013)
(Solís-Lemus, Ané, 2016, PLoS Genetics)

Trait models of evolution in networks



Brownian Motion
+ weighted
average in hybrid

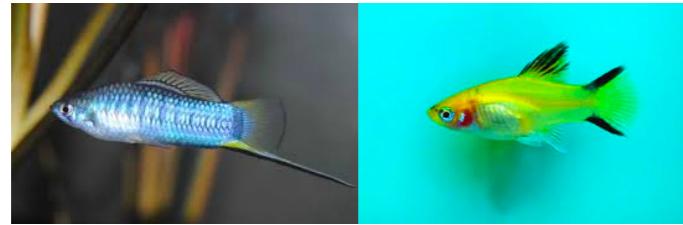
$$X_h = \gamma_1 X_{p_1} + \gamma_2 X_{p_2}$$

(Bastide et al, 2018, Syst Bio)

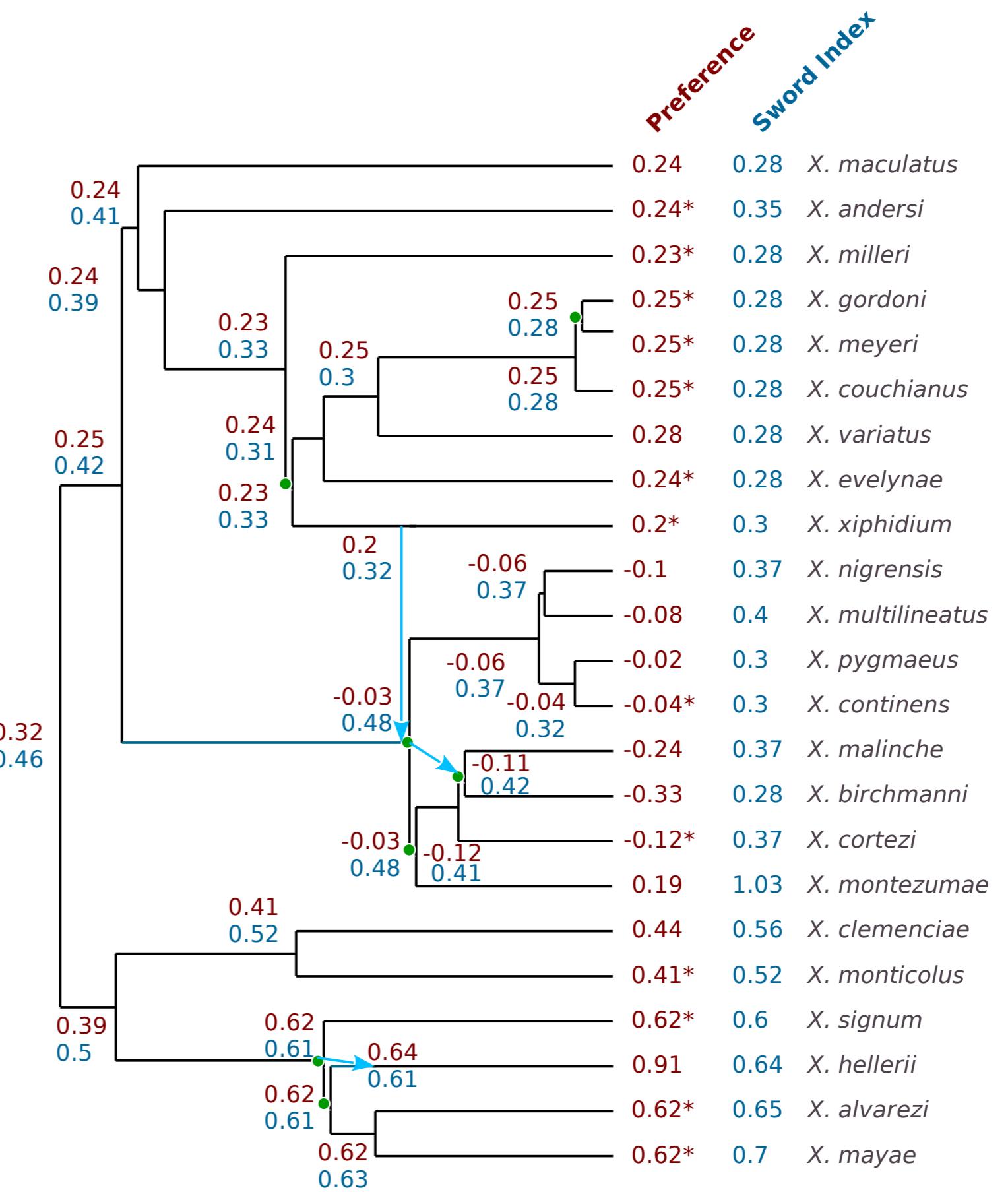
$$\mathbf{X} \sim N(X_{root}, \sigma^2 \mathbf{V})$$

- Phylogenetic signal
- Ancestral reconstruction
- Phylogenetic regression
- Phylogenetic ANOVA

- Sword index
- Female preference

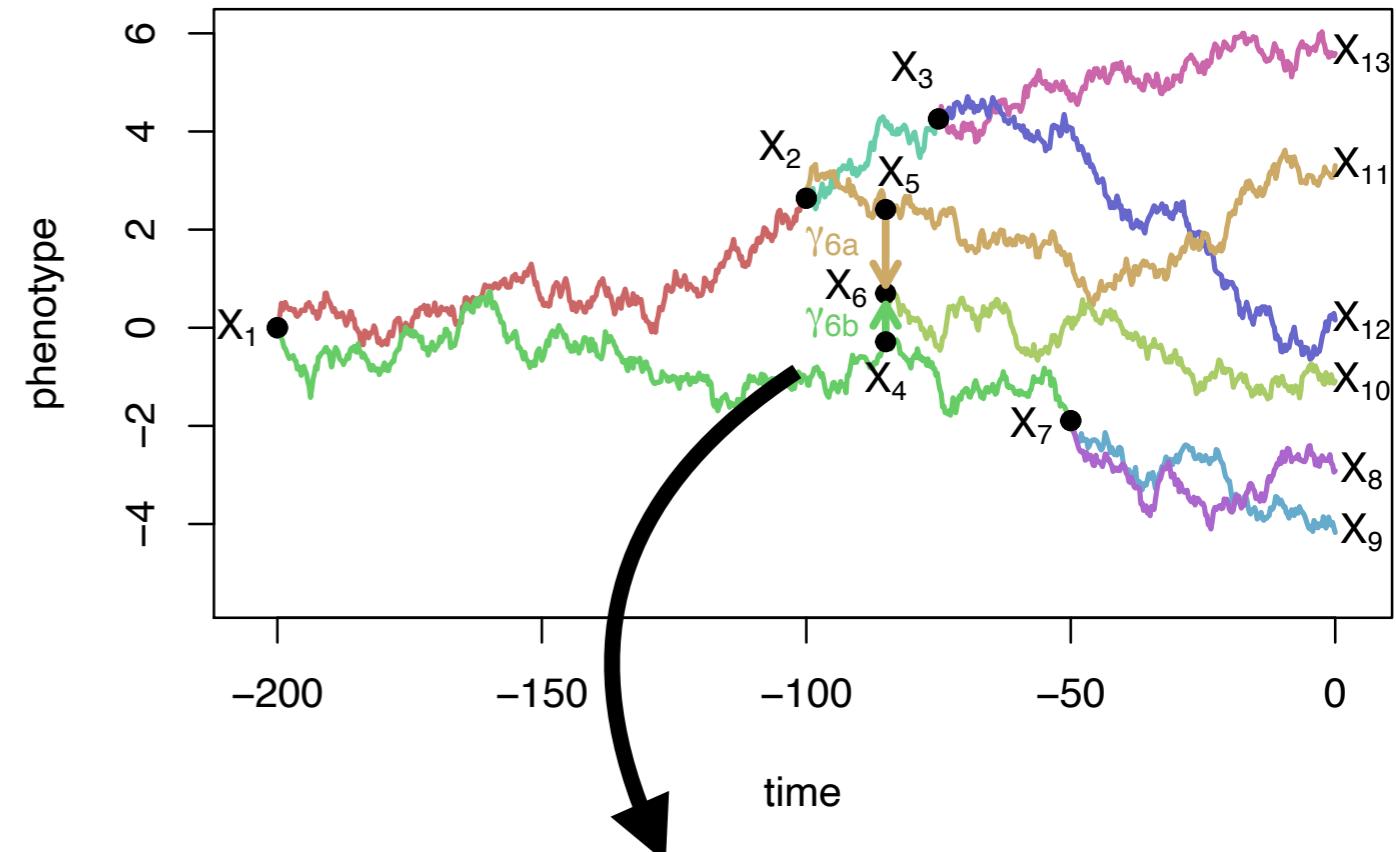
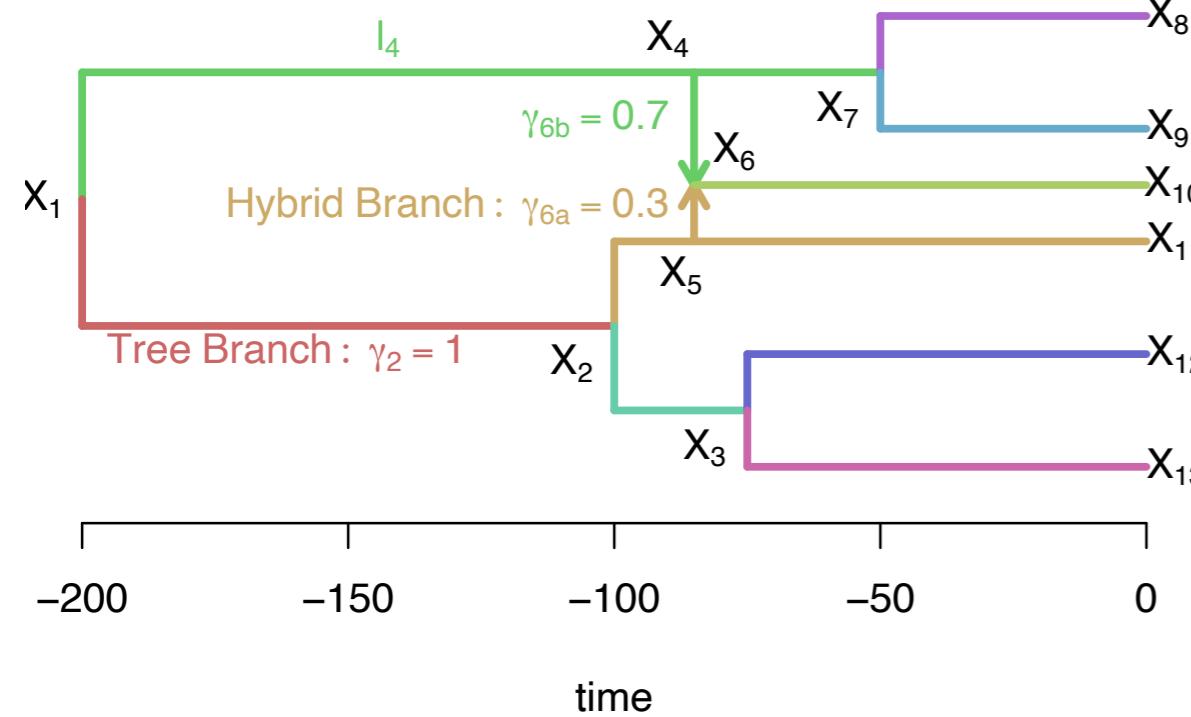


- **Ancestral reconstruction:** common ancestor likely had sword
- **Phylogenetic regression:** positive association between sword index and female preference but not significant ($p = 0.106$)



Test for transgressive evolution

$$X_h = \gamma_1 X_{p_1} + \gamma_2 X_{p_2} + \Delta_h$$



$\Delta_h = 0$ No transgressive evolution

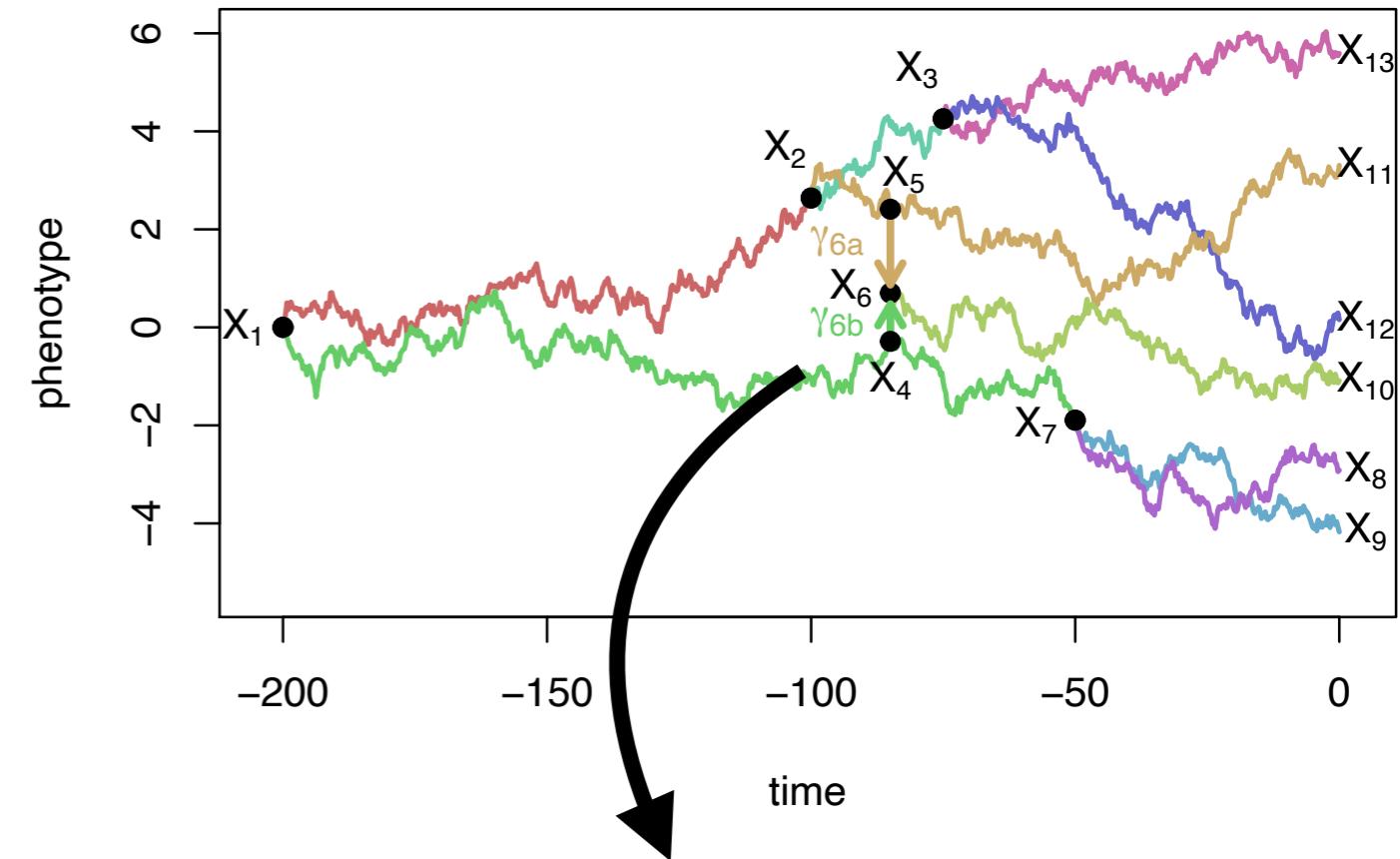
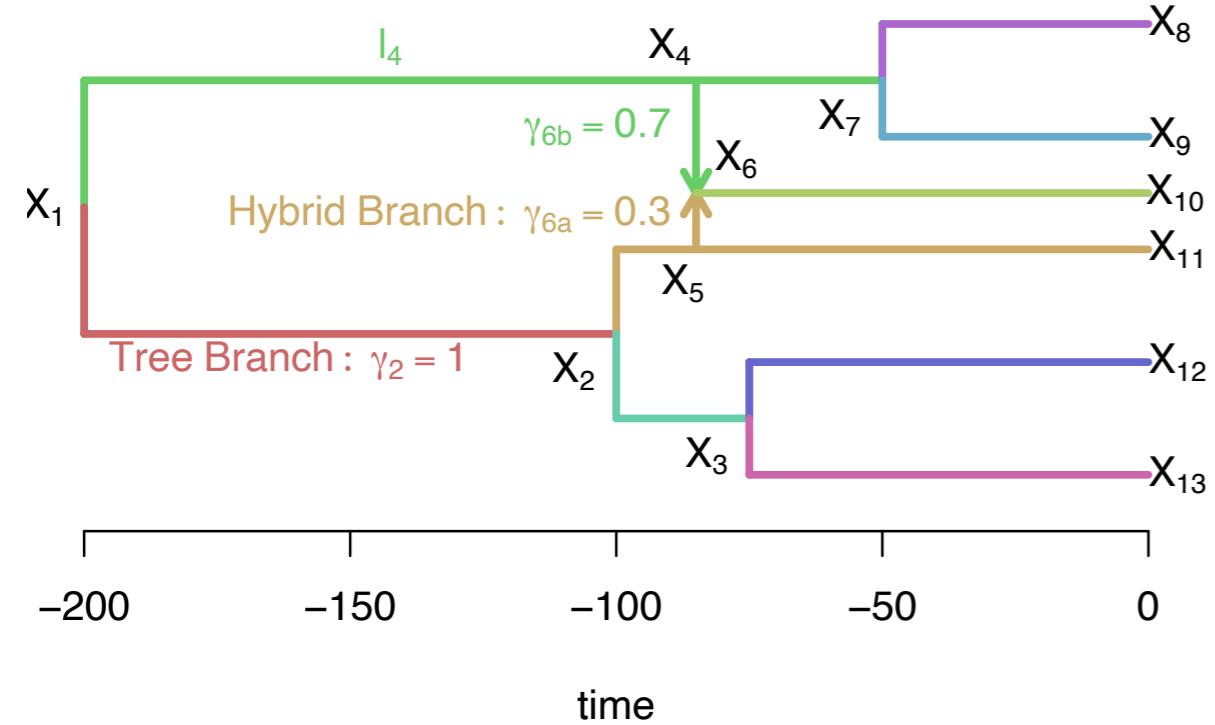
$\Delta_h = \Delta$ Single-effect transgressive evolution

Δ_h Multi-effect transgressive evolution

F tests

Hybrid value:
shift from
parents range

Test for transgressive evolution



- Sword index: $p=0.55$
- Female preference: $p=0.0064$

Hybrid value:
shift from
parents range

PhyloNetworks: analysis for phylogenetic networks

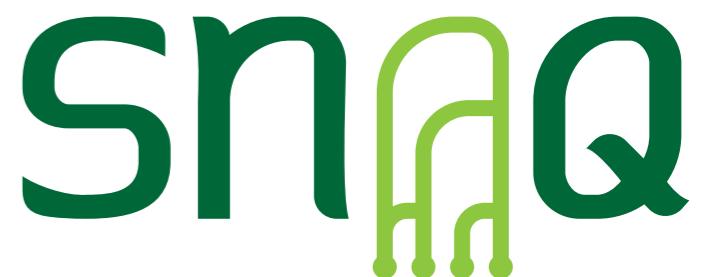
build passing docs stable docs dev codecov 81% coverage 67%

Overview



PhyloNetworks is a [Julia](#) package with utilities to:

- read / write phylogenetic trees and networks, in (extended) Newick format. Networks are considered explicit: nodes represent ancestral species. They can be rooted or unrooted.
- manipulate networks: re-root, prune taxa, remove hybrid edges, extract the major tree from a network, extract displayed networks / trees
- compare networks / trees with dissimilarity measures (Robinson-Foulds distance on trees)
- summarize samples of bootstrap networks (or trees) with edge and node support
- estimate species networks from multilocus data (see below)
- phylogenetic comparative methods for continuous trait evolution on species networks / trees



- Step-by-step tutorial
- Online documentation
- Google user group



(Solis-Lemus & Ane, 2016; Solis-Lemus. et al, 2017)



<https://solislemuslab.github.io/>



mstdn.social/@solislemuslab



crsl4



@thestatistician

PhyloNetworks: analysis for phylogenetic networks

build passing docs stable docs dev codecov 81% coverage 67%

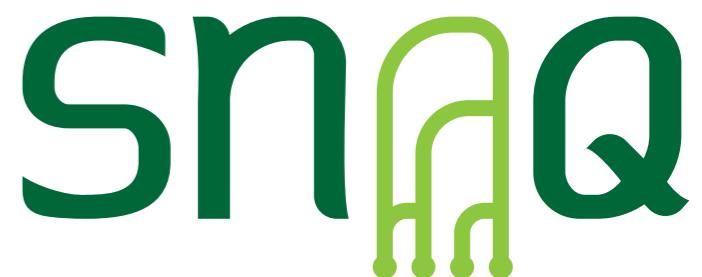
Overview

Get your stickers!



PhyloNetworks is a [Julia](#) package with utilities to:

- read / write phylogenetic trees and networks, in (extended) Newick format. Networks are considered explicit: nodes represent ancestral species. They can be rooted or unrooted.
- manipulate networks: re-root, prune taxa, remove hybrid edges, extract the major tree from a network, extract displayed networks / trees
- compare networks / trees with dissimilarity measures (Robinson-Foulds distance on trees)
- summarize samples of bootstrap networks (or trees) with edge and node support
- estimate species networks from multilocus data (see below)
- phylogenetic comparative methods for continuous trait evolution on species networks / trees



- Step-by-step tutorial
- Online documentation
- Google user group



(Solis-Lemus & Ane, 2016; Solis-Lemus. et al, 2017)



<https://solislemuslab.github.io/>



<mstdn.social/@solislemuslab>



crsl4



@thestatistician

Solis-Lemus lab



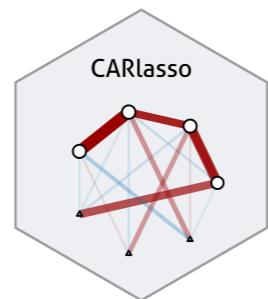
WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON



Thank you!



Bayesian
NetworkRegression.jl



snaQ

phylo
networks



<https://solislemuslab.github.io/>



<mstdn.social/@solislemuslab>



[solislemuslab](#)



@thestatistician

Questions?