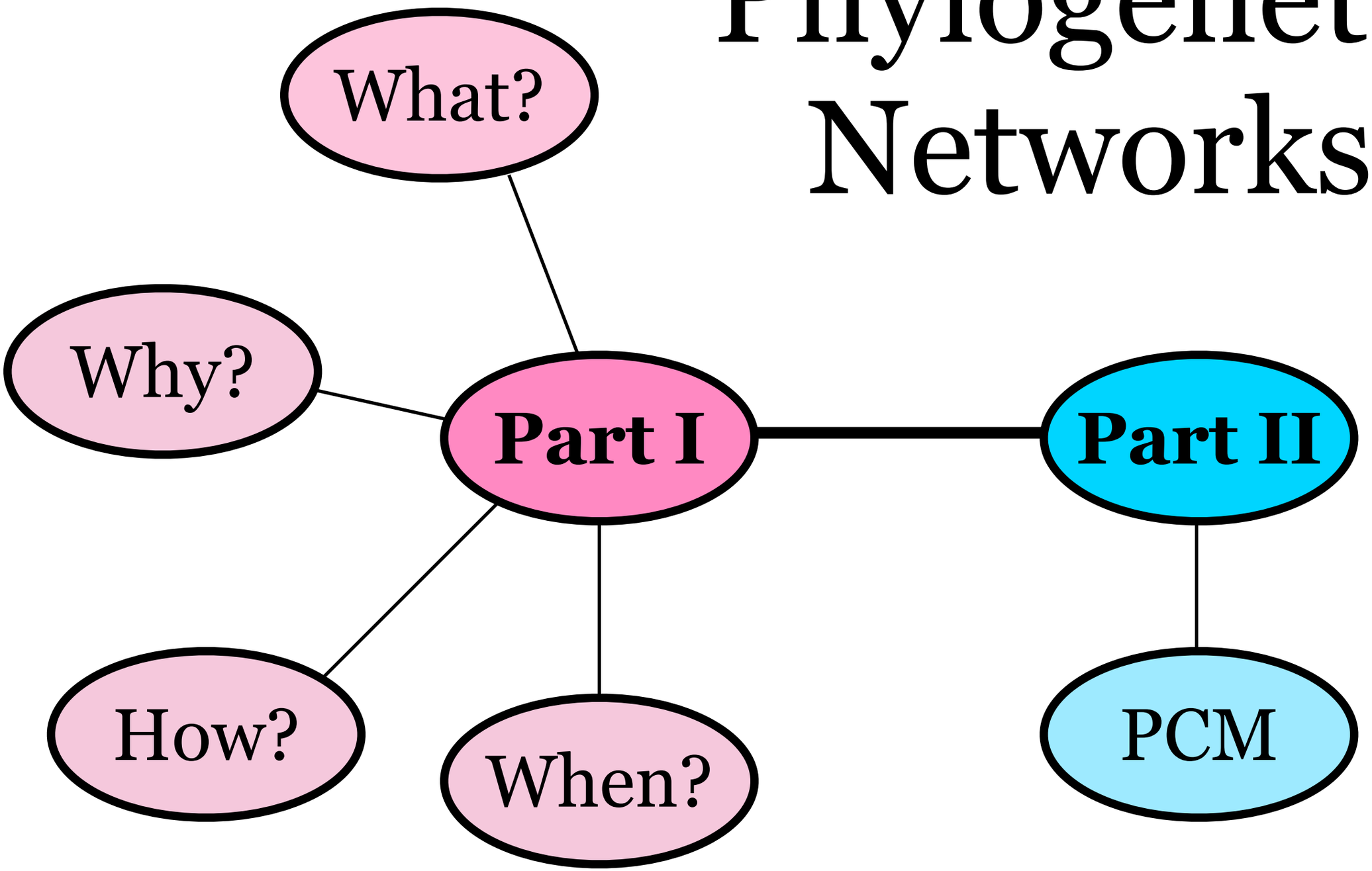
Statistical models on phylogenetic networks

Claudia Solís-Lemus, PhD
University of Wisconsin-Madison
Wisconsin Institute for Discovery
Department of Plant Pathology

May 30, 2023

https://solislemuslab.github.io/ mstdn.social/@solislemuslab crsl4 @thestatistician

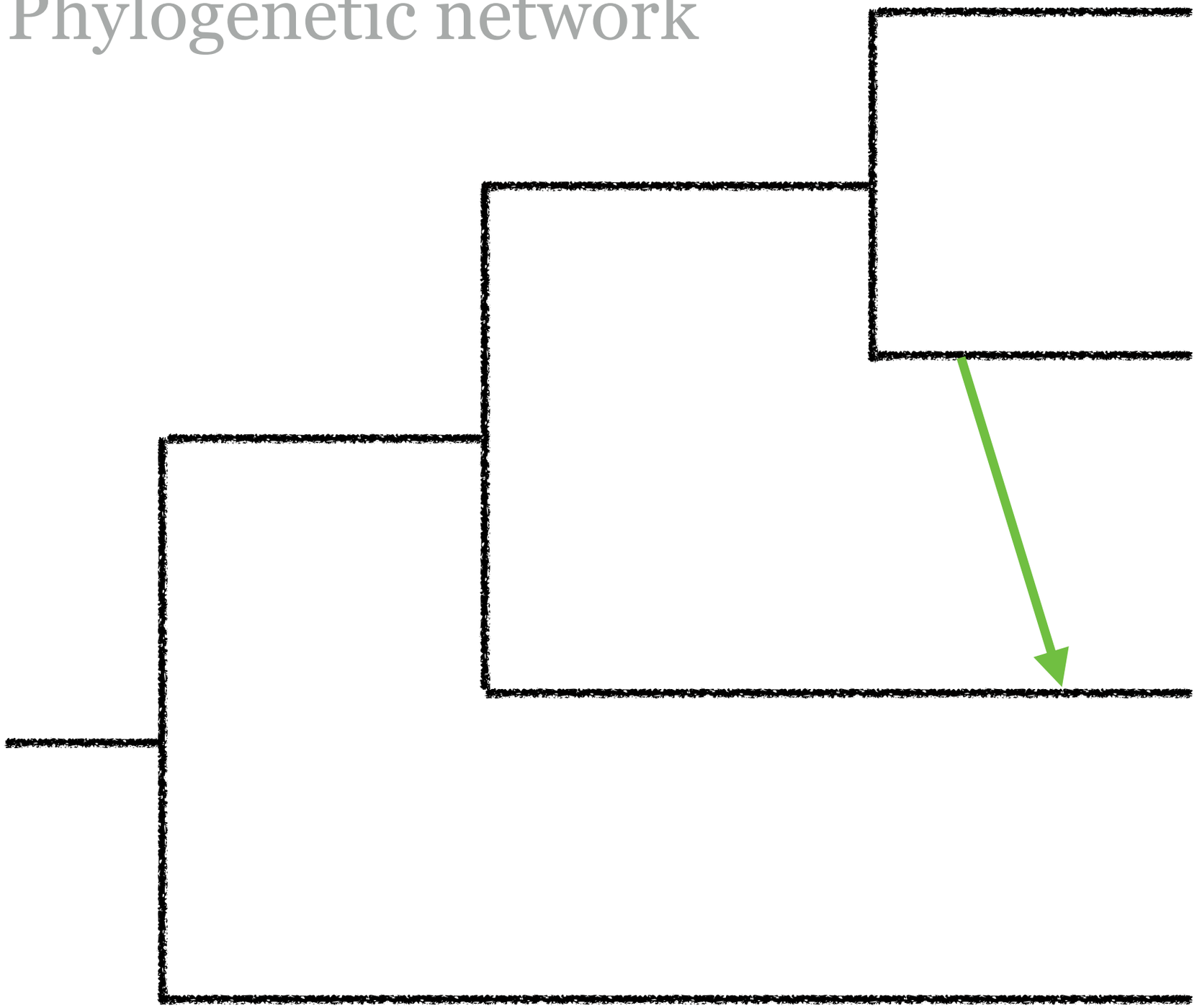
Phylogenetic Networks



https://solislemuslab.github.io/ mstdn.social/@solislemuslab crsl4 @thestatistician

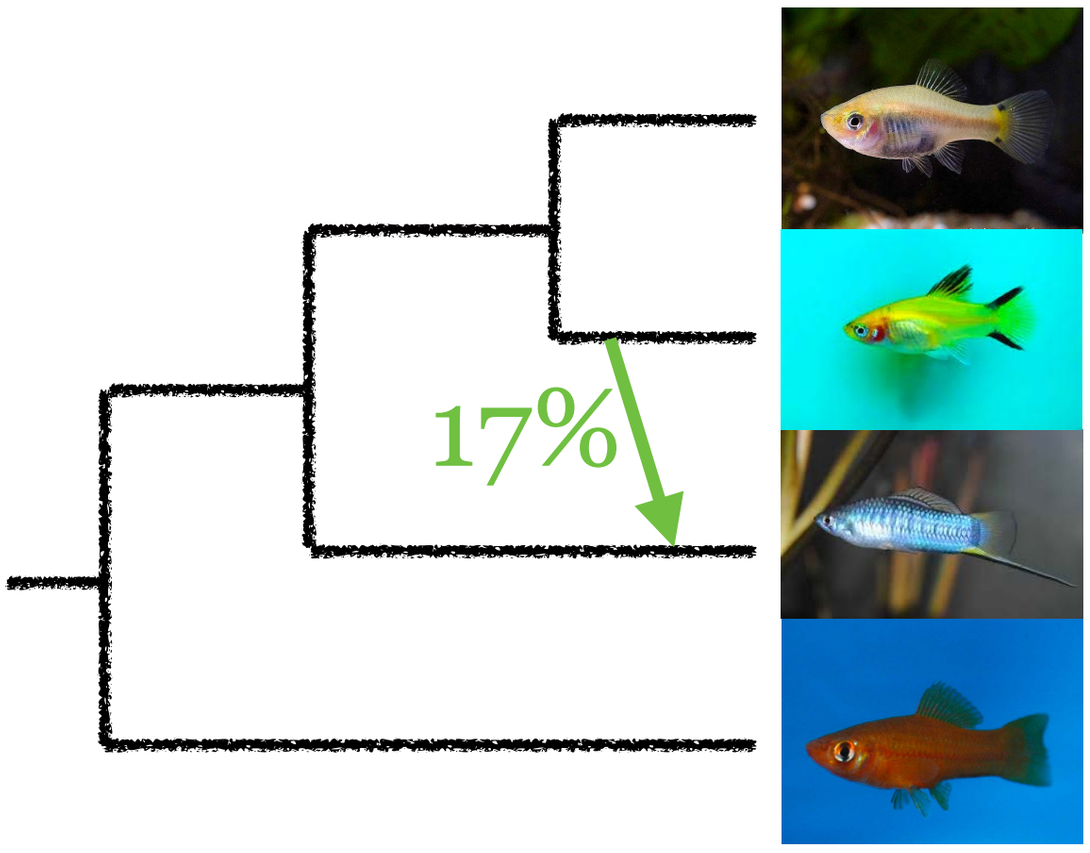
What?

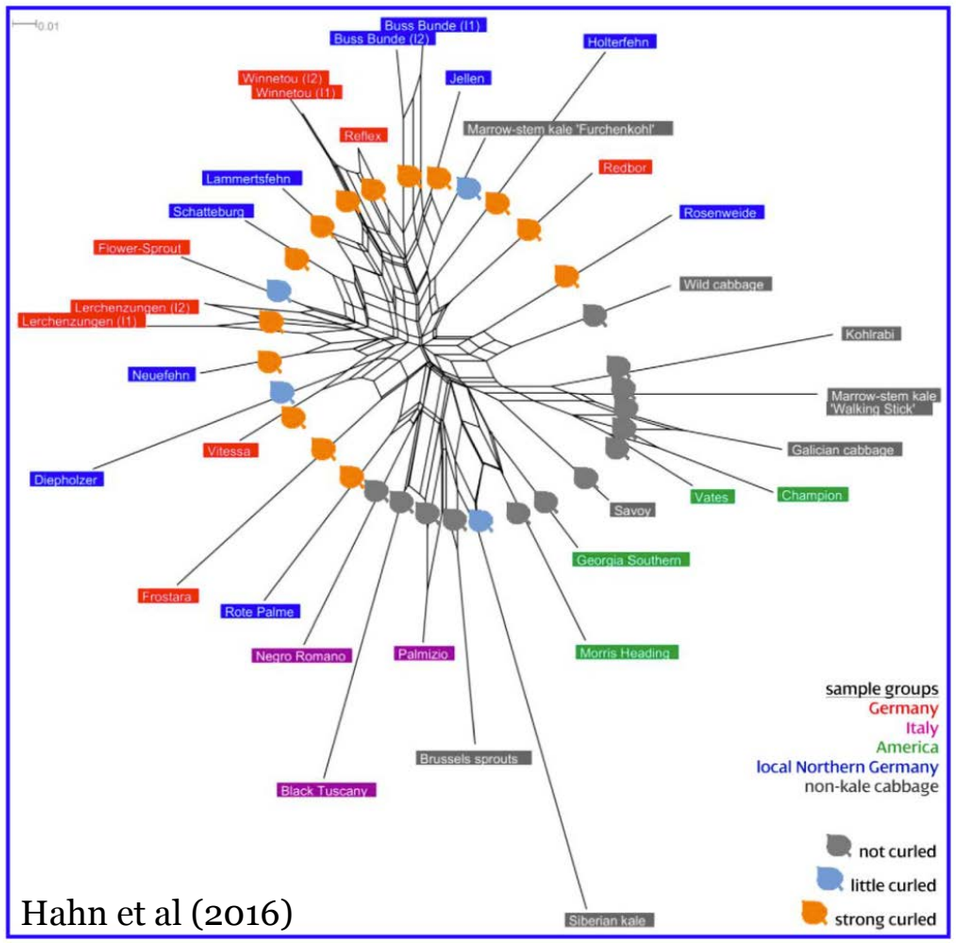
Phylogenetic network



What?

Phylogenetic network





Hahn et al (2016)

Explicit

Implicit

Why?

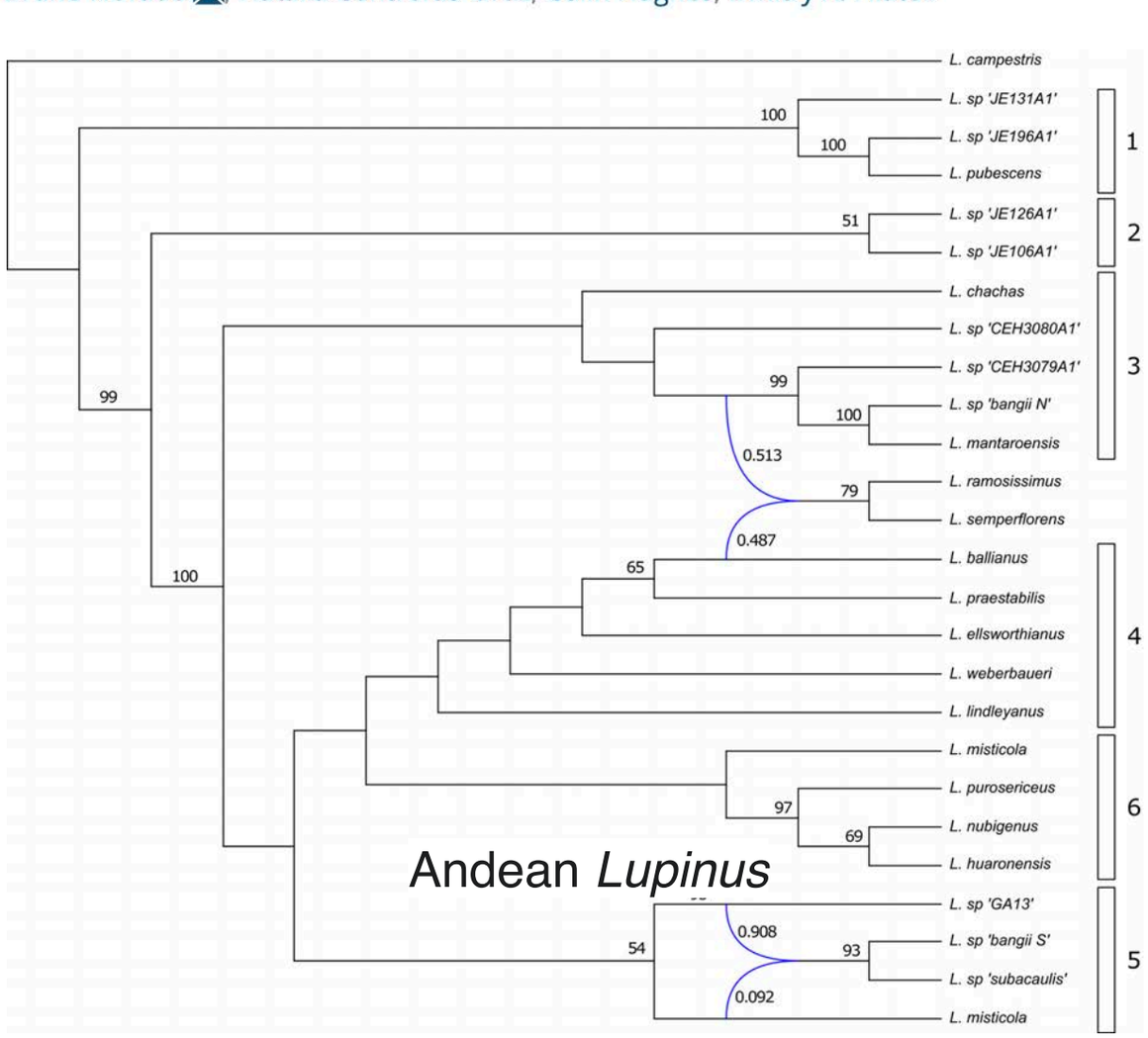
Phylogenetic network

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

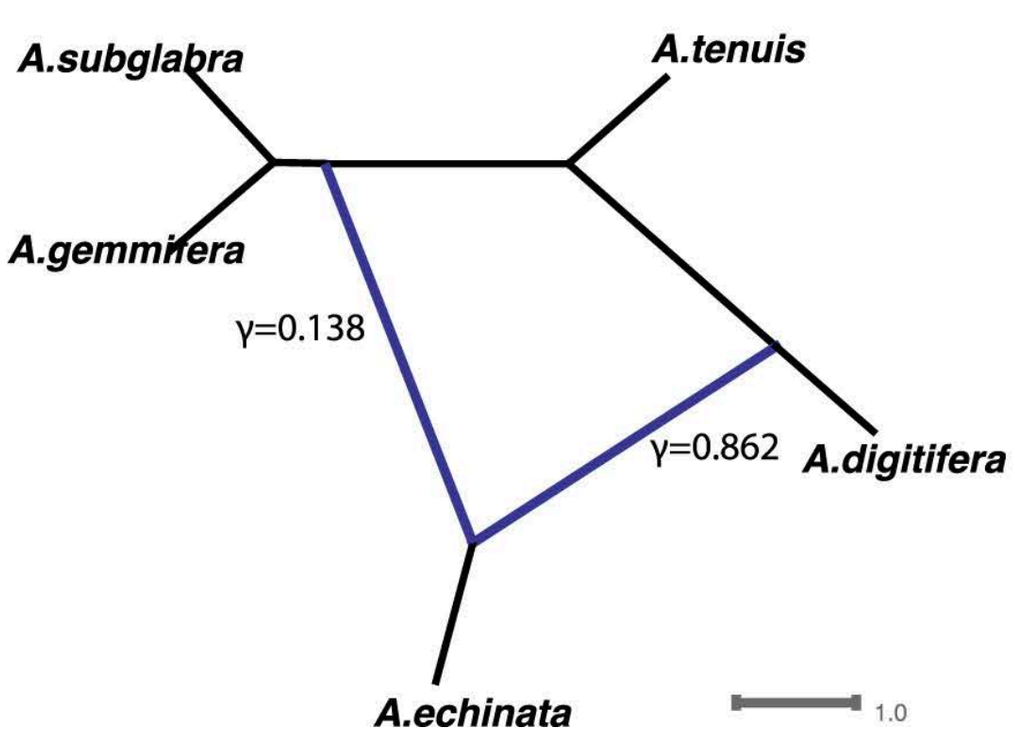
Volume 28, Issue 21, 5 November 2018, Pages 3373-3382.e5

CellPress

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]

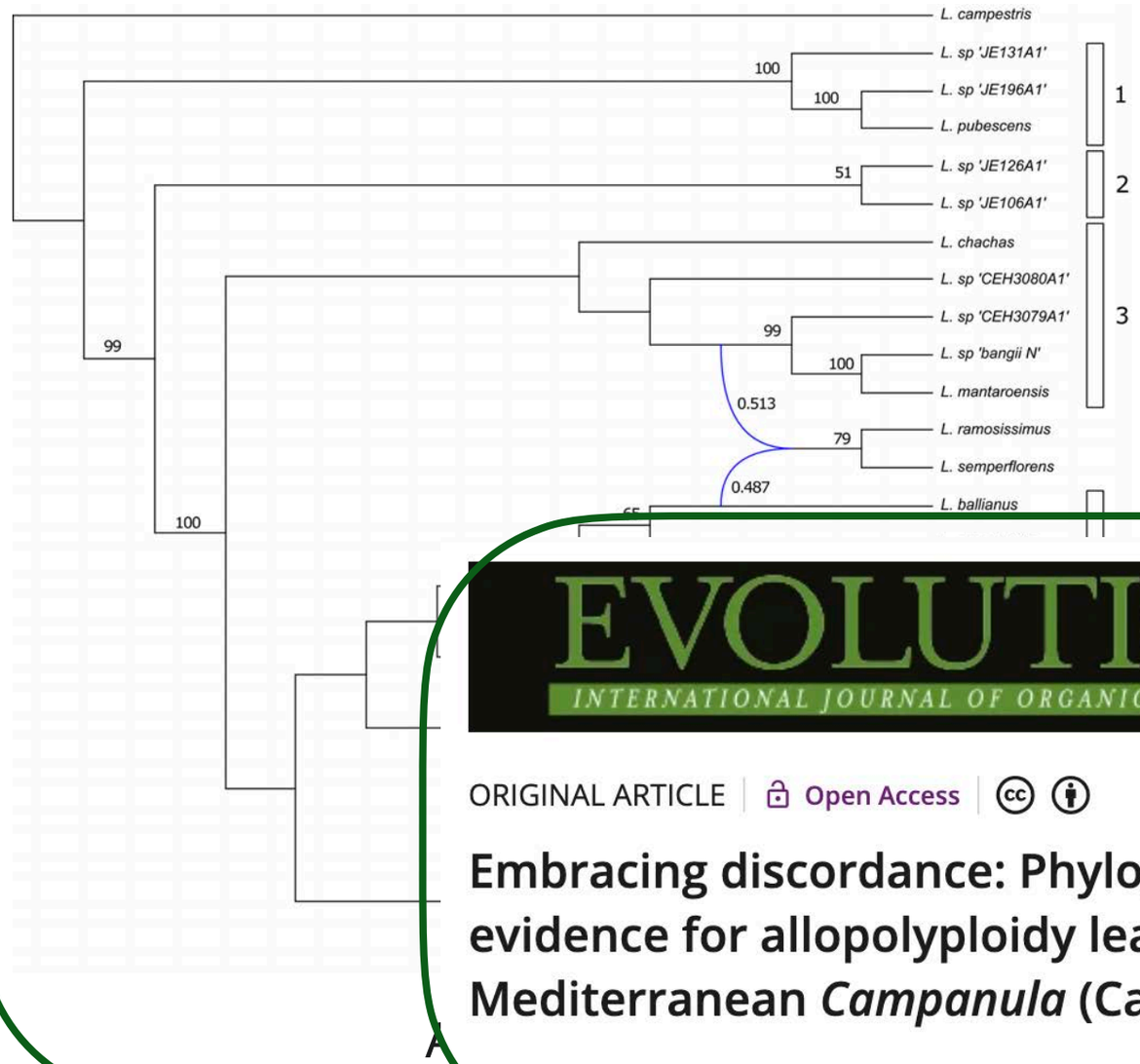


New Phytologist

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



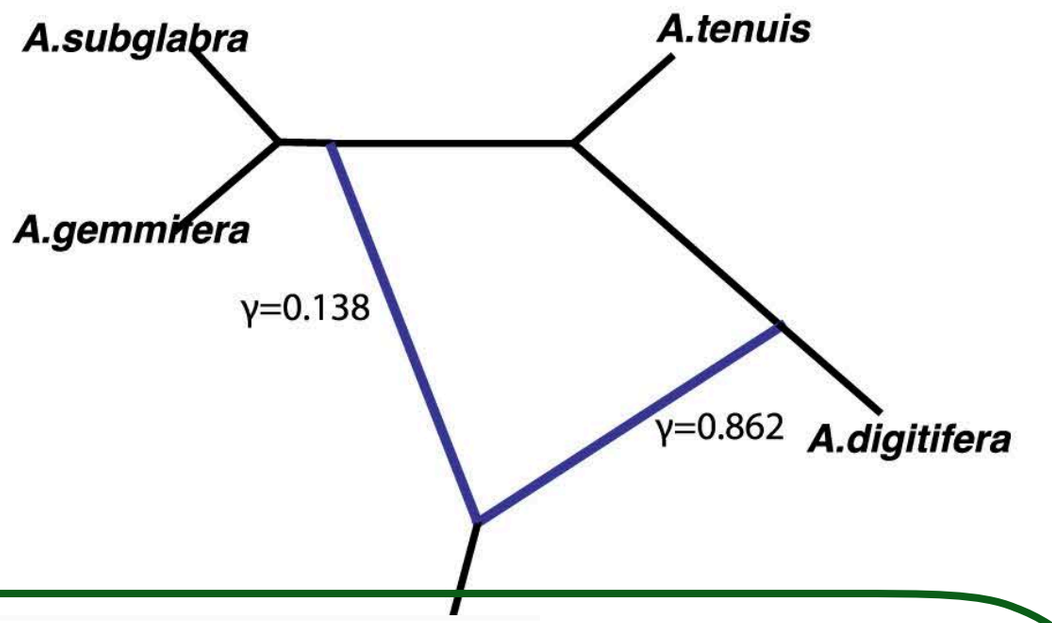
Current Biology

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]

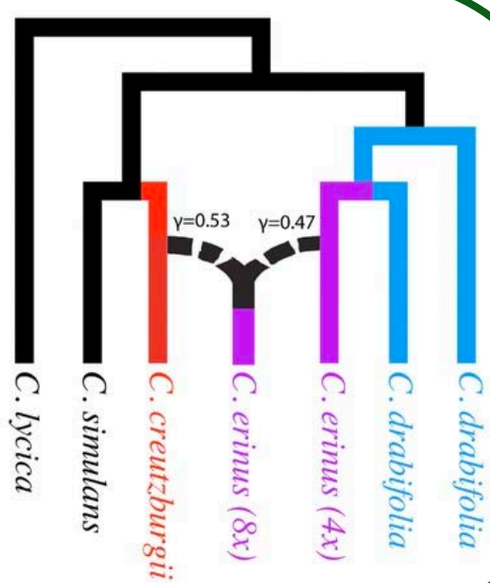


EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION
SSE
SOCIETY for the STUDY of EVOLUTION

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



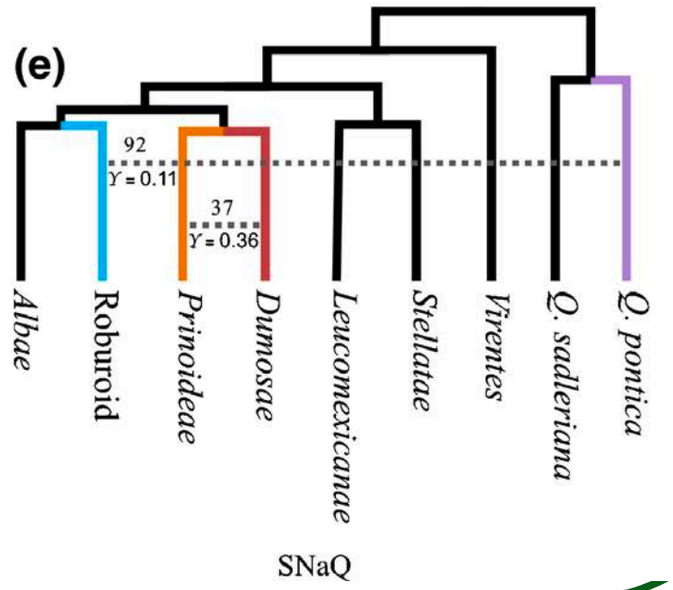
New Phytologist

Full paper | Free Access

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

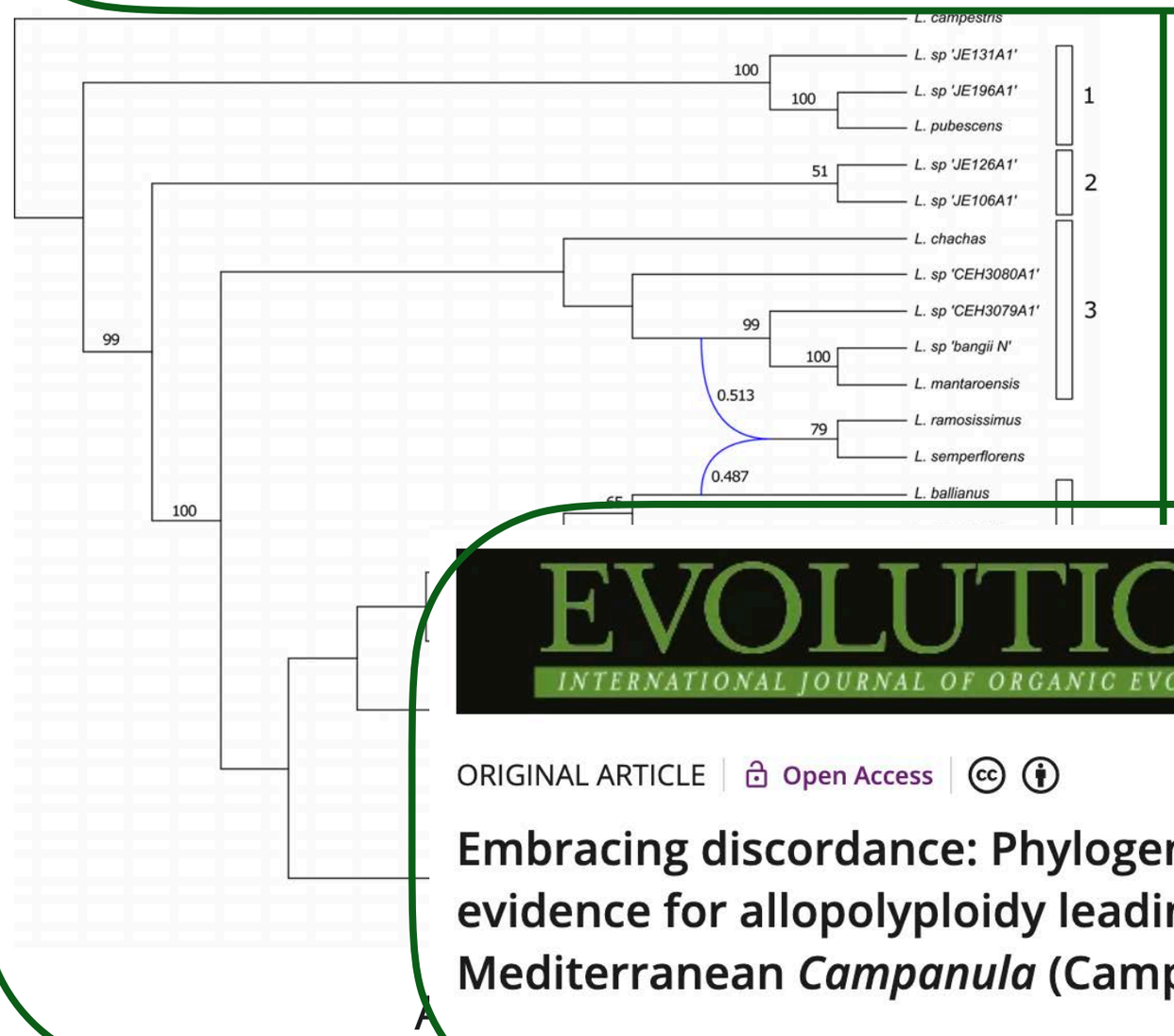
Andrew A. Crowl, Paul S. Manos, John D. McVay, Alan R. Lemmon, Emily Moriarty Lemmon, Andrew L. Hipp ... See fewer authors

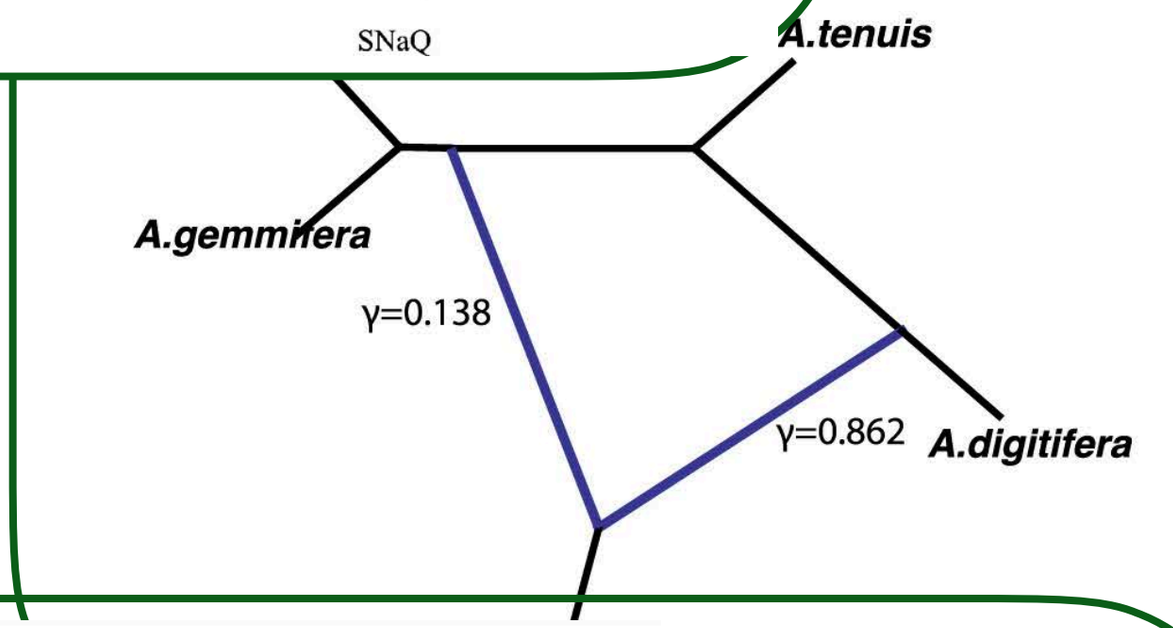
First published: 08 April 2019 | https://doi.org/10.1111/nph.15842 | Citations: 32



CellPress

imate Change in
Corals





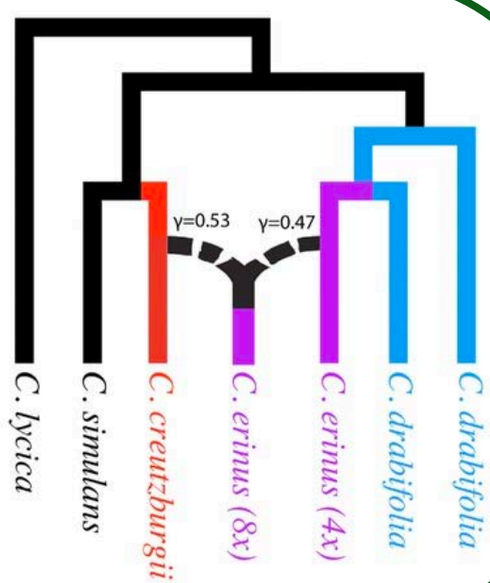
EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

SSE
SOCIETY for the STUDY of EVOLUTION

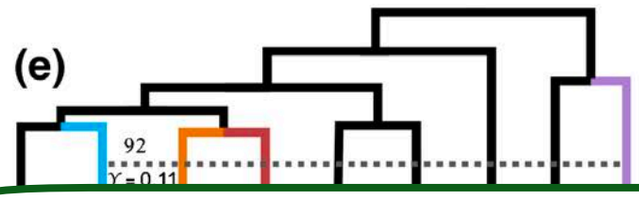
ORIGINAL ARTICLE | Open Access

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



New Phytologist



CellPress

Full p

F

n

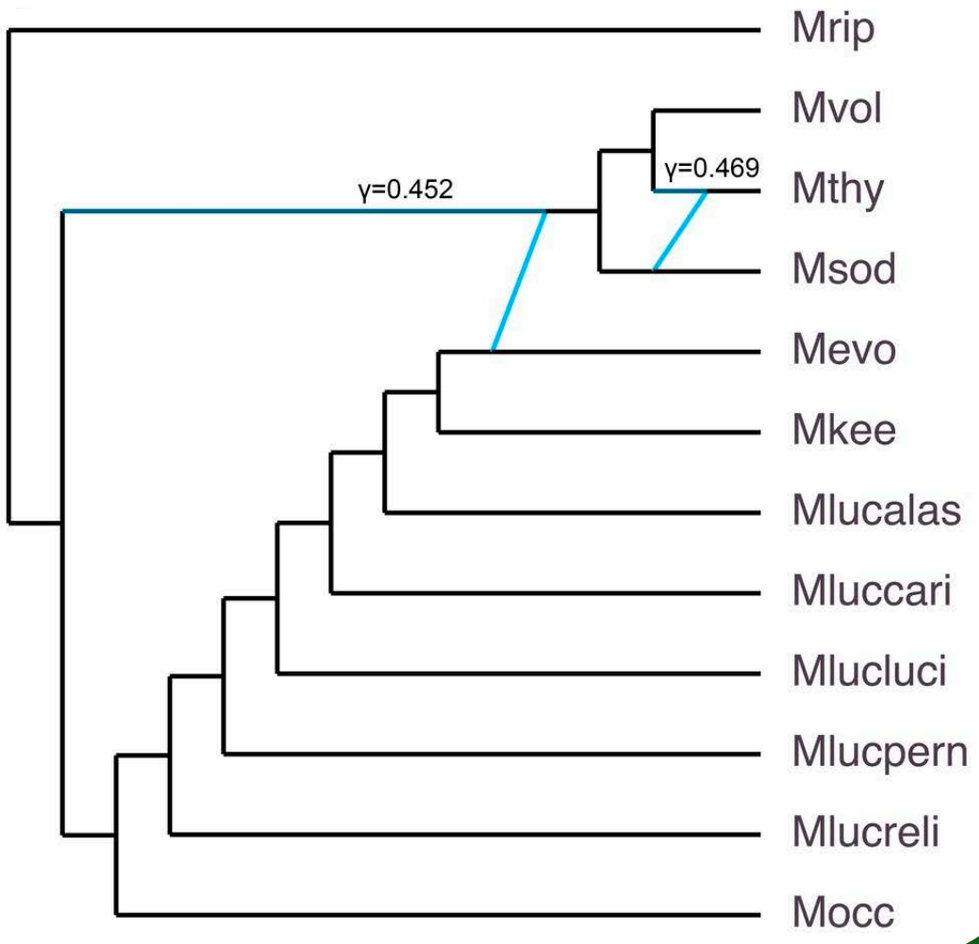
in

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

FREE

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** ▾



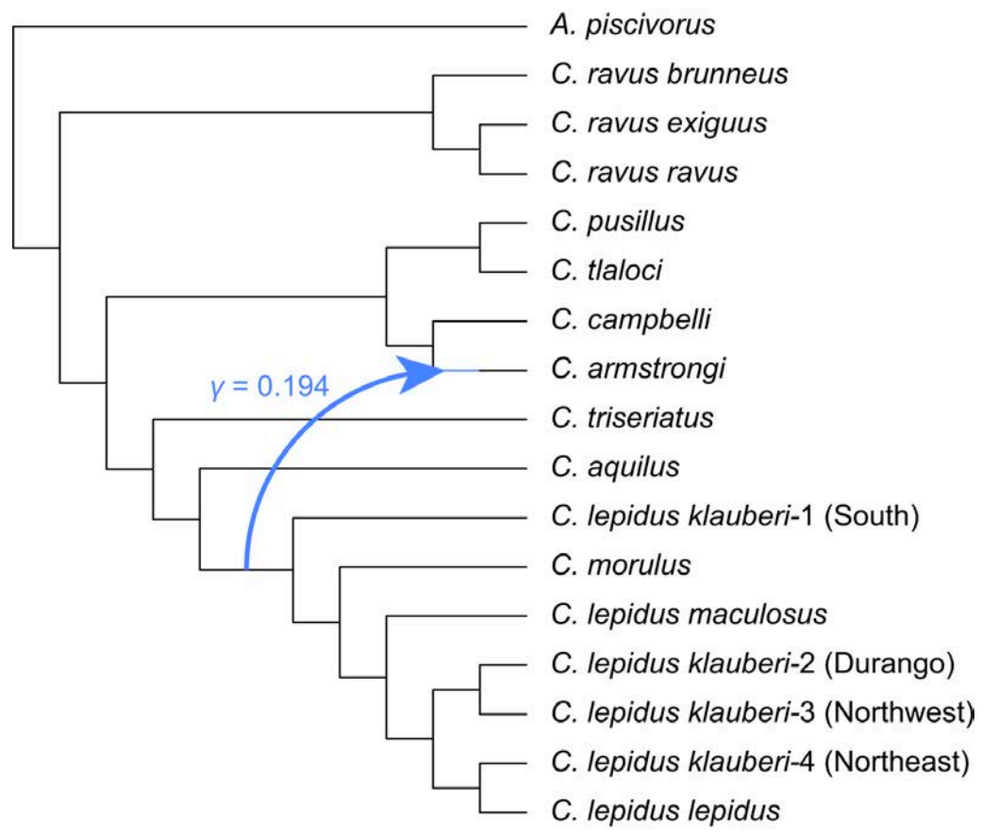
MOLECULAR ECOLOGY RESOURCES

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

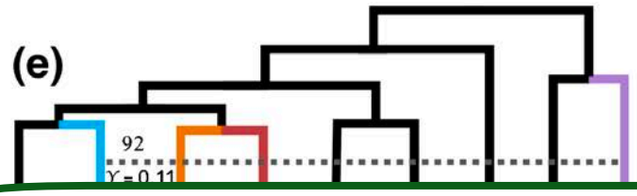


Mediterranean *Campanula* (Campan

Andrew A. Crowl ✉, Cody Myers, Nico Cellinese

New Phytologist

CellPress



MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE | Full Access

Cryptic diversity in the Mexican highlands: Thousands of UCE loci

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

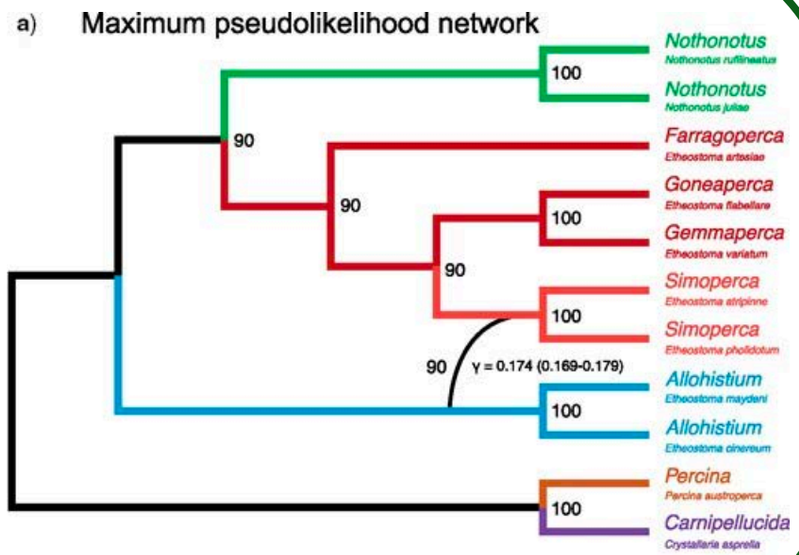
Ariadna E Morales ✉, Bryan C Carstens

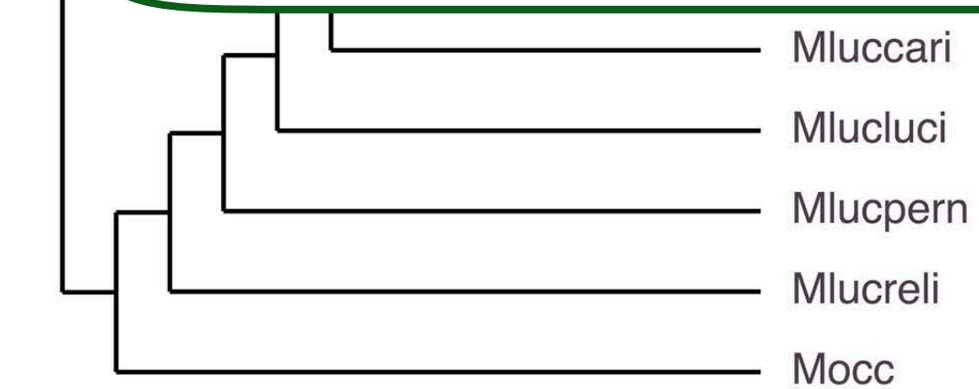
Systematic Biology, Volume 68, Issue 2, March 2019, Pages 329–346,
https://doi.org/10.1093/sysbio/syy074
Published: 03 December 2018 **Article history** ▾

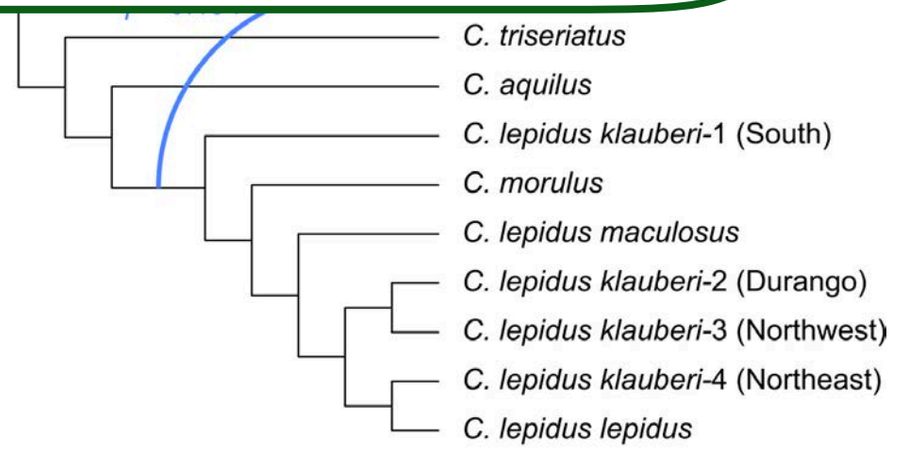
Phylogenomic Signatures of Ancient Introgression in a Rogue Lineage of Darters (Teleostei: Percidae)

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** ▾



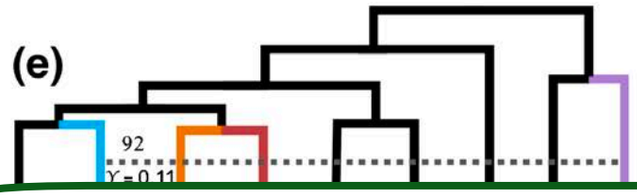




Mediterranean *Campanula* (Campanul

Andrew A. Crowl ✉, Cody Myers, Nico Cellinese

New Phytologist



CellPress

Full p

F

n

in

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

FREE

Ariadna E Morales ✉, Bryan C Carstens

Systemati

https

Pu

MOLECULAR ECOLOGY
RESOURCES

RESOURCE ARTICLE | Full Access

Cryptic diversity in the Mexican highlands: Thousands of UCE loci

its and

otalus)

E.

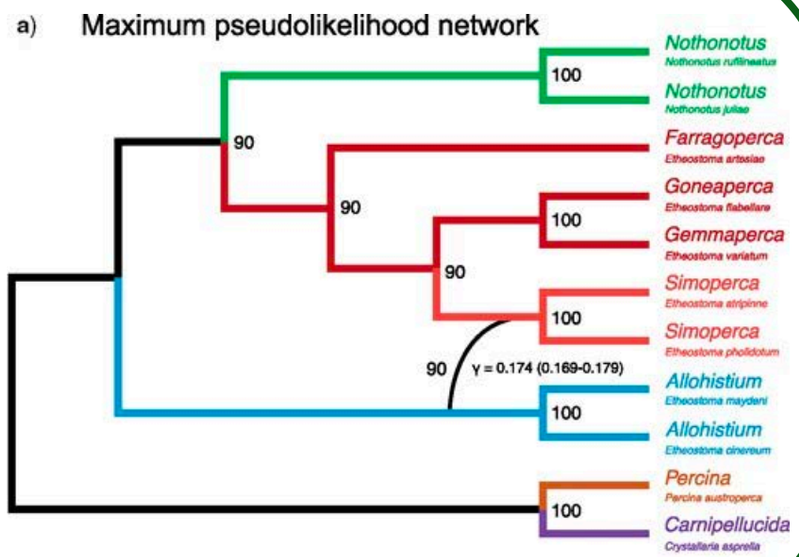
0

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

FREE

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** ▾



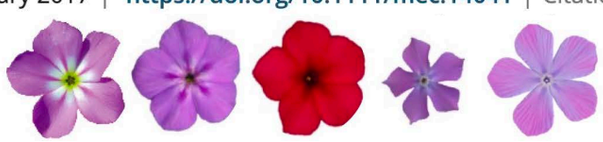
MOLECULAR ECOLOGY

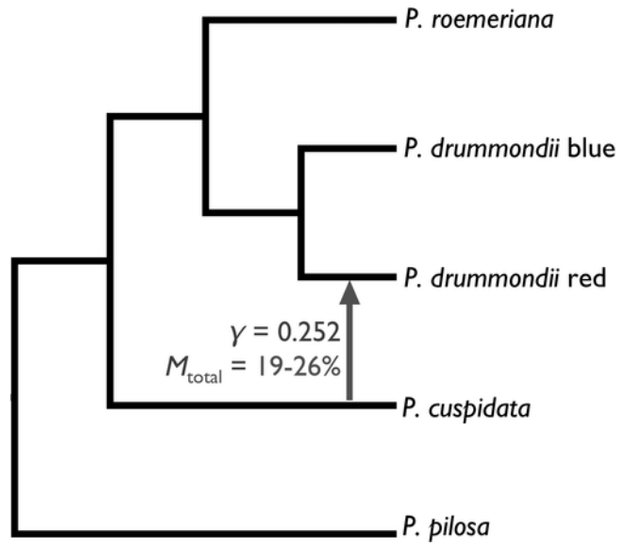
Original Article | Full Access

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





h)

ngo)

west)

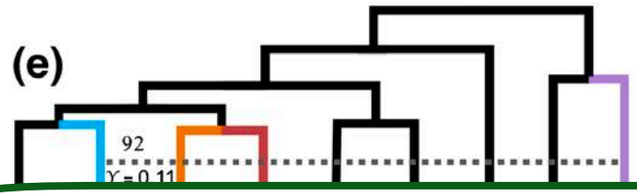
east)

drabifolia

bifolia

(4x)

New Phytologist



CellPress

MOLECULAR ECOLOGY RESOURCES

Evidence that *Myotis lucifugus* "Subs… Nonsister Species, Despite Gene F…

Ariadna E Morales ✉, Bryan C Carstens

Systemati…
https…
Pu…

Phylogenomic Signat… in a Rogue Lineage of…

FREE

Daniel J MacGuigan ✉, Thomas J N…

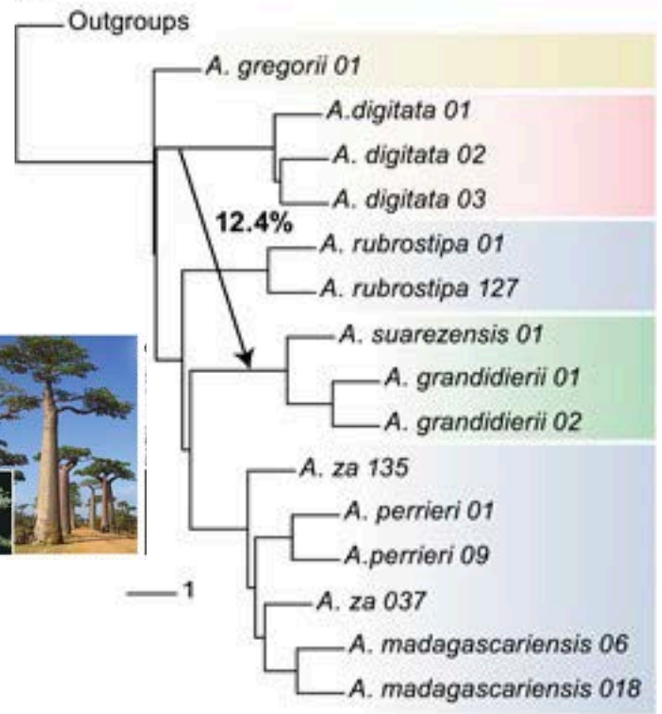
Systematic Biology, Volume 68, Issue…
https://doi.org/10.1093/sysbio/syy0…
Published: 03 December 2018 **A…**

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

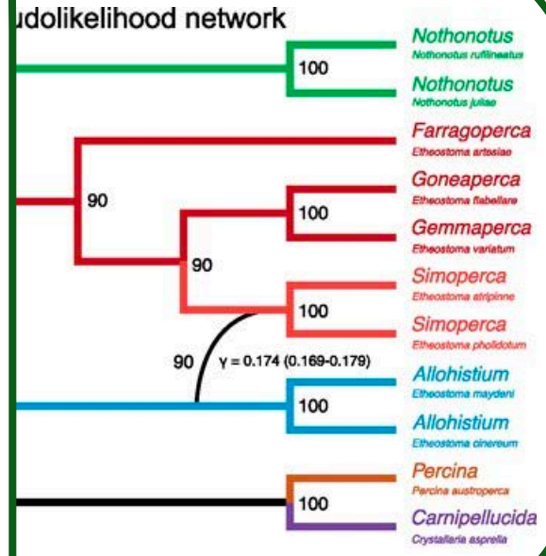
FREE

Nisa Karimi ✉, Corrinne 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 ▾**



…dolikelihood network



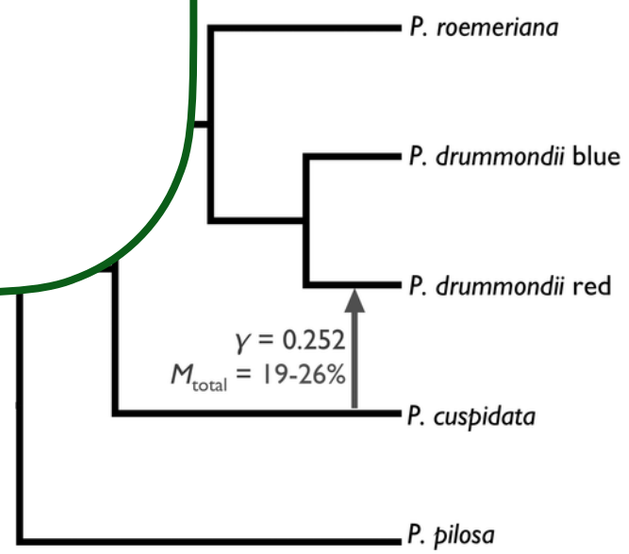
MOLECU…

Original Article | Full Access

Genomic evidence of ge… *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



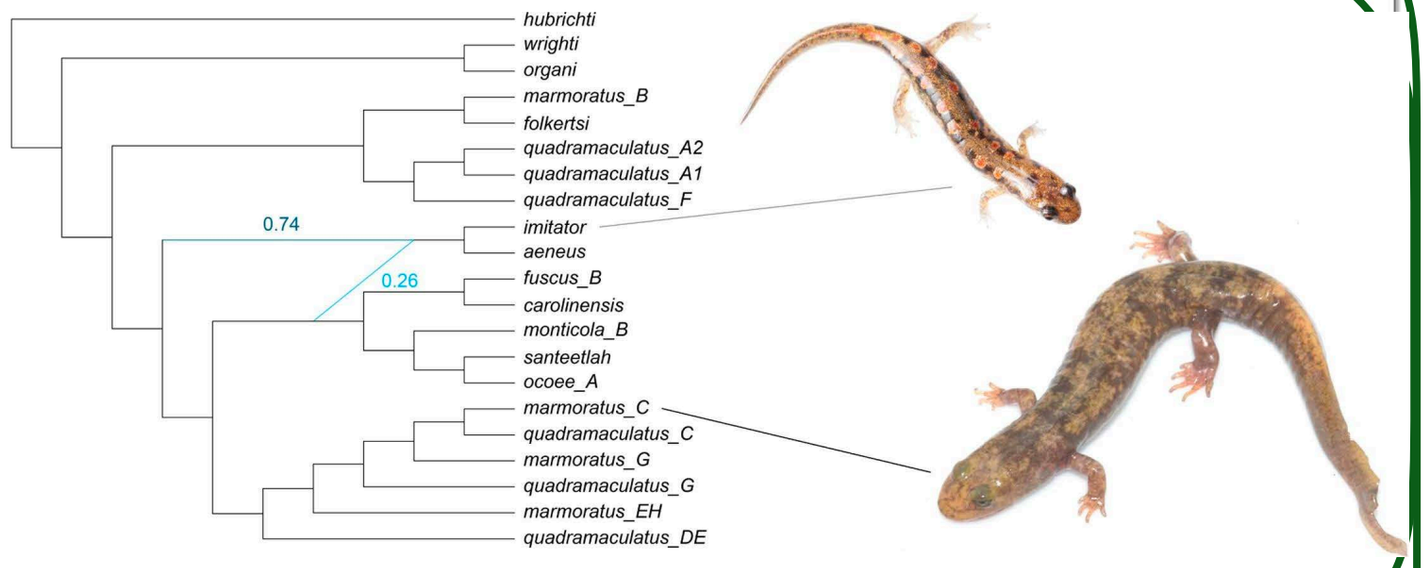
ELSEVIER

Molecular Phylogenetics and Evolution

Volume 146, May 2020, 106751

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]

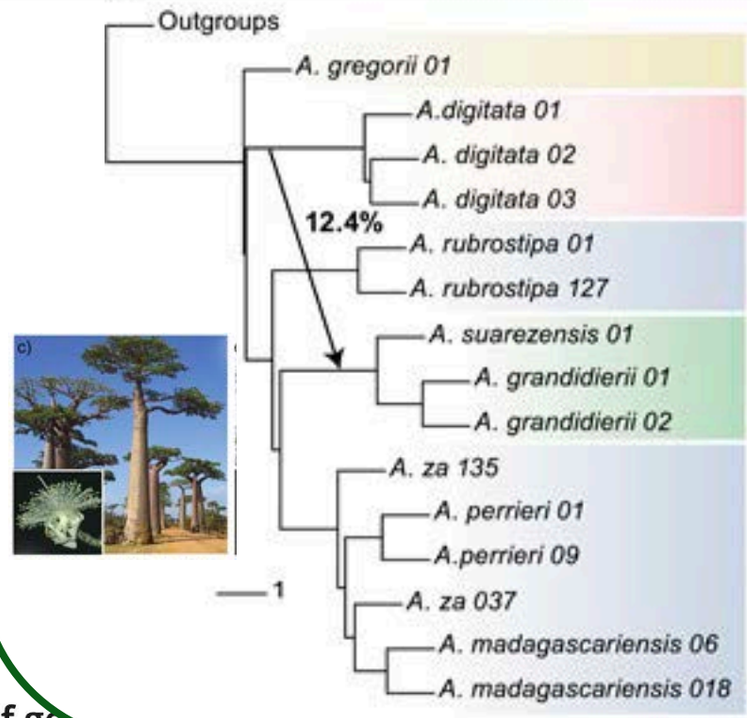


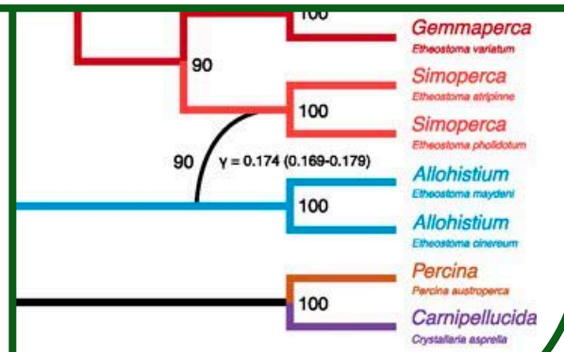
Daniel J MacGuigan, Thomas J N

Systematic Biology, Volume 68, Issue

https://doi.org/10.1093/sysbio/syy0

Published: 03 December 2018 **A**





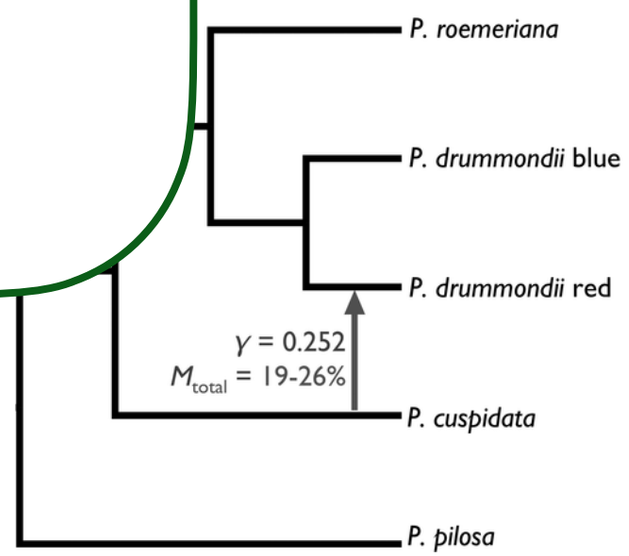
MOLECU

Original Article | Full Access

Genomic evidence of ge… *Phlox*

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

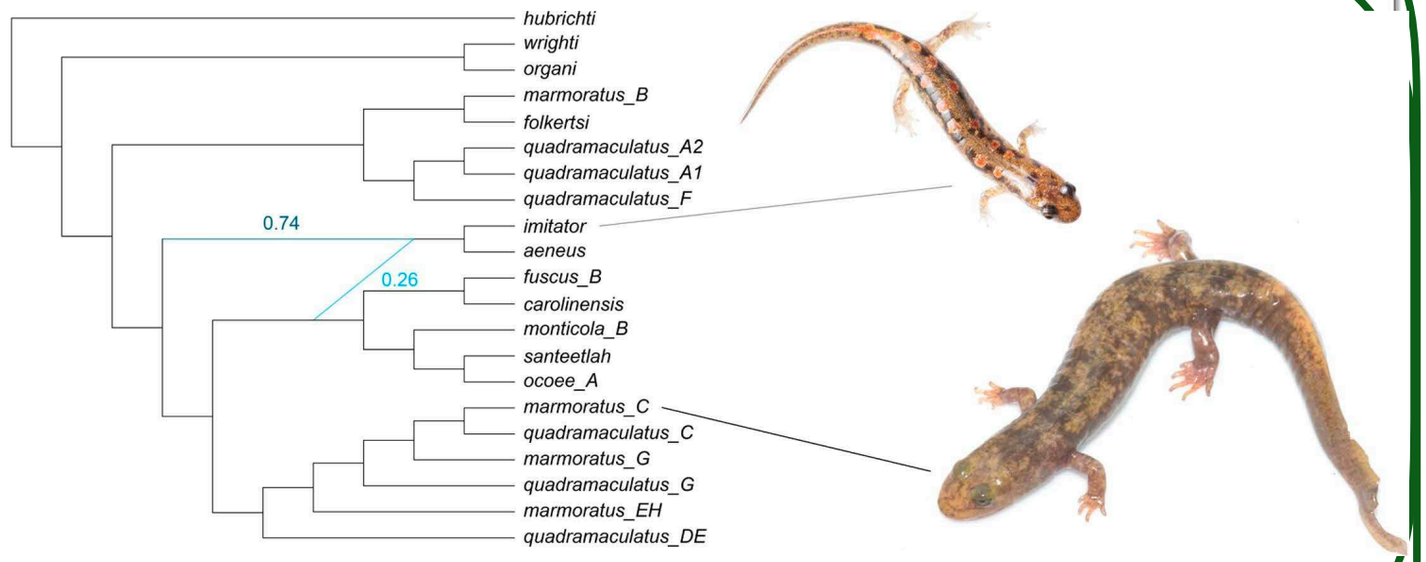
First published: 31 January 2017 | https://doi.org/10.1111/mec.14041 | Citations: 15



Molecular Phylogenetics and Evolution
Volume 146, May 2020, 106751

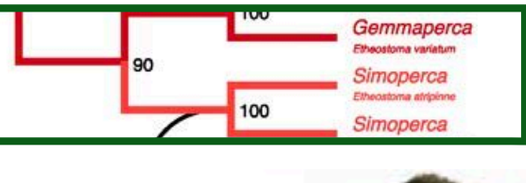
Phylogenomic data reveal reticulation and incongruence among mitochondrial candidate species in Dusky Salamanders (*Desmognathus*)

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
ht

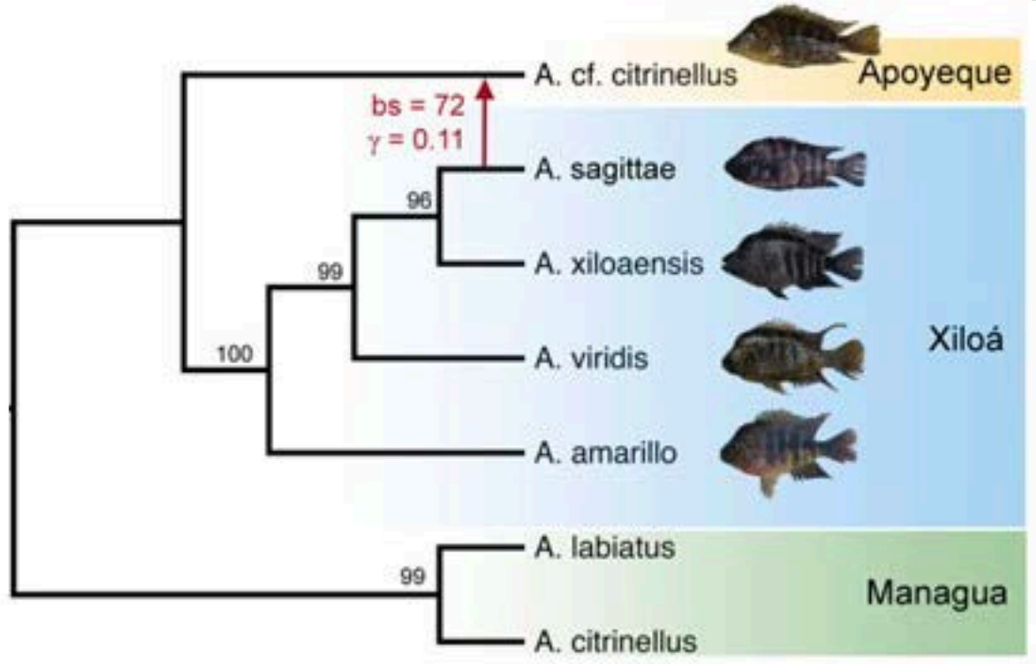


Implementing Large Genomic Single Nucleotide Polymorphism Data Sets in Phylogenetic Network Reconstructions: A Case Study of Particularly Rapid Radiations of Cichlid Fish

FREE

Melisa Olave, Axel Meyer

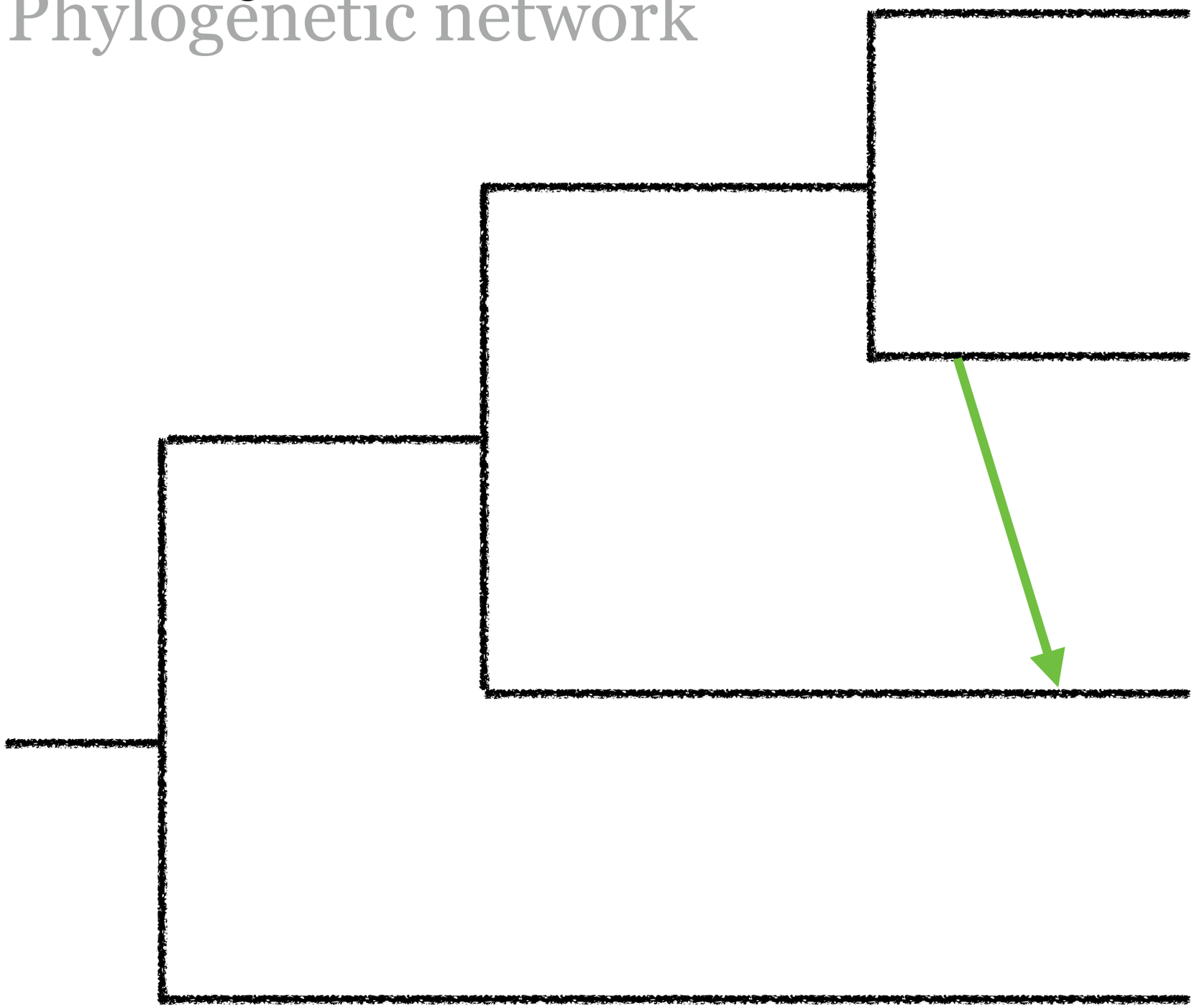
Systematic Biology, Volume 69, Issue 5, September 2020, Pages 848–862,
https://doi.org/10.1093/sysbio/syaa005
Published: 03 February 2020 **Article history**



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

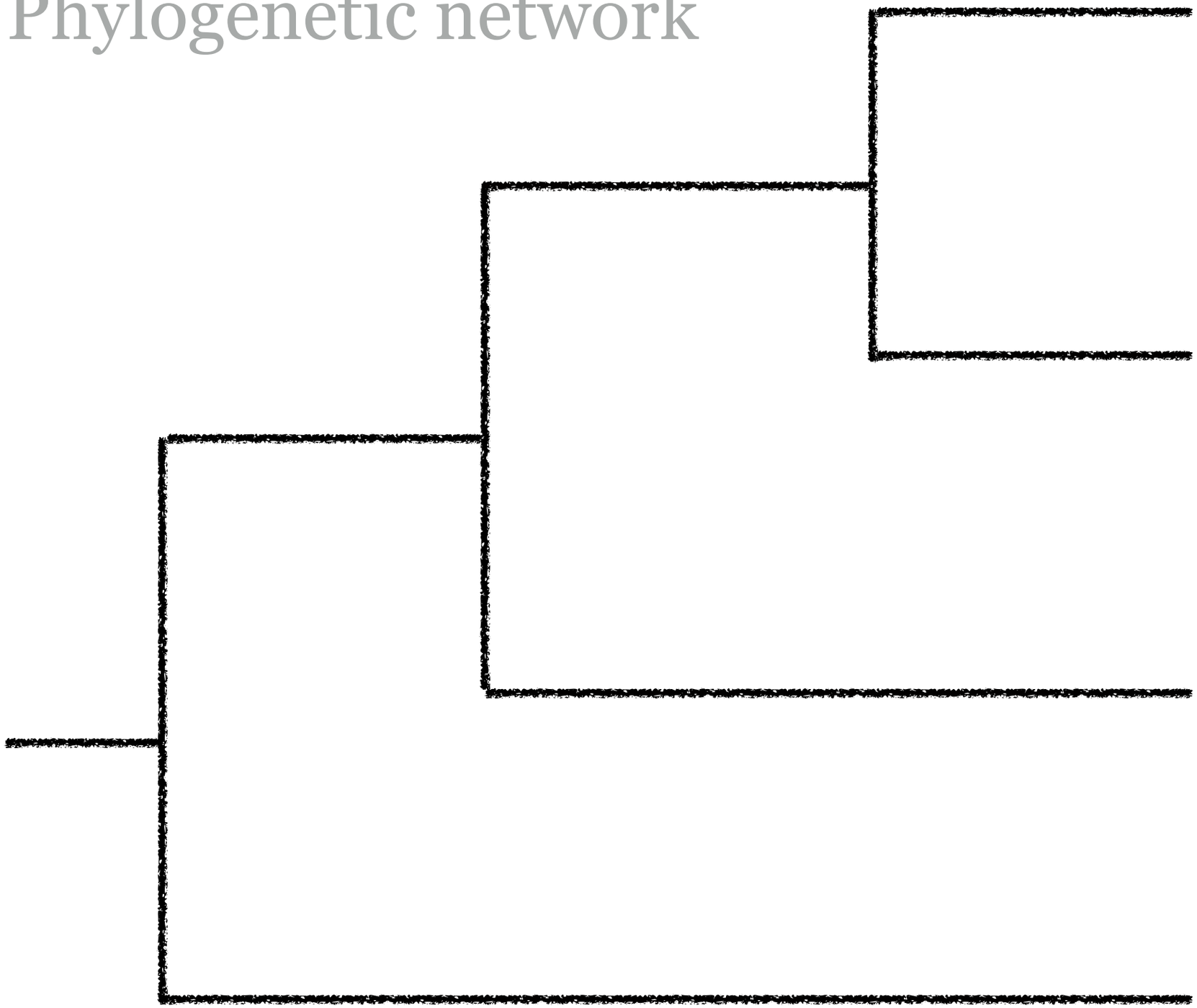
Why?

Phylogenetic network



Why?

Phylogenetic network

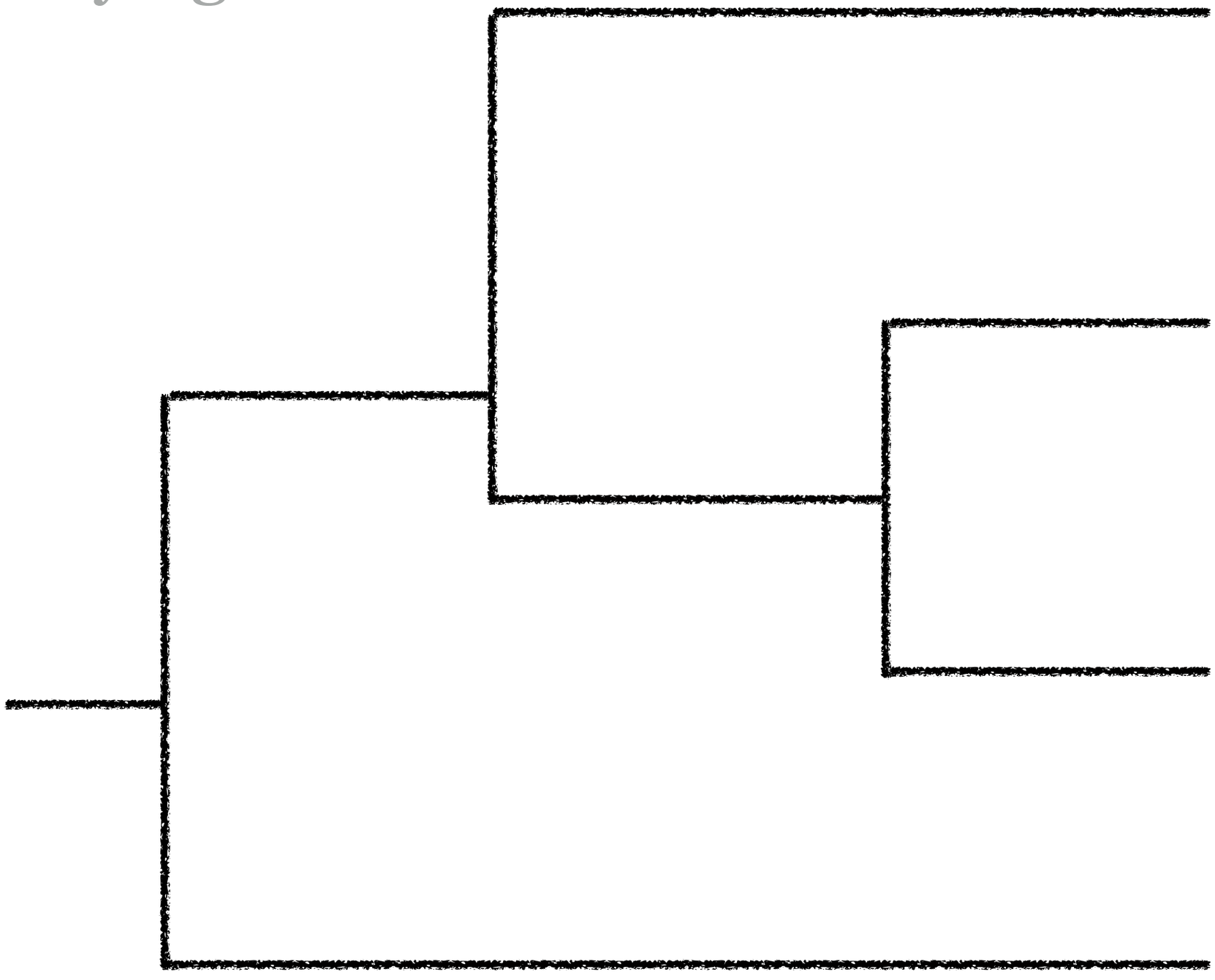


Main tree

Why?

Phylogenetic network

Ignore gene flow
=>Wrong tree!

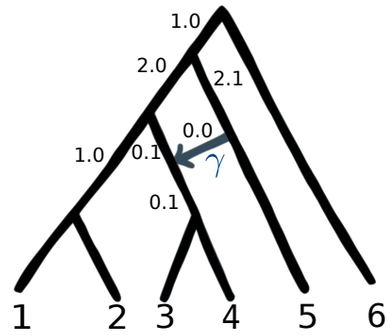


(S.-L., Yang, Ané, 2016, Syst Bio)

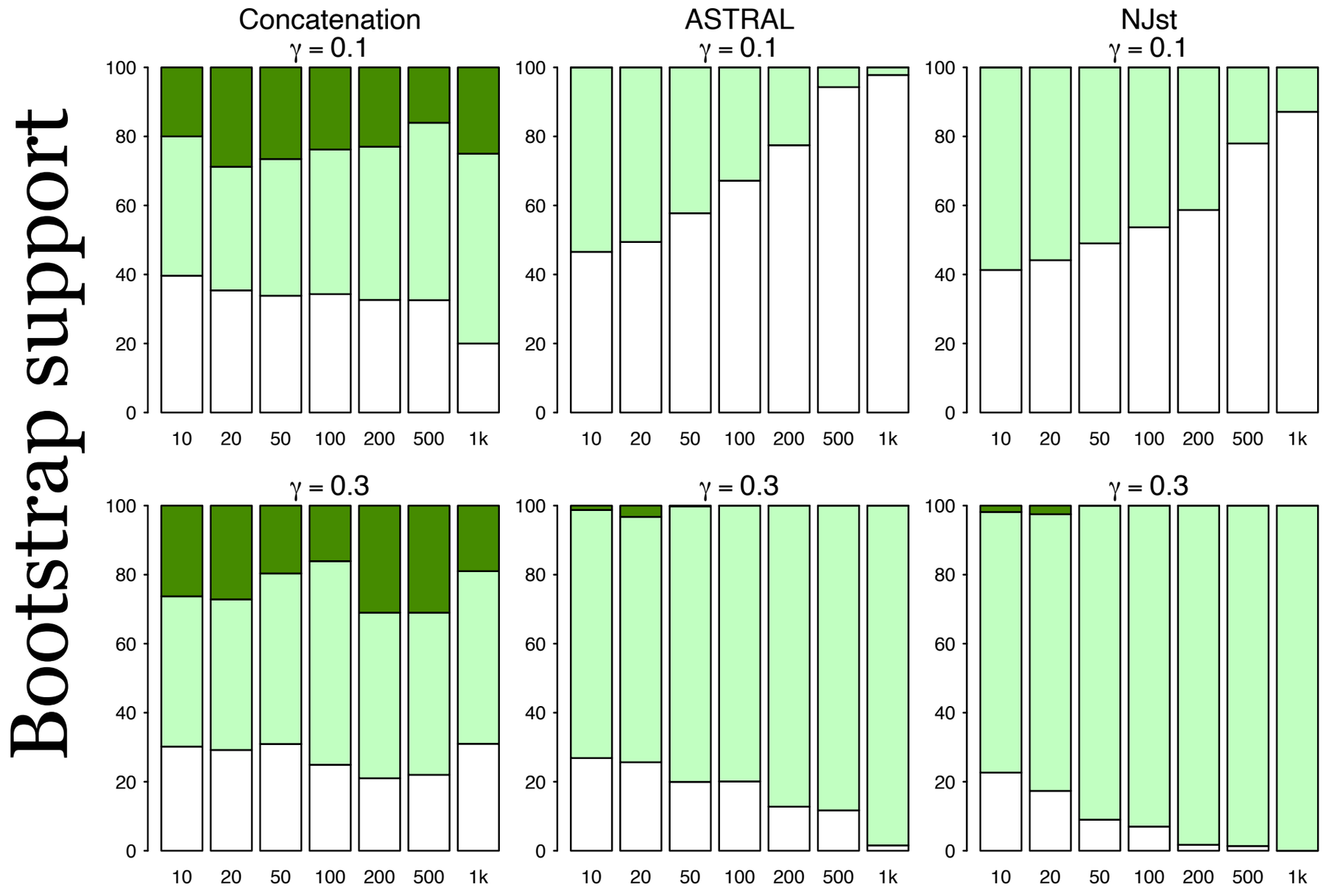
Why?

Phylogenetic network

Coalescent tree methods not robust to gene flow



White: true tree



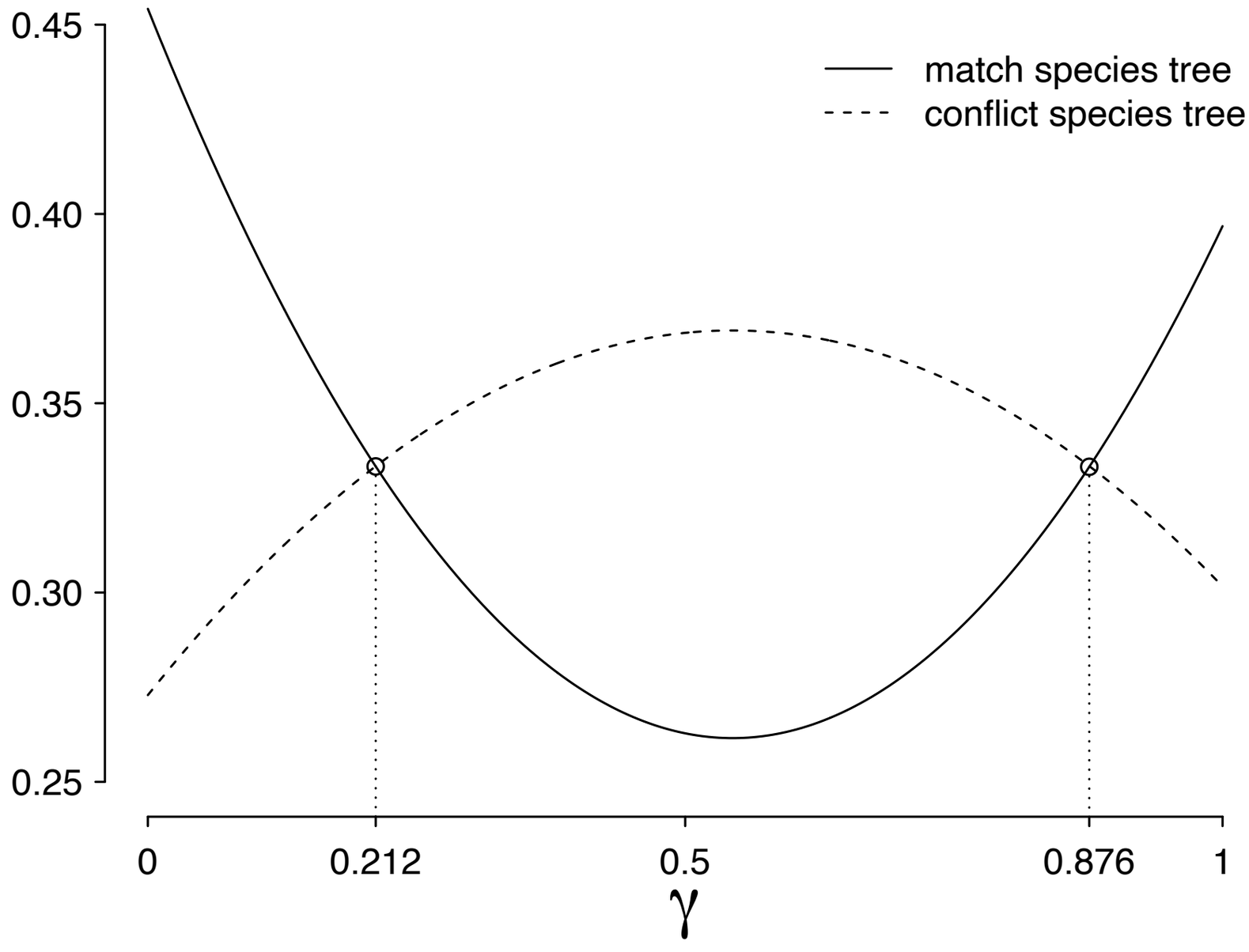
(S.-L., Yang, Ané, 2016, Syst Bio)

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

Why?

Phylogenetic network

Anomaly zone with gene flow

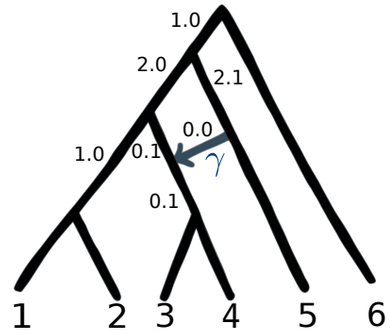


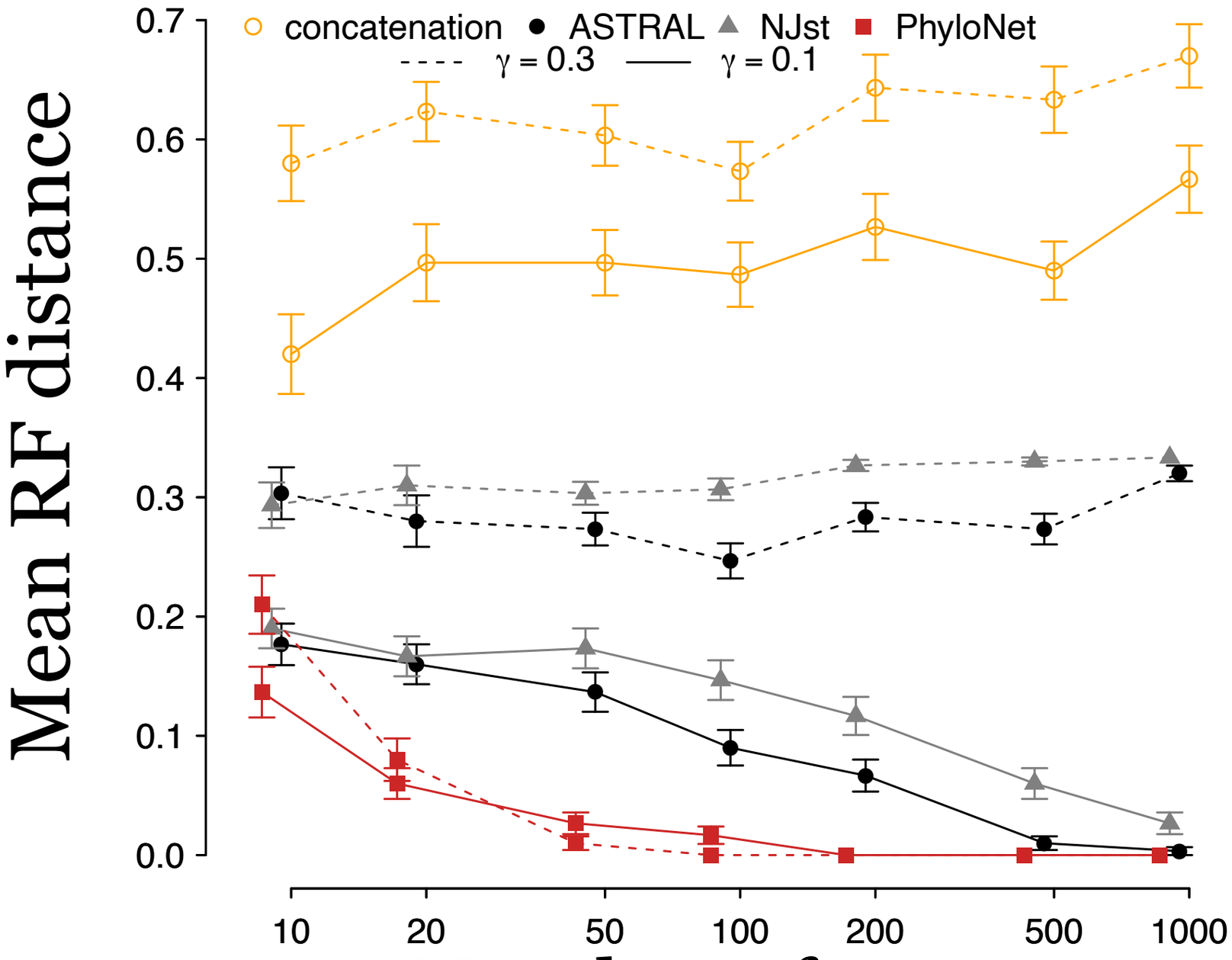
(S.-L., Yang, Ané, 2016, Syst Bio)

Why?

Phylogenetic network

Coalescent tree methods not robust to gene flow





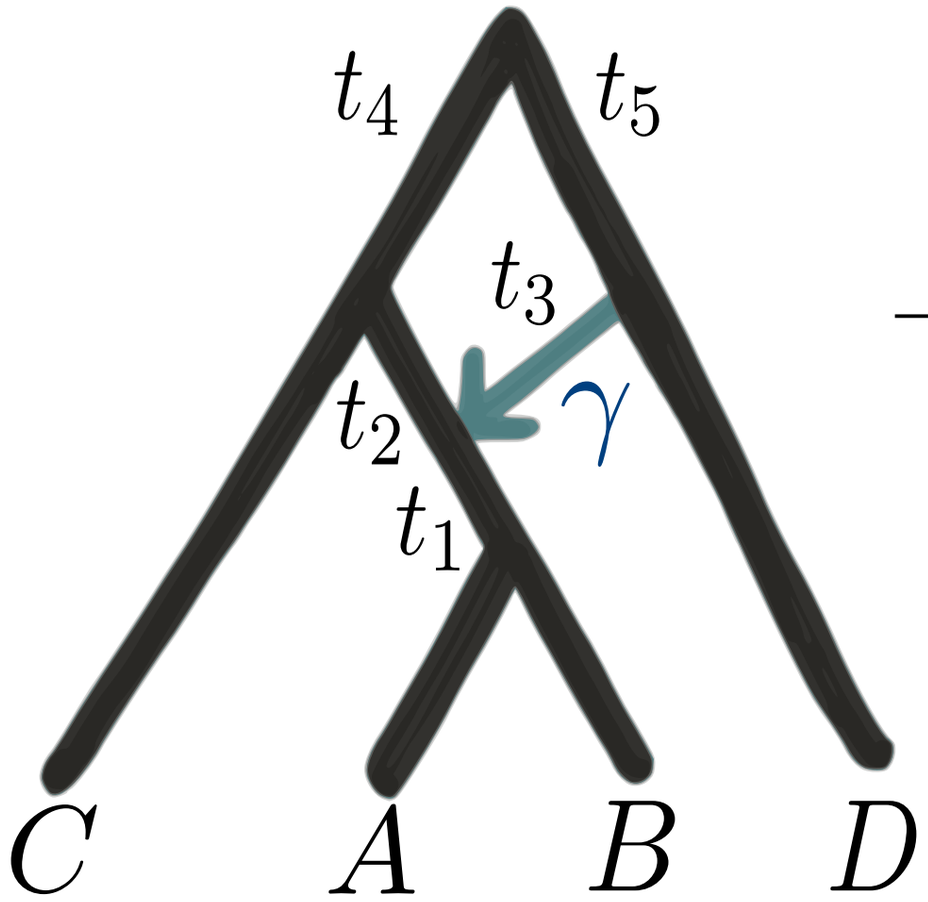
(S.-L., Yang, Ané, 2016, Syst Bio)

ASTRAL (Mirarab et al, 2014)
NJst (Liu&Yu,2011)
PhyloNet (Yu et al 2012, 2014)

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)

See also Long & Kubatko (2018) for AUGT under continuous gene flow between sister species

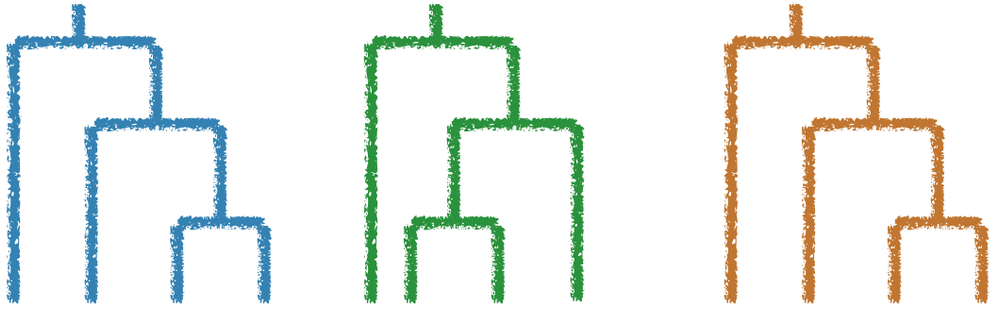
So far…

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

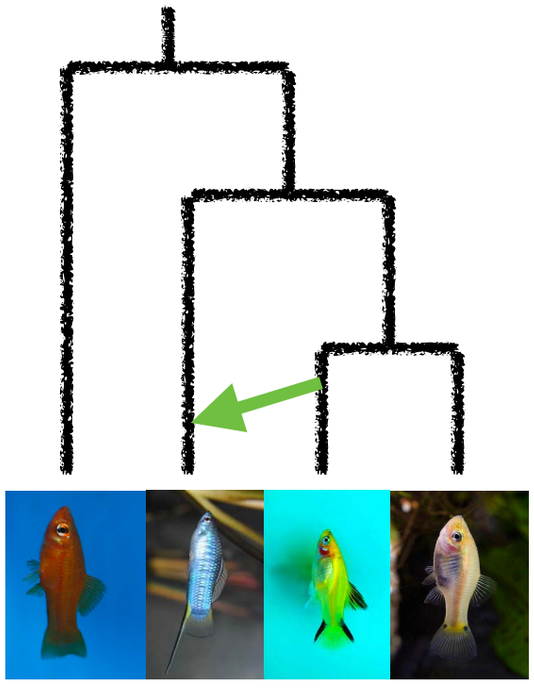
How?

Phylogenetic network

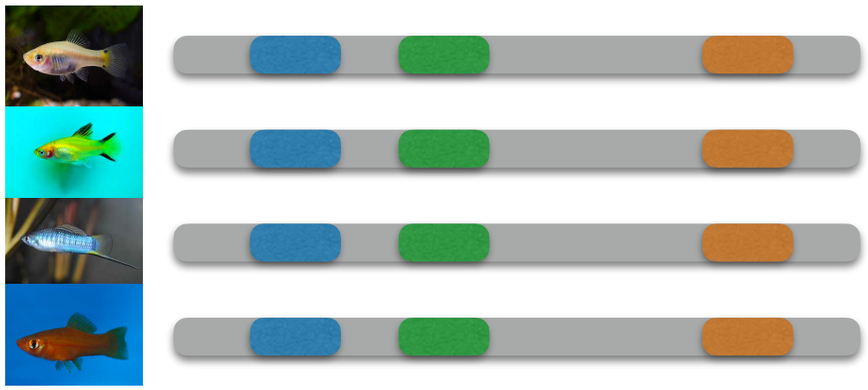
MrBayes
(Huelsenbeck, Ronquist, 2001)
RAxML
(Stamatakis, 2014)
PhyML
(Guindon et al, 2010)
RevBayes
(Hoehna et al, 2016)
IQ-TREE
Nguyen et al. (2015)



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



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



$$P(N, G, \theta | D) \propto \pi(N)\pi(\theta) \prod_{i=1}^{L} P(D_i|G_i)P(G_i|N)$$

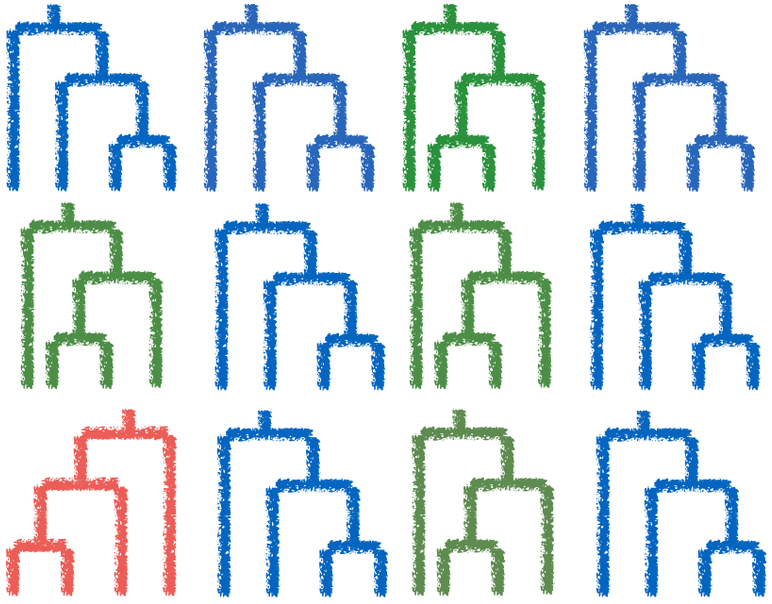
network · gene trees · L loci · **Prior Network** · **Substitution Model** · **Multispecies Coalescent**

Birth-hybridization process

reticulations, cycle diameter

BEAST2
(Zhang et al, 2017)

PhyloNet Bayesian
(Wen et al, 2016)

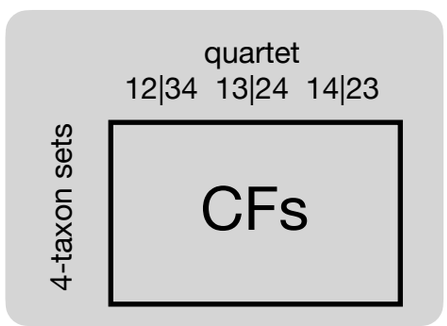


$$P(N, \theta | G) \propto \pi(N)\pi(\theta) \prod_{i=1}^{L} P(G_i|N, \theta)$$

PhyloNet Bayesian
(Wen et al, 2016)

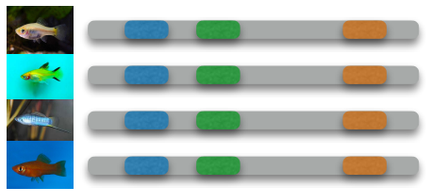
$$L(N, \theta) = \prod_{i=1}^{L} P(G_i|N, \theta)$$

PhyloNet Likelihood
(Yu et al, 2014)



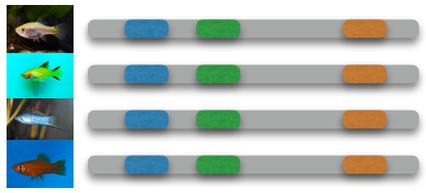
$$\tilde{L}(N, \theta) \propto \prod_{q} L(q|N, \theta)$$

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



BEAST2
(Zhang et al, 2017)

Birth-hybridization process

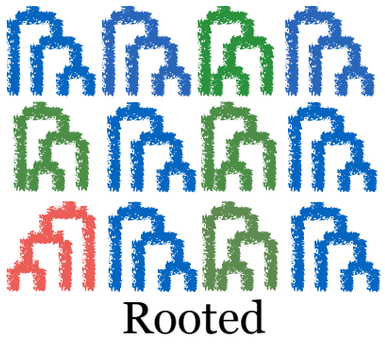


PhyloNet Bayesian
(Wen et al, 2016)

reticulations, cycle diameter

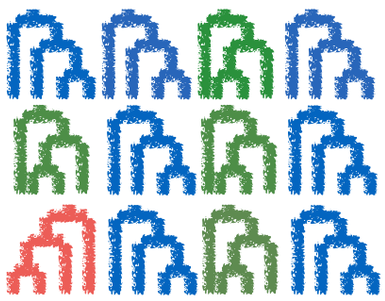
MCMC: Network moves, mixing

Most accurate, not scalable



Rooted

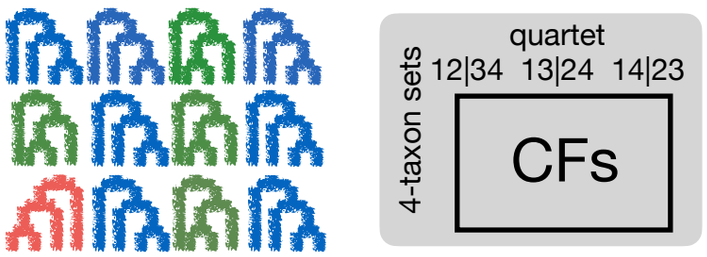
PhyloNet Bayesian
(Wen et al, 2016)



Rooted

PhyloNet Likelihood
(Yu et al, 2014)

Heuristic search: Network moves



Unrooted

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.

Slide by Cecile Ane

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

Marianne Bjørner

arXiv > q-bio > arXiv:2211.00712

Search... Help | Advance

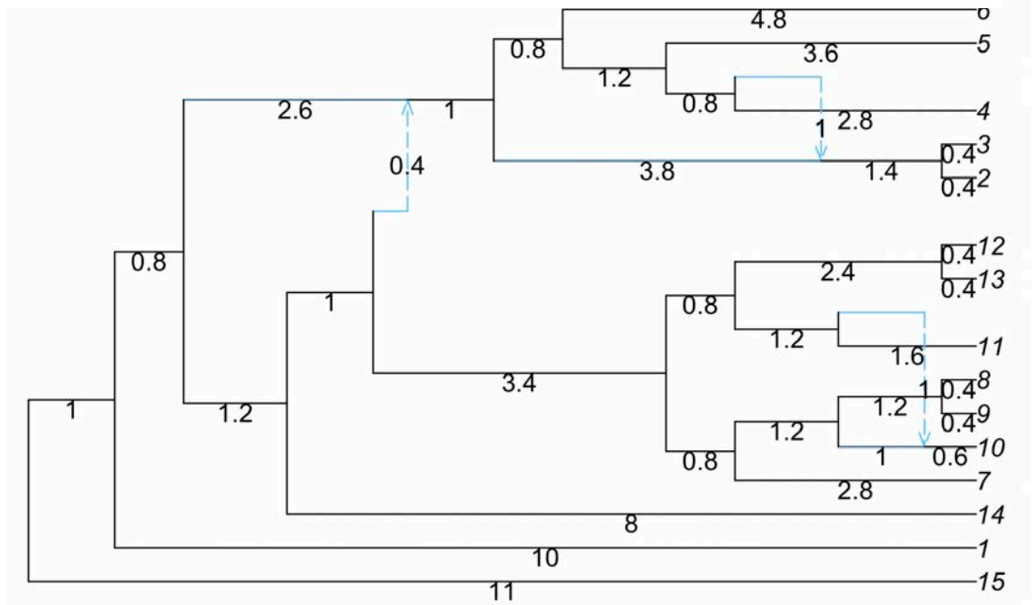
Quantitative Biology > Populations and Evolution

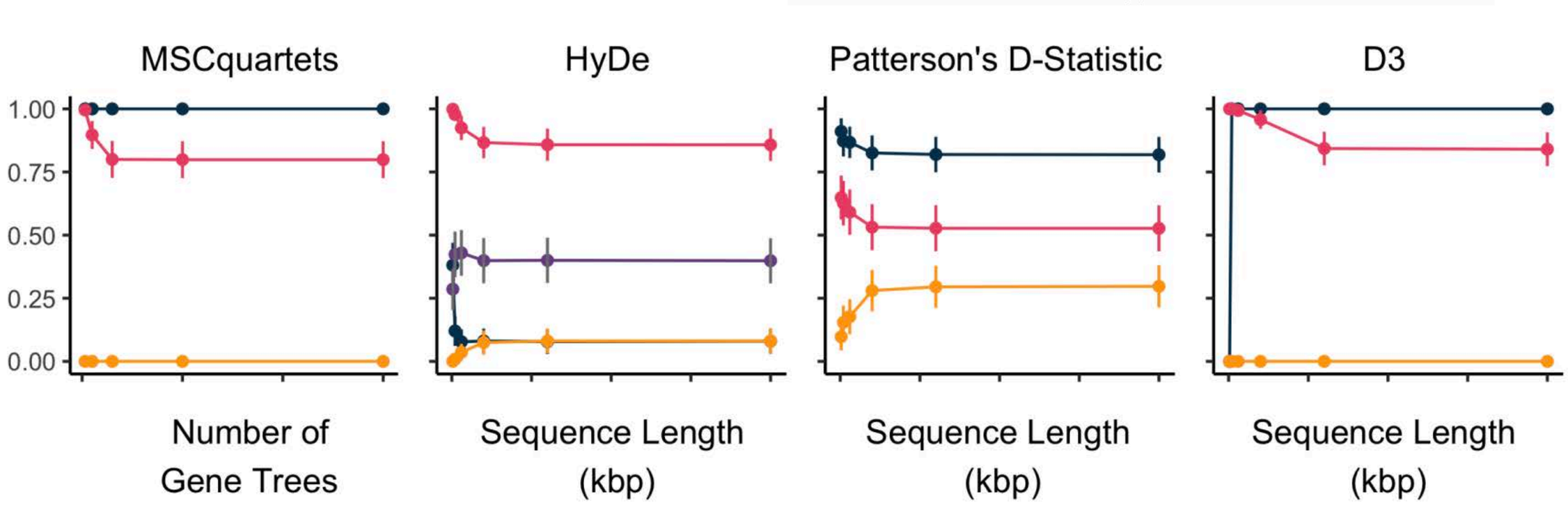
[Submitted on 1 Nov 2022]

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

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

Ancient gene flow distorts discovery rates for hybrid detection methods





(See also Kong & Kubatko, 2021)

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

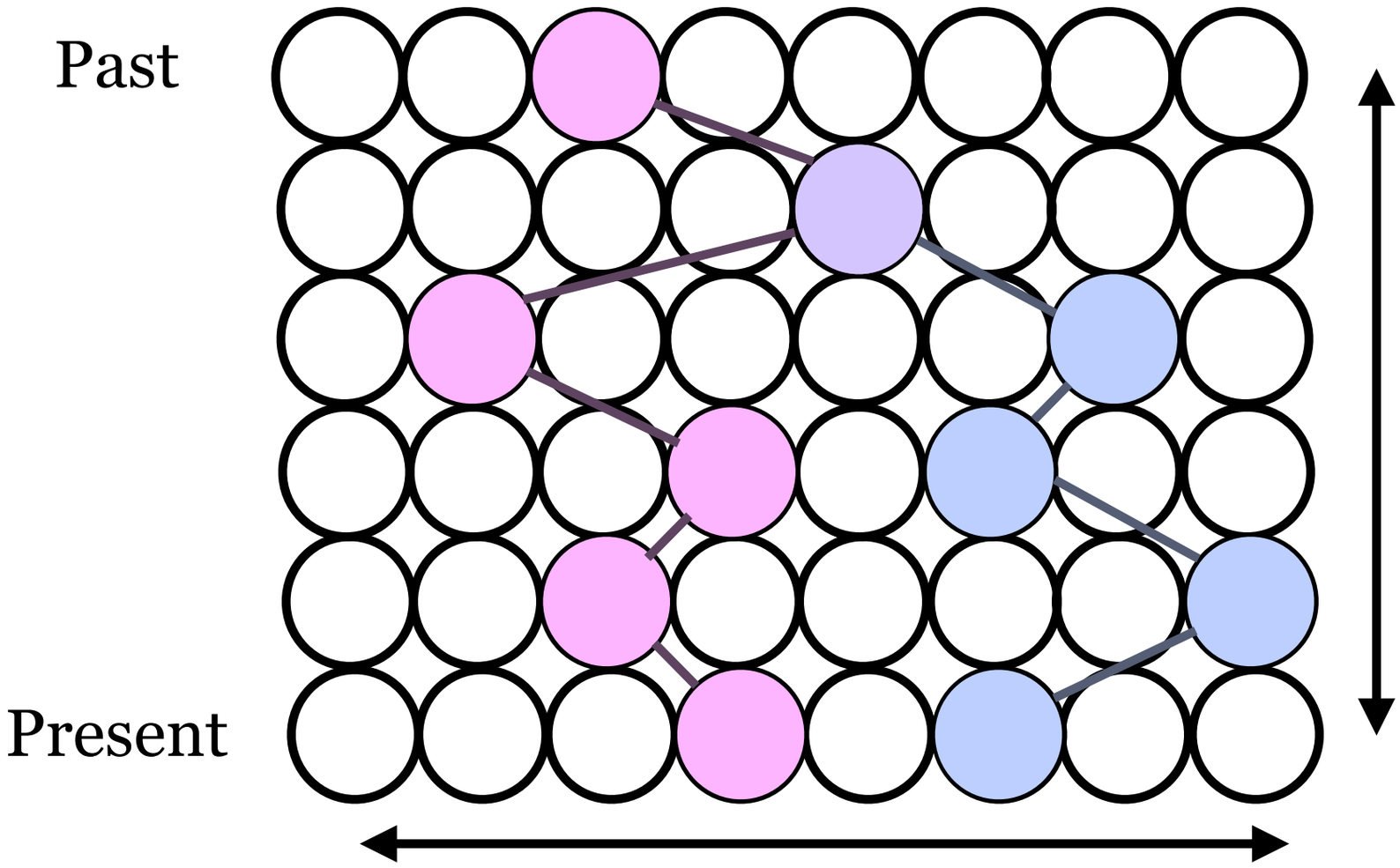
SNaQ

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

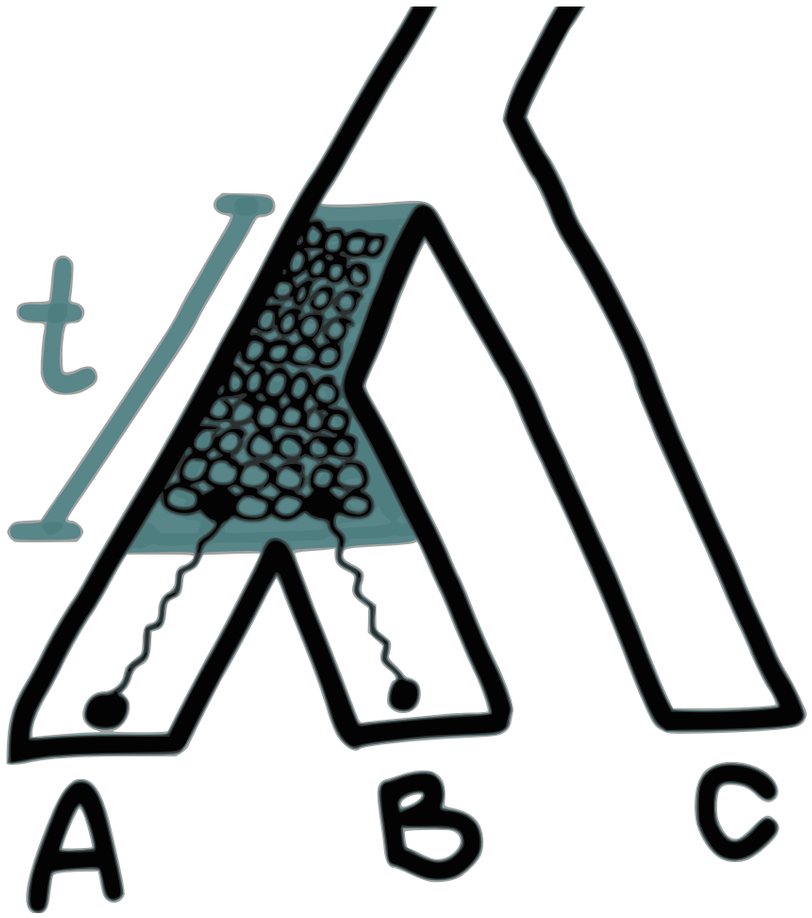
Coalescent model within 1 population



Probability of no coalescence in g generations: $\left(1-\frac{1}{N}\right)^{g}$

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

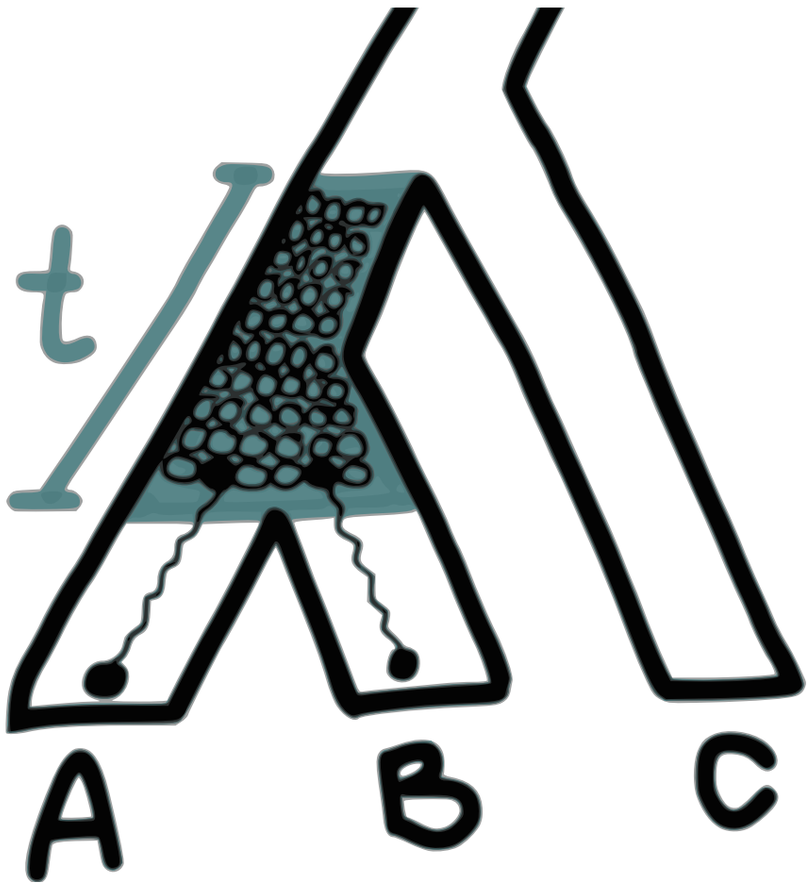
Multispecies coalescent on a tree



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

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

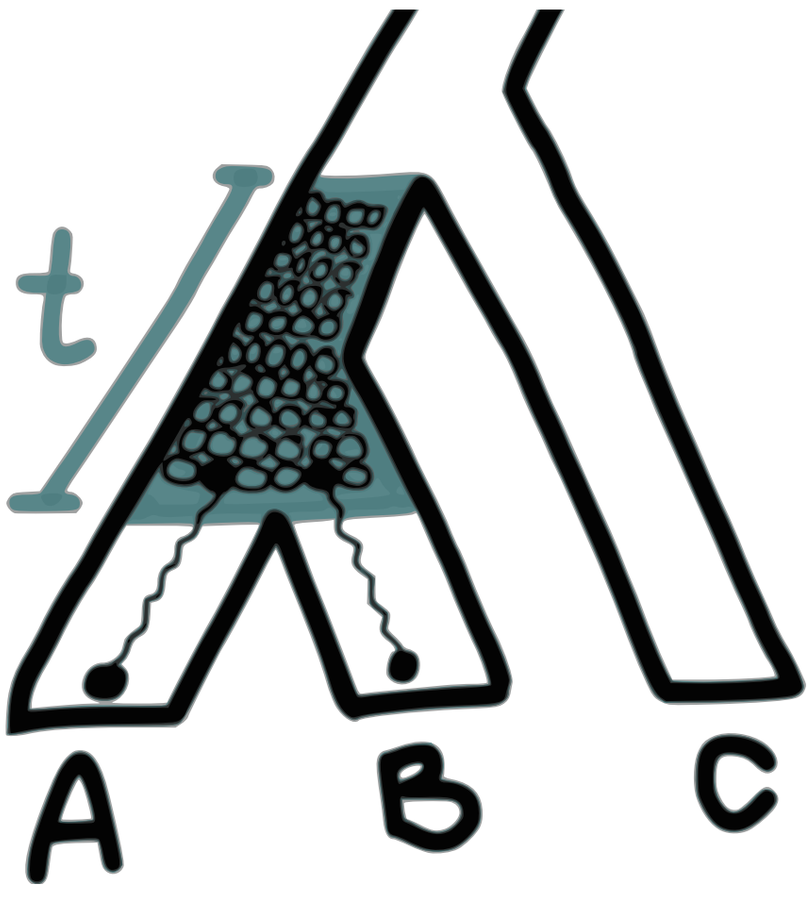
Multispecies coalescent on a tree



$$\mathbb{P}\left(\begin{array}{ccc} & \wedge & \\ A & B & C \end{array}\right) =$$

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

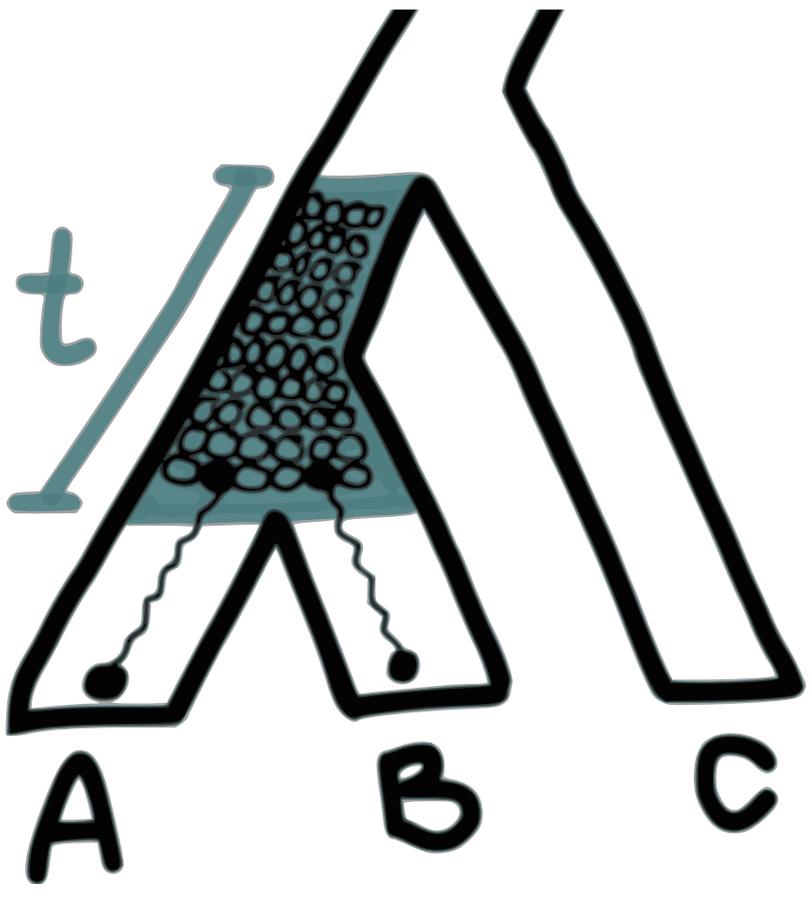
Multispecies coalescent on a tree



$$\mathbb{P}\left(\begin{array}{ccc} & \wedge & \\ A & B & C \end{array}\right) = 1 - e^{-t}$$

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

Multispecies coalescent on a tree



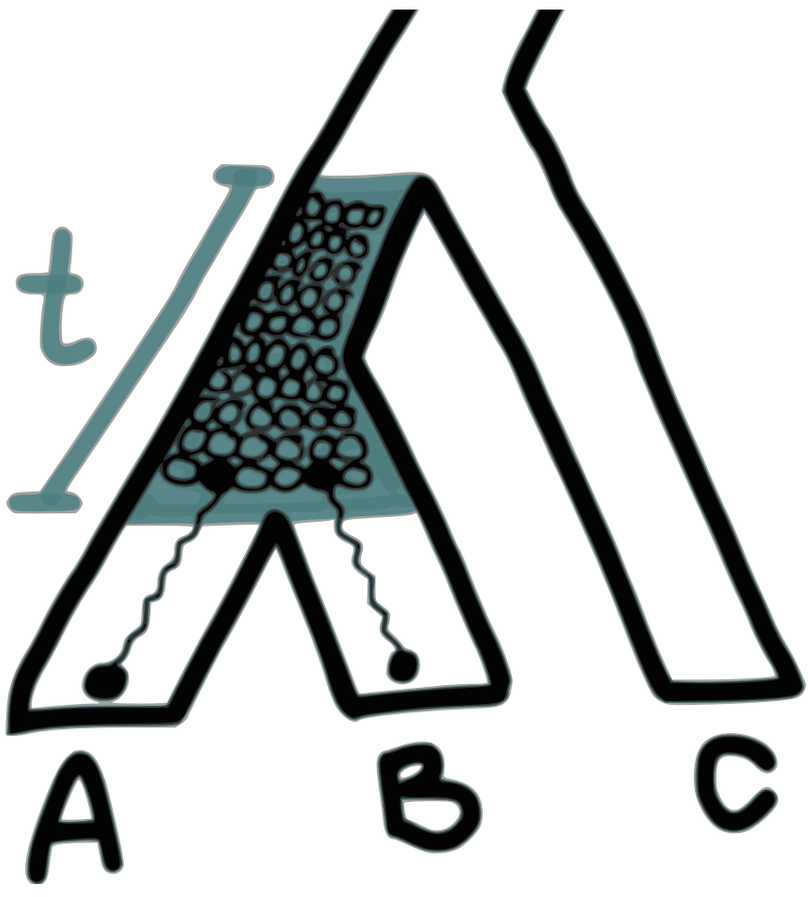
$$P(T > t) = e^{-t}$$

$$\mathbb{P}\left(\begin{array}{ccc} & \wedge & \\ A & B & C \end{array}\right) =$$

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

$$+$$

Multispecies coalescent on a tree



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

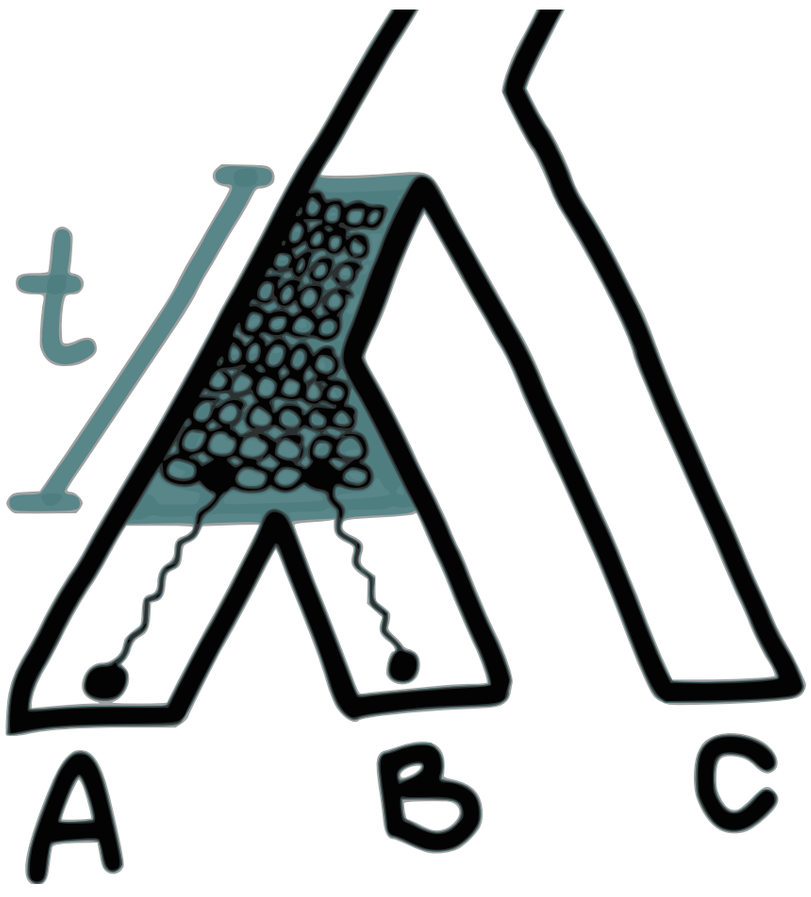
$$\mathbb{P}\left(\begin{array}{ccc} & \wedge & \\ A & B & C \end{array}\right) =$$

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

$$+$$

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

Multispecies coalescent on a tree



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

$$\mathbb{P}\left(\begin{array}{ccc} & \wedge & \\ A & B & C \end{array}\right) =$$

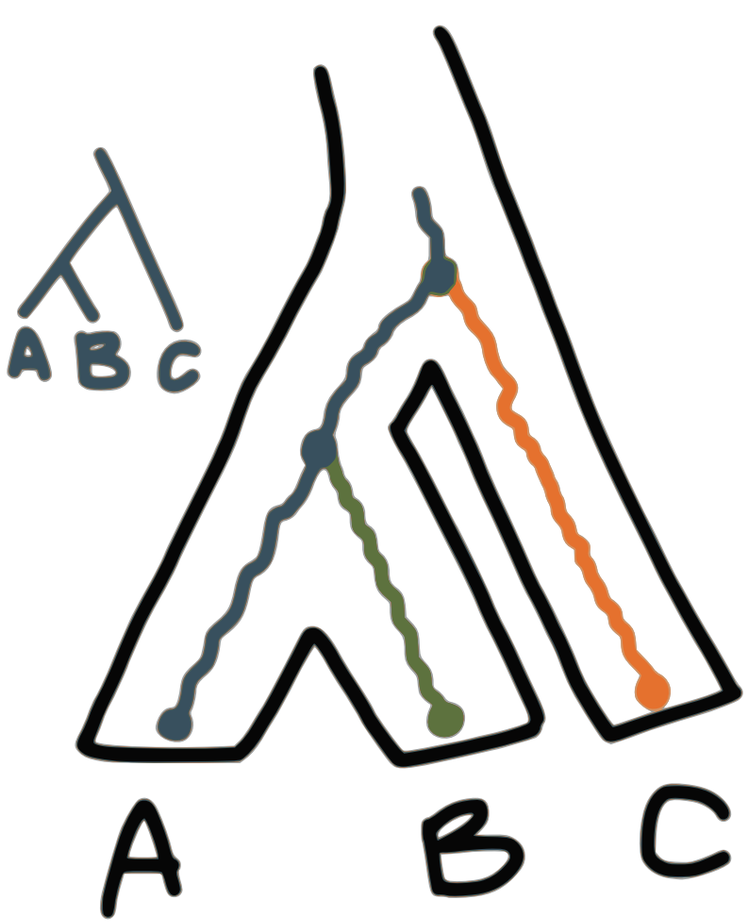
$$1 - e^{-t}$$

$$+$$

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

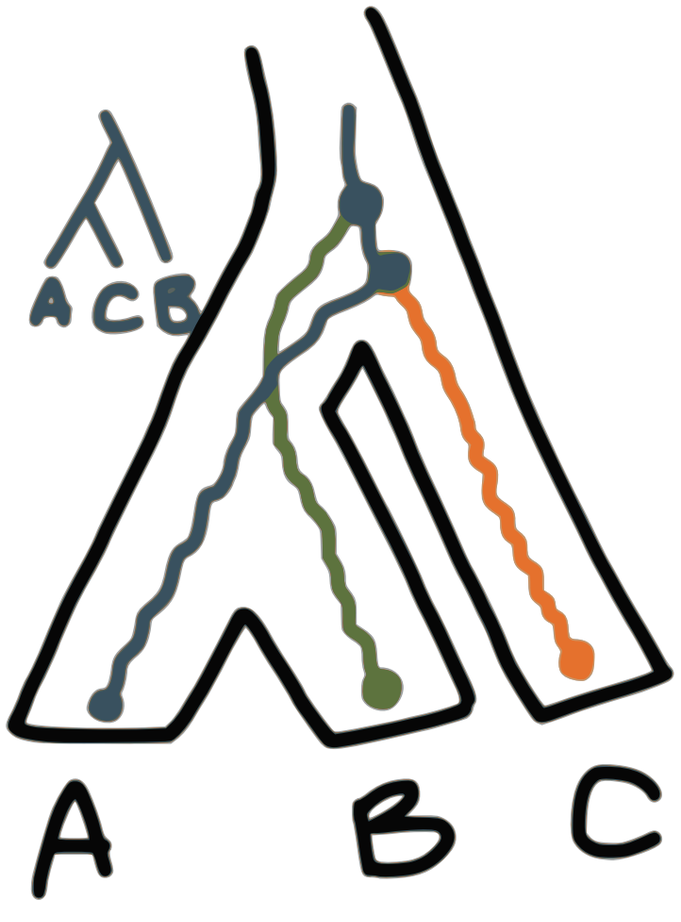
$$= 1 - \frac{2}{3} e^{-t}$$

Multispecies coalescent on a tree



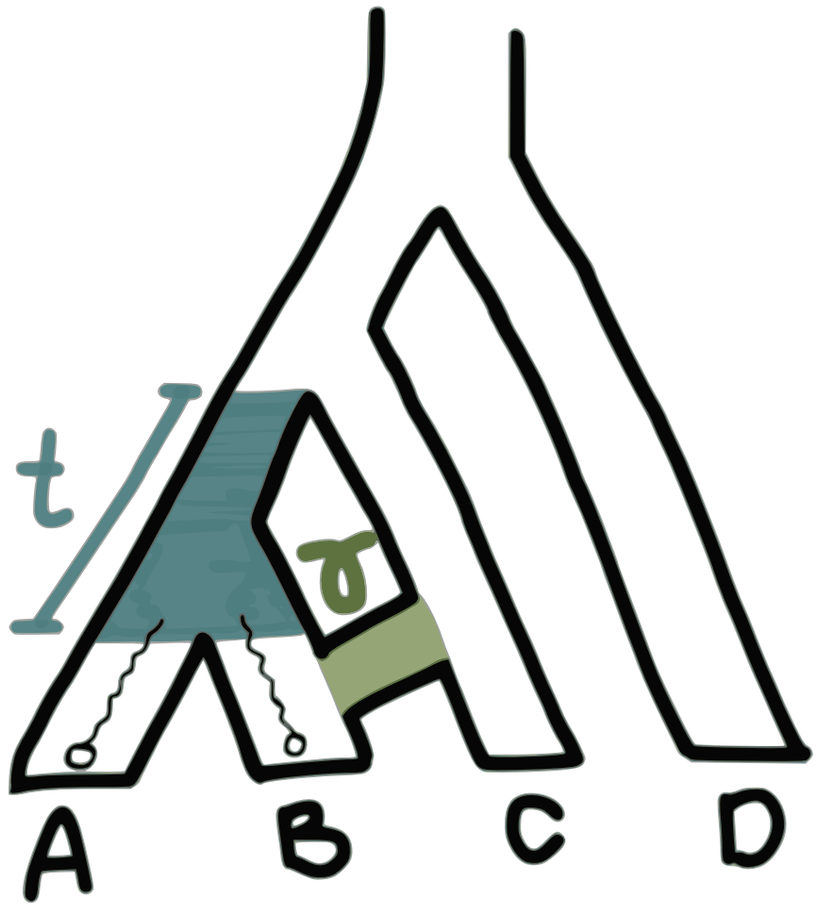
$$1 - \frac{2}{3}e^{-t}$$

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



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

Multispecies coalescent on a network

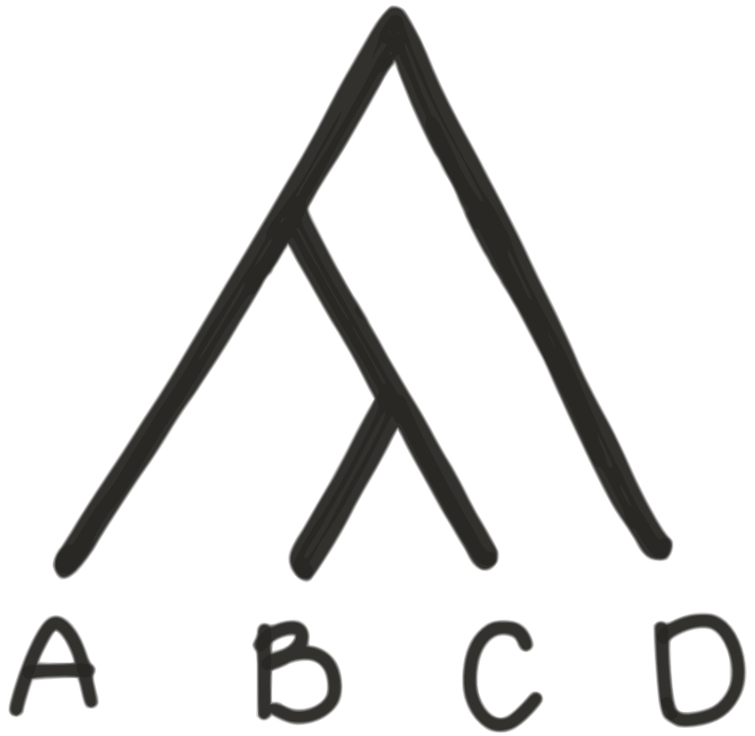
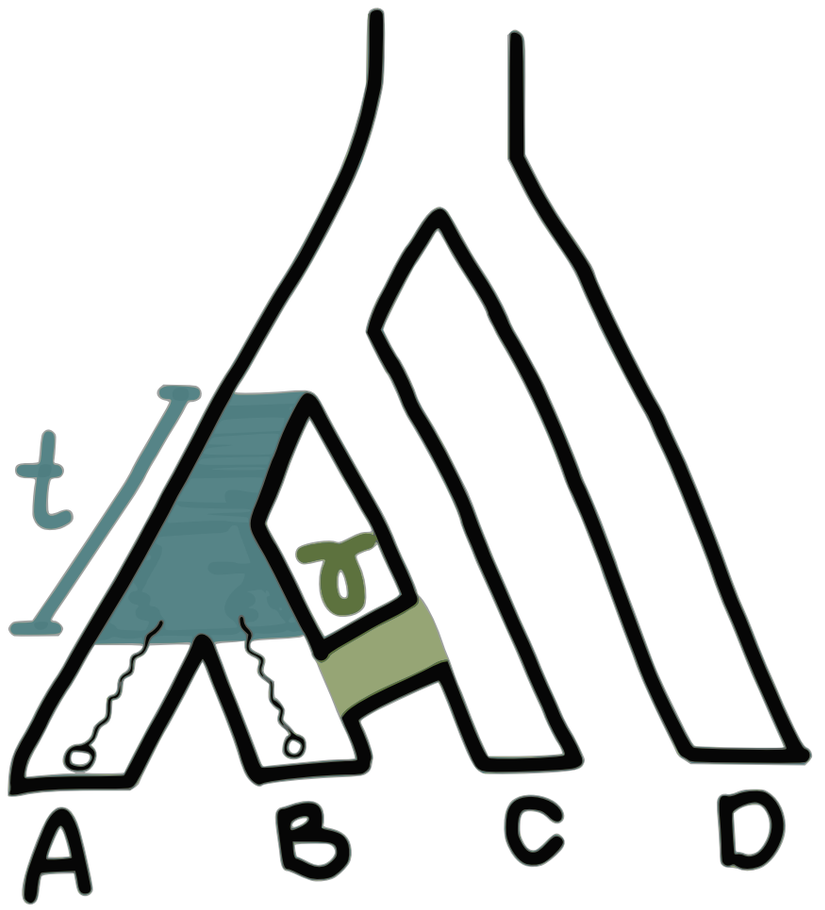


(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

Multispecies coalescent on a network

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

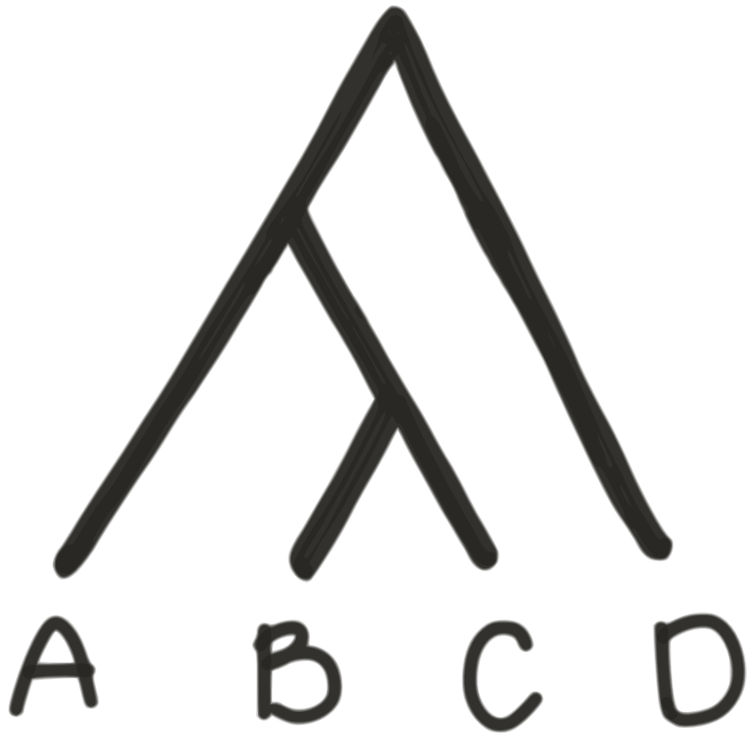
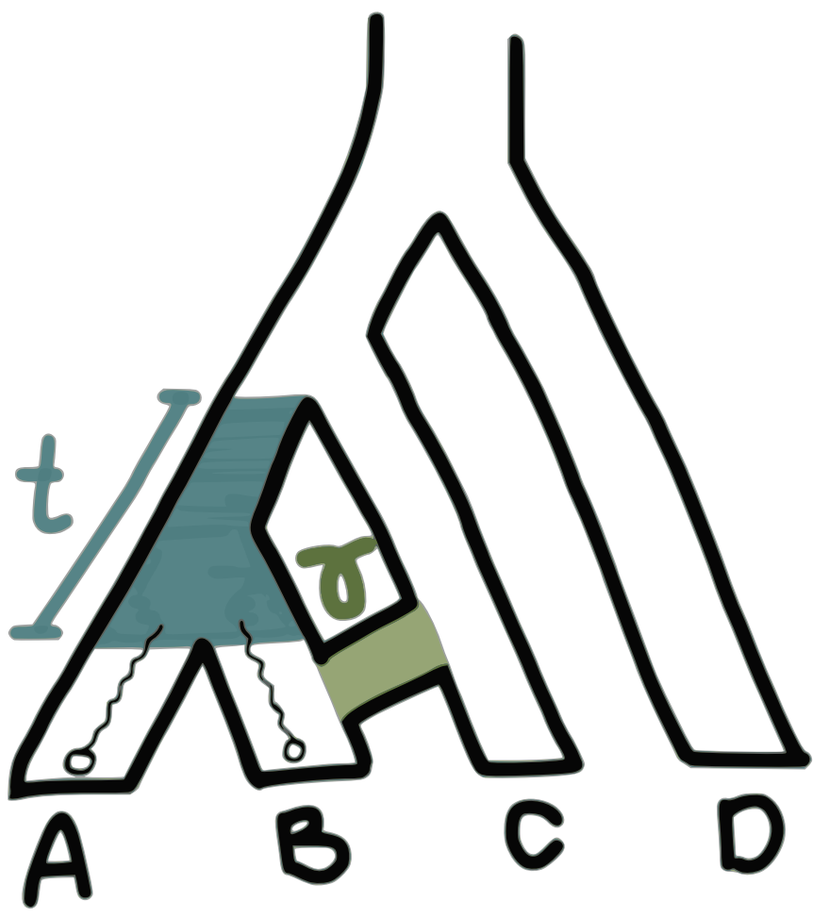
Multispecies coalescent on a network

$P($  $|$  $)$

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

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

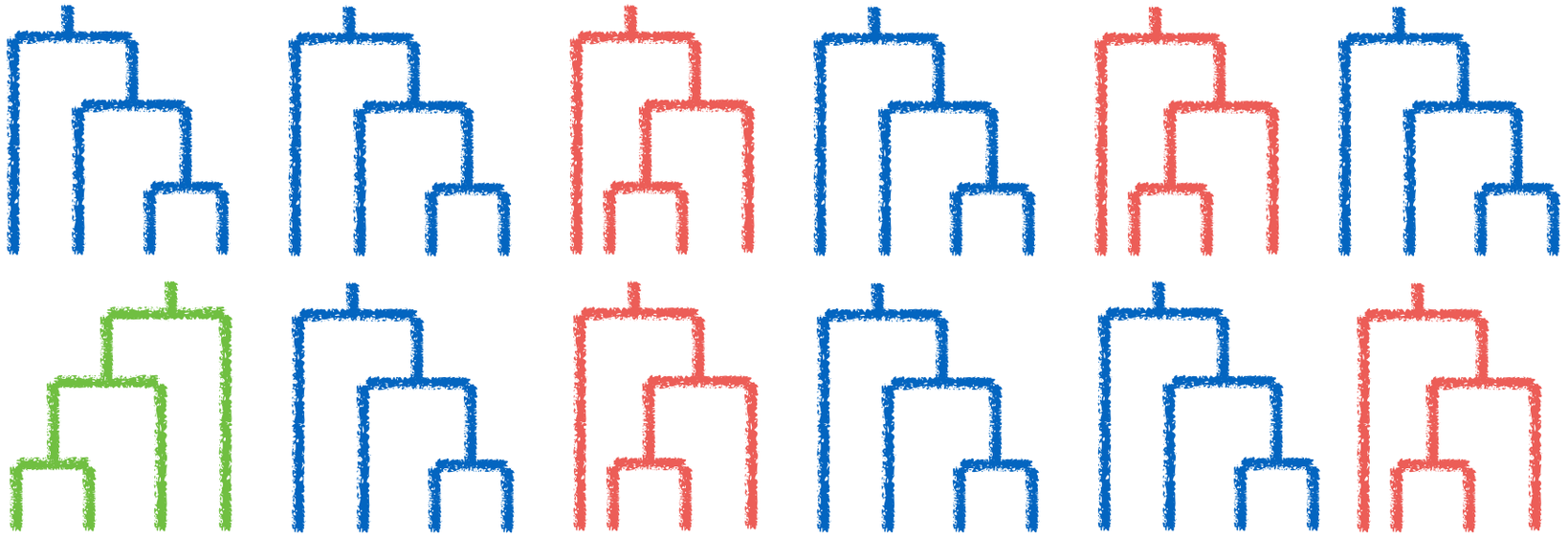
Multispecies coalescent on a network

$P($  $|$  $)$

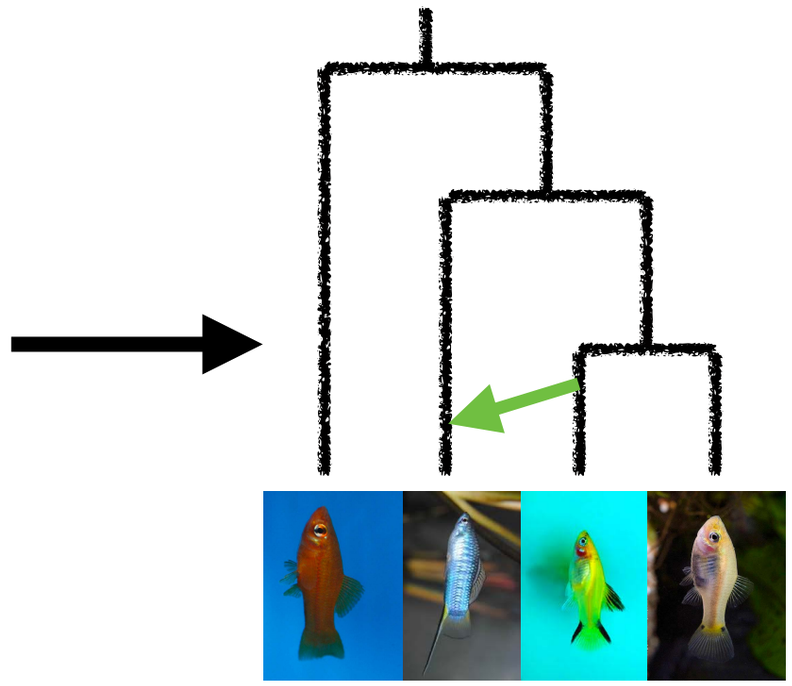
$$CF_{BC|AD}(t, t_2, \gamma) = (1-\gamma)\frac{1}{3}e^{-t} + \gamma(1 - \frac{2}{3}e^{-t_2})$$

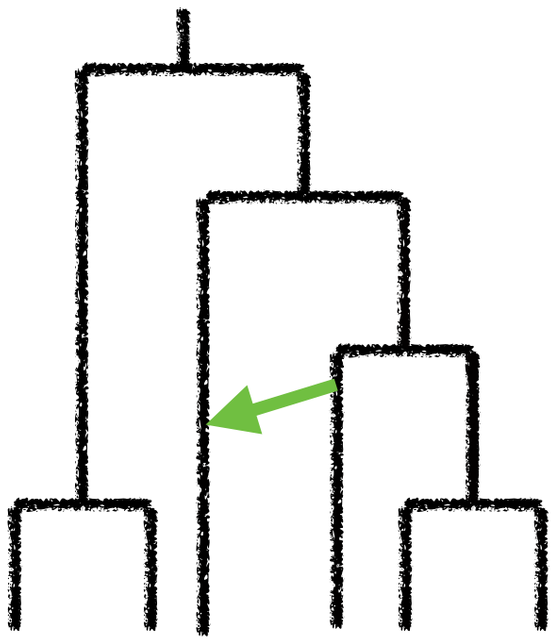
(Meng, Kubatko, 2009)
(Yu, Degnan, Nakhleh, 2012)

Maximum **pseudo**likelihood



Data





Quartet-based inference

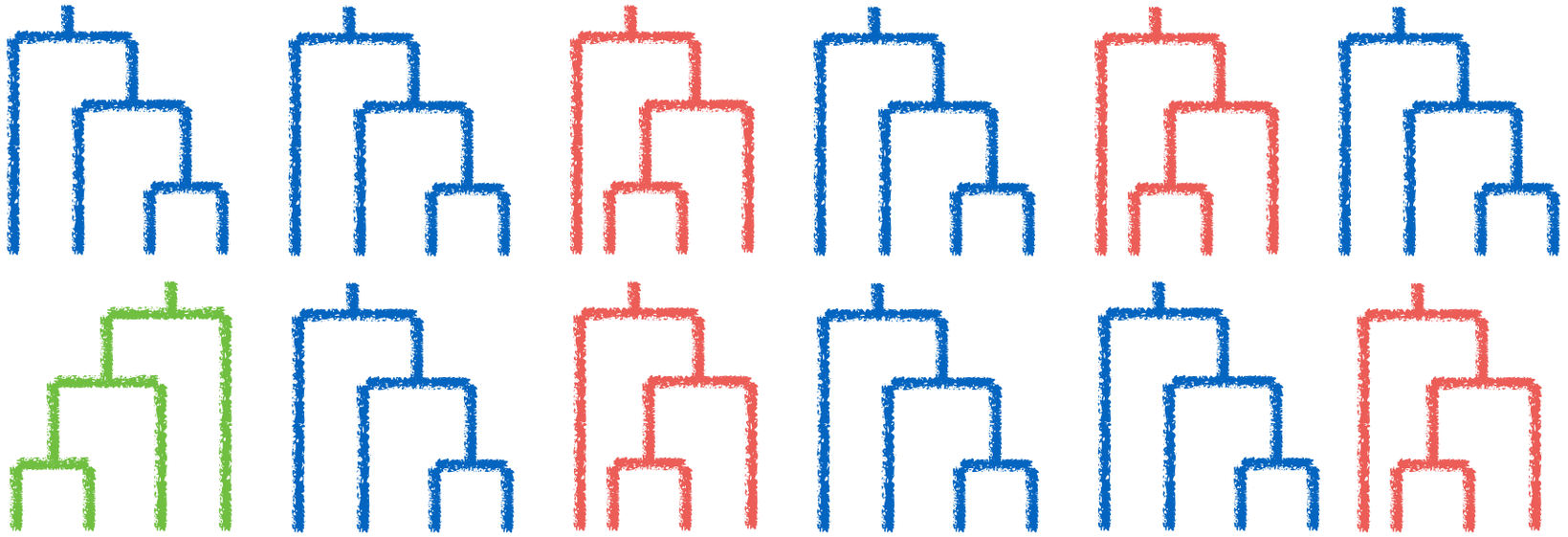
$$\tilde{L}(network) = \prod L(quartet)$$

(S-L, Ané, 2016, PLoS Genetics)

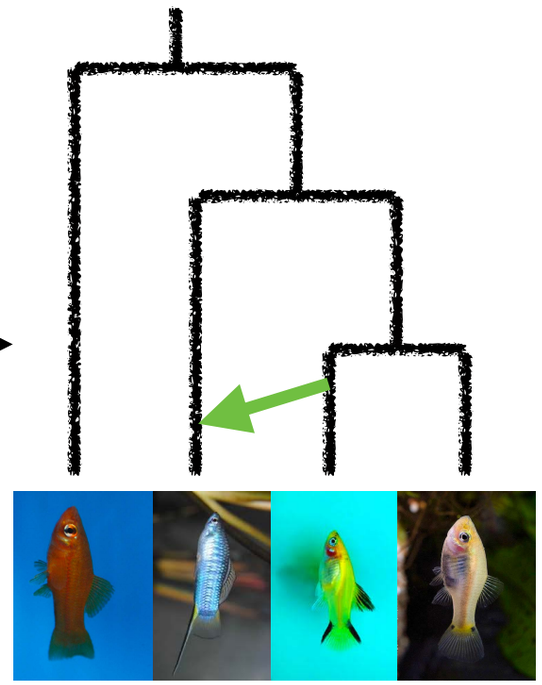
`www.github.com/CRSL4/PhyloNetworks`

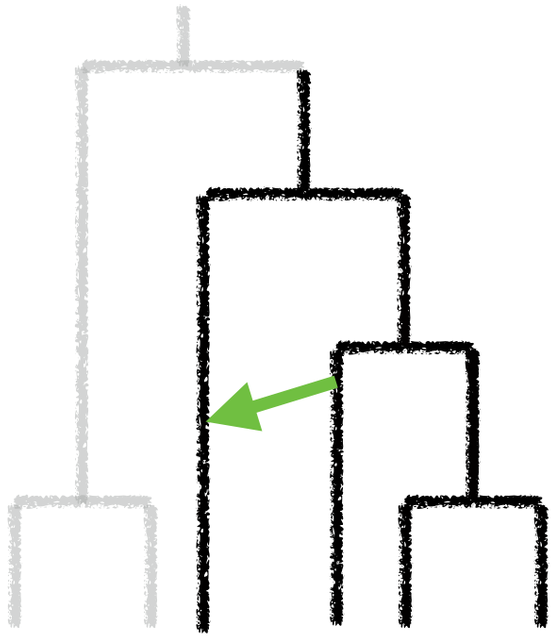
SNAQ julia

Maximum **pseudo**likelihood



Data





Quartet-based inference

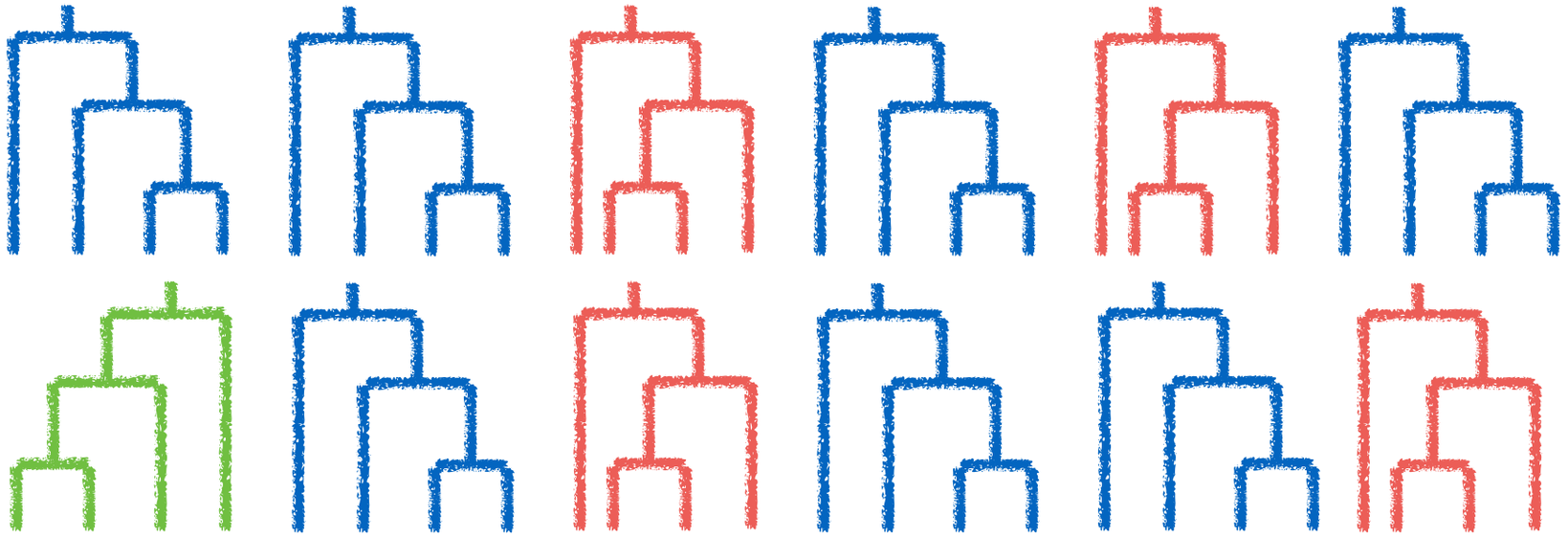
$$\tilde{L}(network) = \prod L(quartet)$$

(S-L, Ané, 2016, PLoS Genetics)

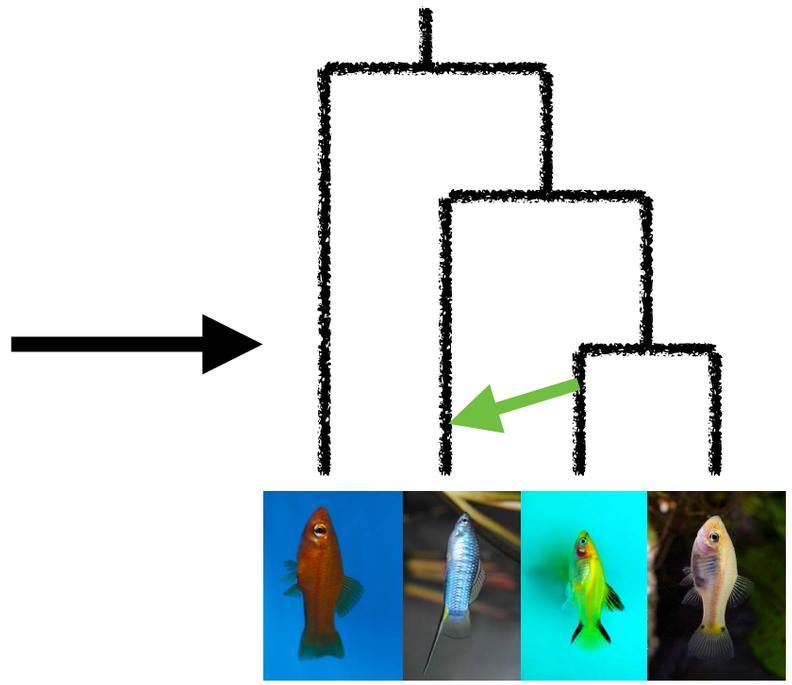
`www.github.com/CRSL4/PhyloNetworks`

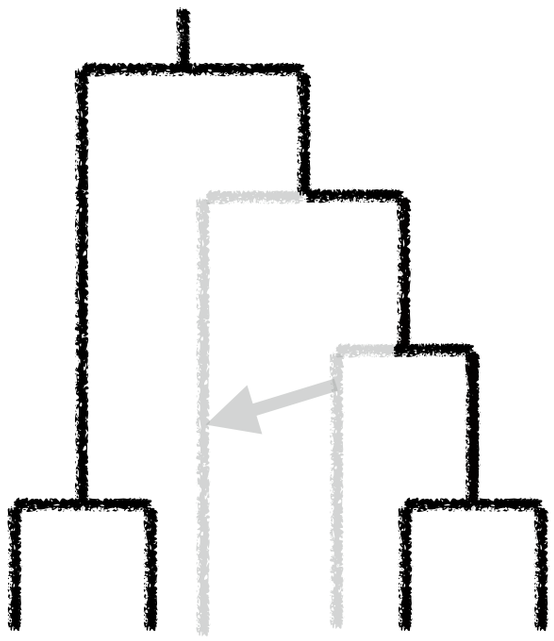
SNAQ julia

Maximum **pseudo**likelihood



Data





Quartet-based inference

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

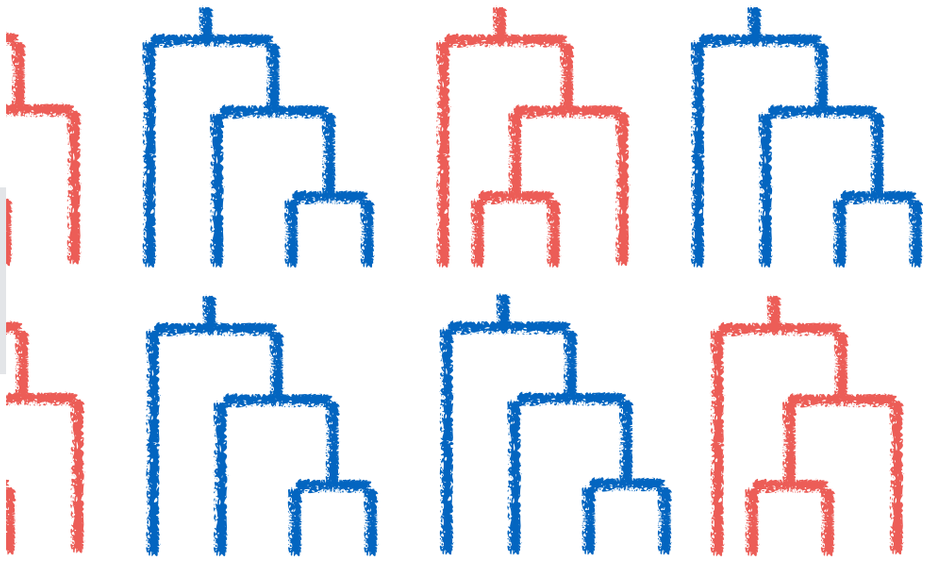
(S-L, Ané, 2016, PLoS Genetics)

`www.github.com/CRSL4/PhyloNetworks`

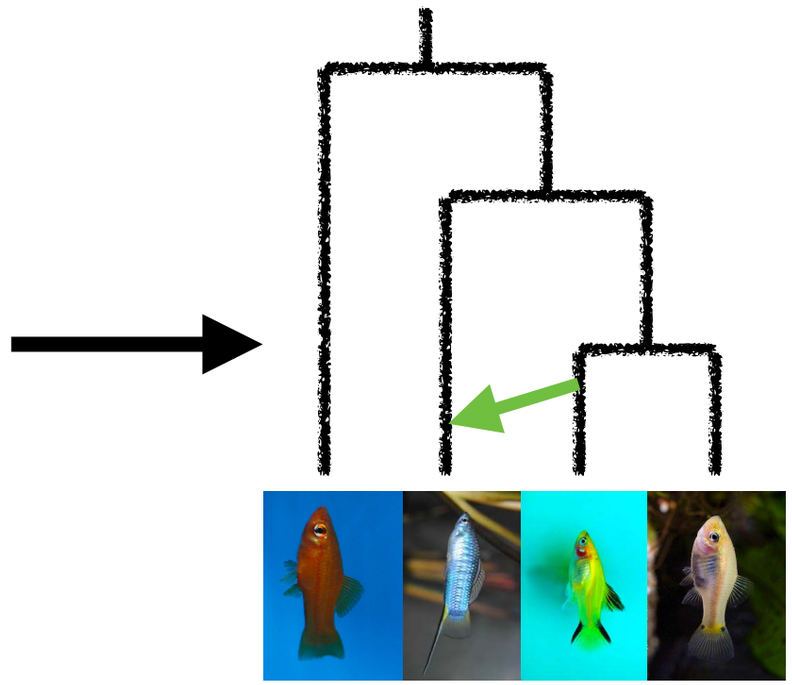
SNAQ julia

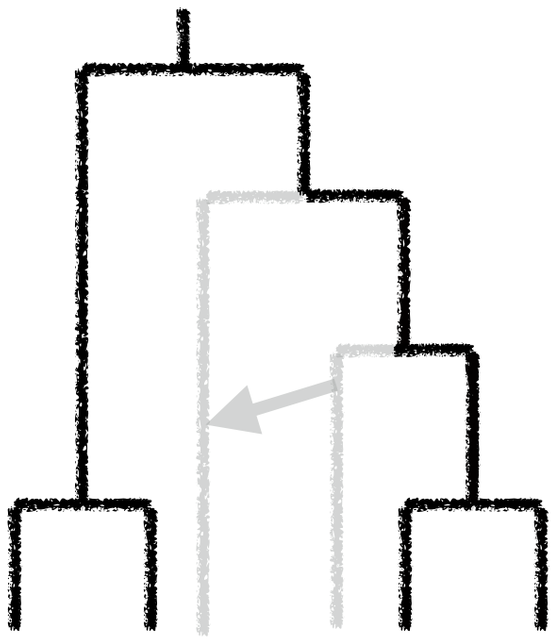
Maximum **pseudo**likelihood

Unrooted gene trees	No rooting error
No branch lengths	No molecular clock assumption
Concordance factors	Account for tree estimation error



Data





Quartet-based inference

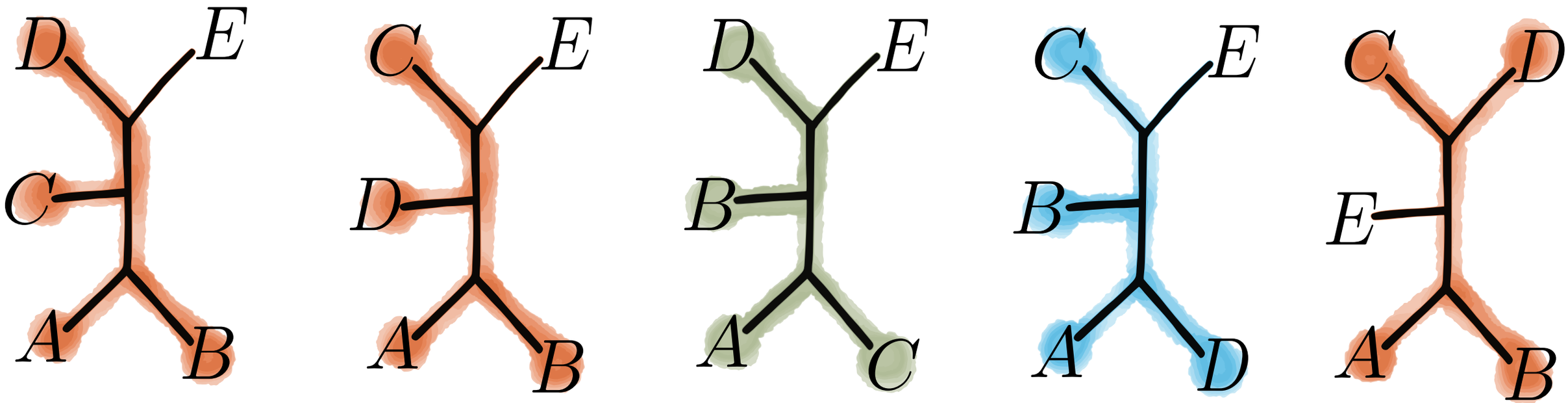
$$\tilde{L}(network) = \prod L(quartet)$$

(S-L, Ané, 2016, PLoS Genetics)

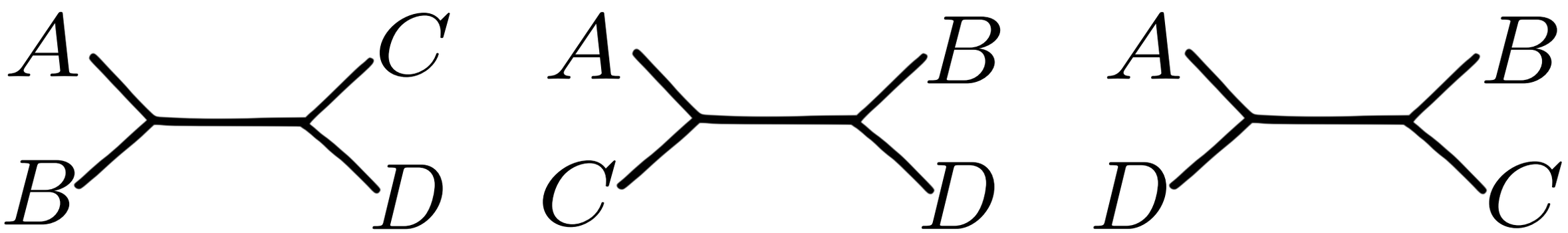
`www.github.com/CRSL4/PhyloNetworks`

SNAQ julia

Quartet-based inference



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



3/5 1/5 1/5

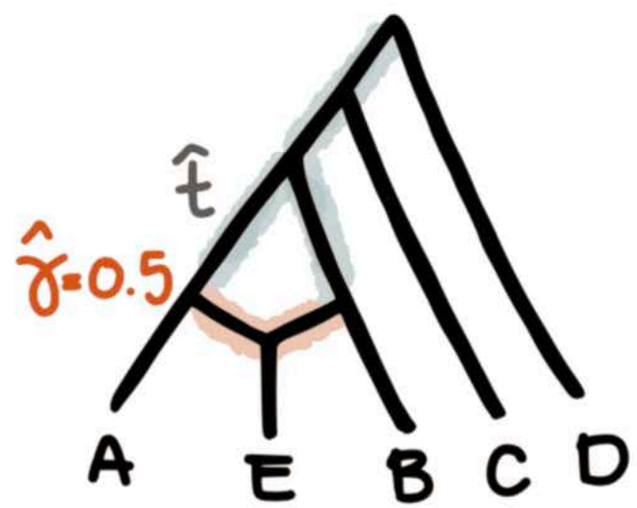
Quartet-based inference

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

$\longrightarrow$

inferred network:

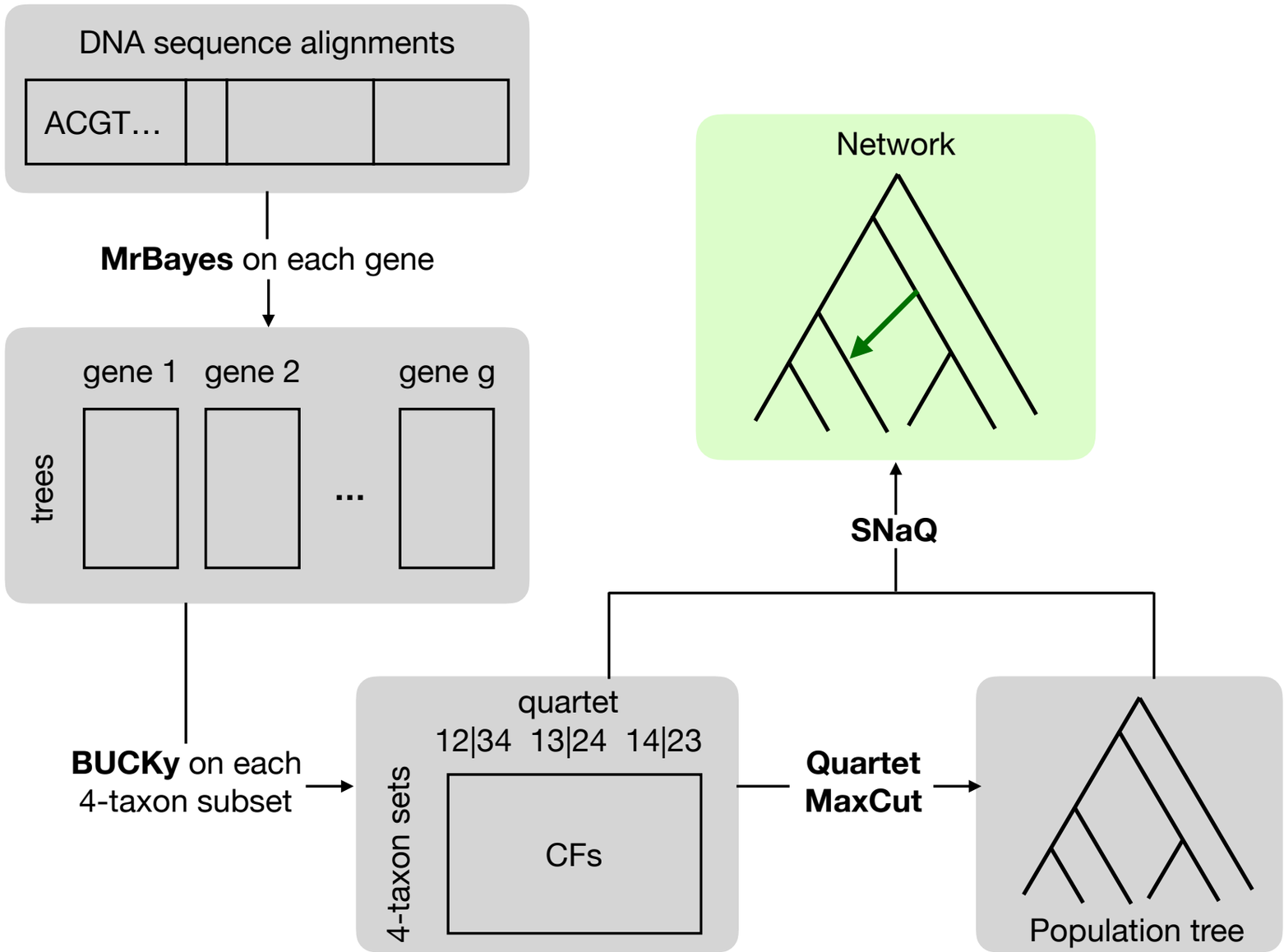


Maximum Pseudo-Likelihood:

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

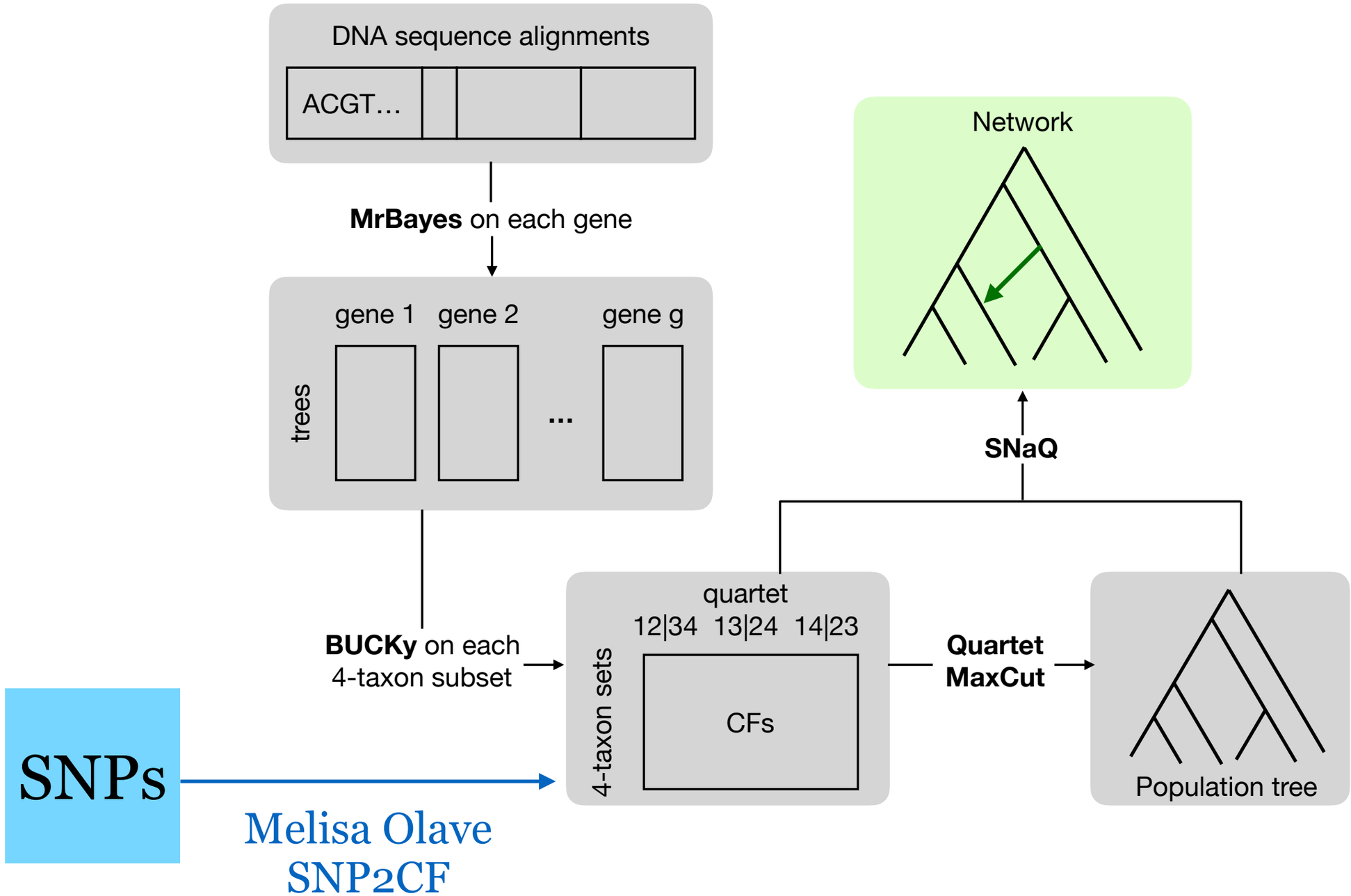
How?

Phylogenetic network



How?

Phylogenetic network



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

Hybrid detection + Network estimation

Bella
Wu

arXiv > q-bio > arXiv:2211.16647

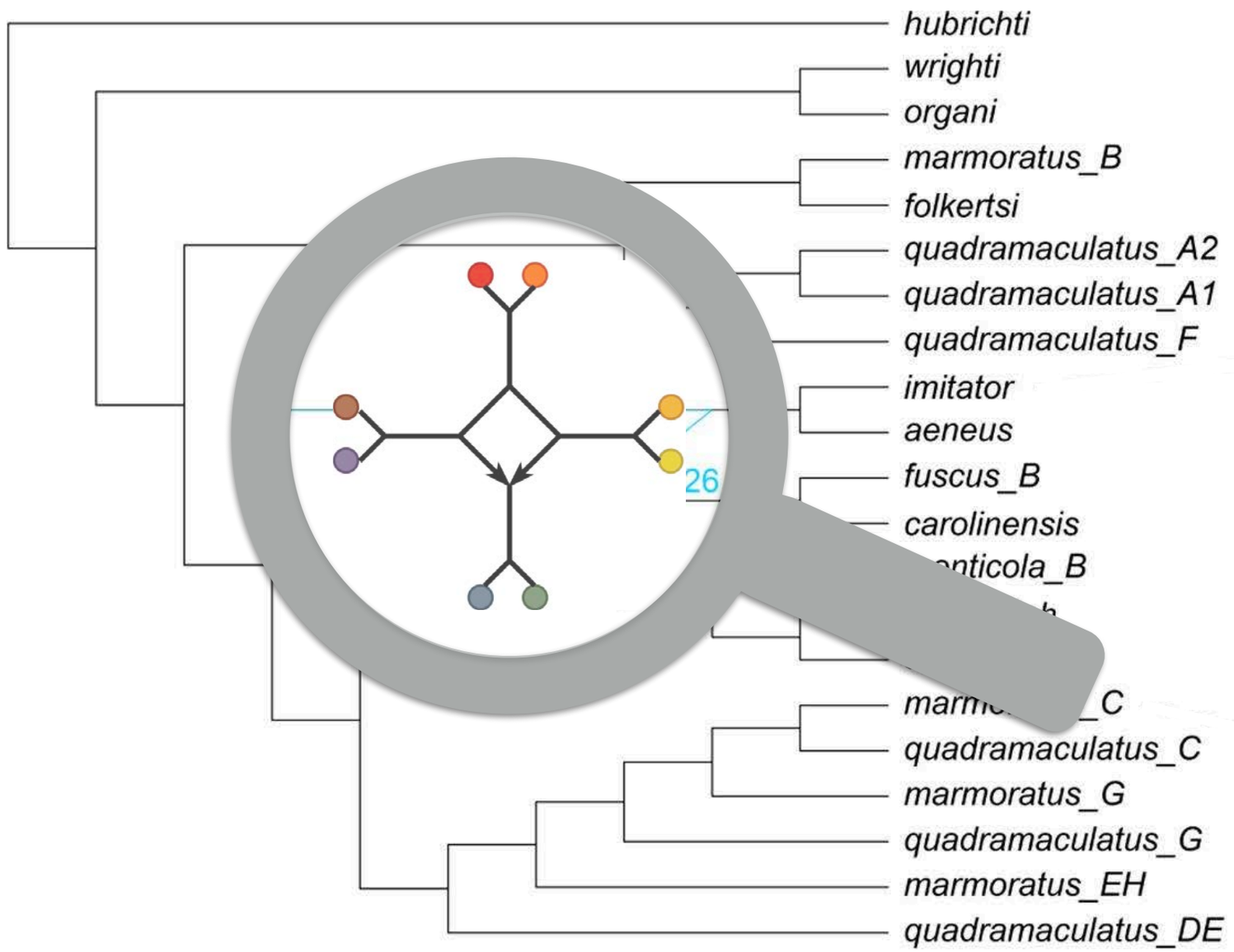
Quantitative Biology > Populations and Evolution

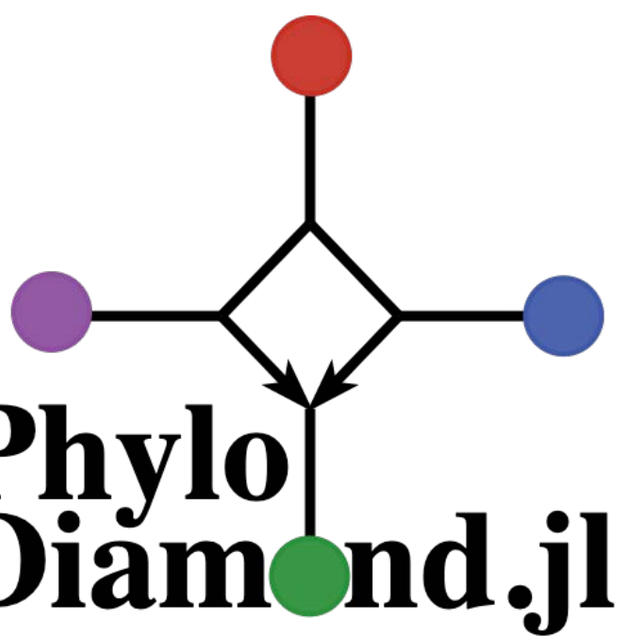
[Submitted on 30 Nov 2022]

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

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes





Bella Wu

arXiv > q-bio > arXiv:2211.16647

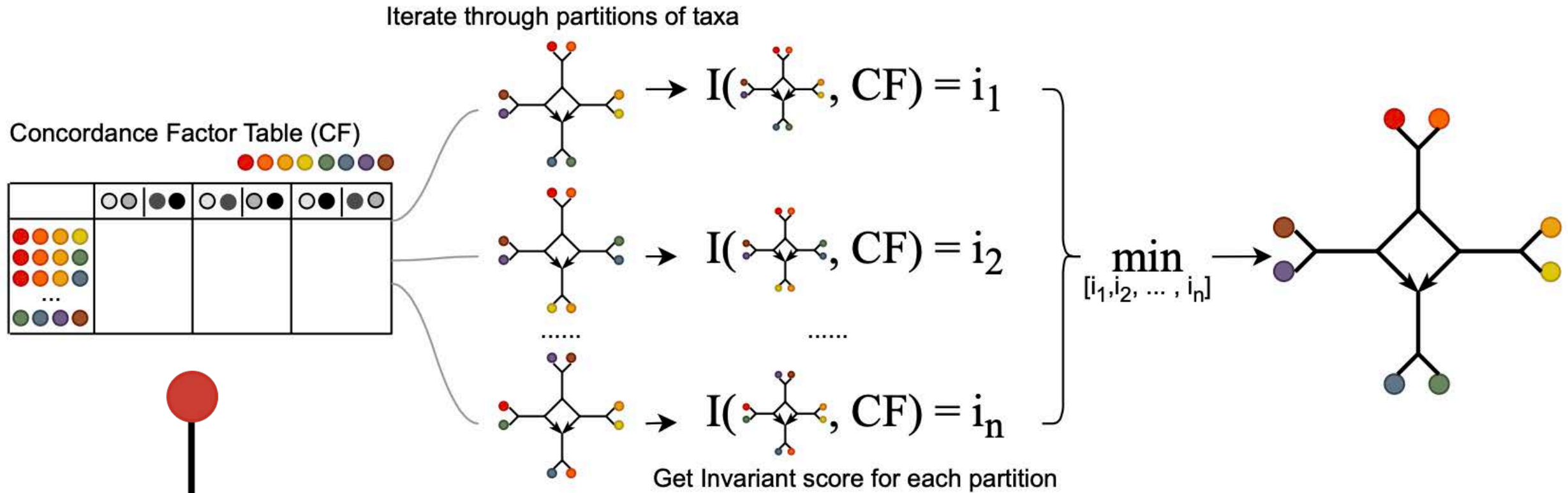
Quantitative Biology > Populations and Evolution

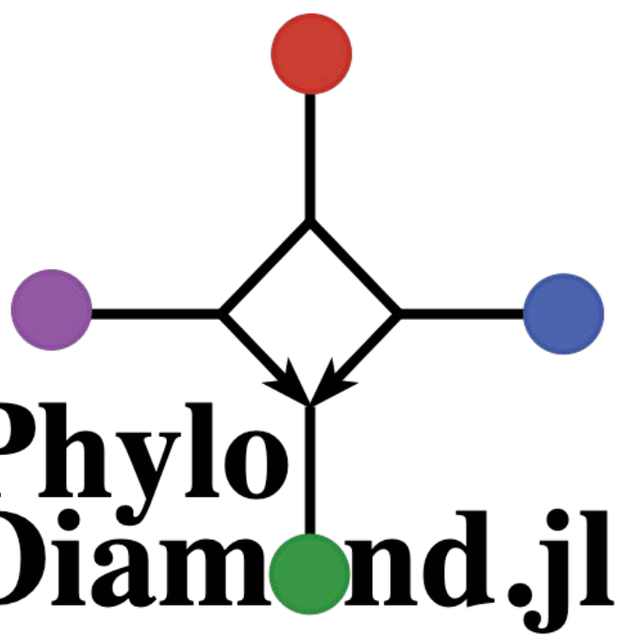
[Submitted on 30 Nov 2022]

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

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes





Bella Wu

arXiv > q-bio > arXiv:2211.16647

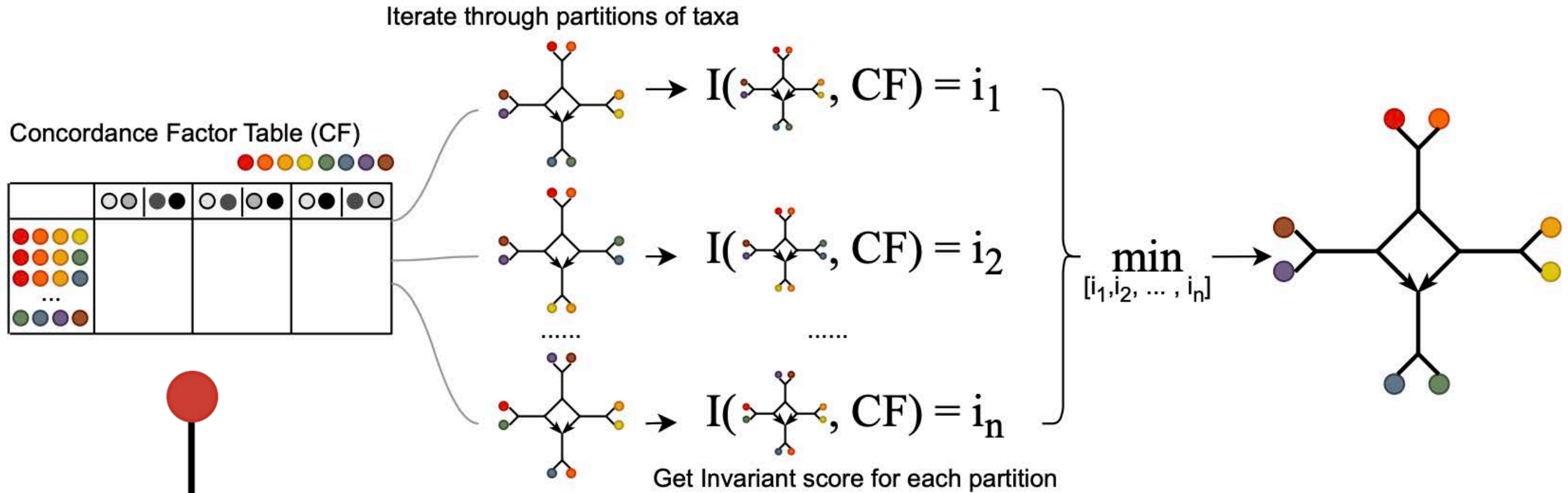
Quantitative Biology > Populations and Evolution

[Submitted on 30 Nov 2022]

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

Zhaoxing Wu, Claudia Solis-Lemus

Identifies hybridization cycles of 4 nodes



PhyloDiamond.jl

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

Bella Wu

arXiv > q-bio > arXiv:2211.16647

Quantitative Biology > Populations and Evolution

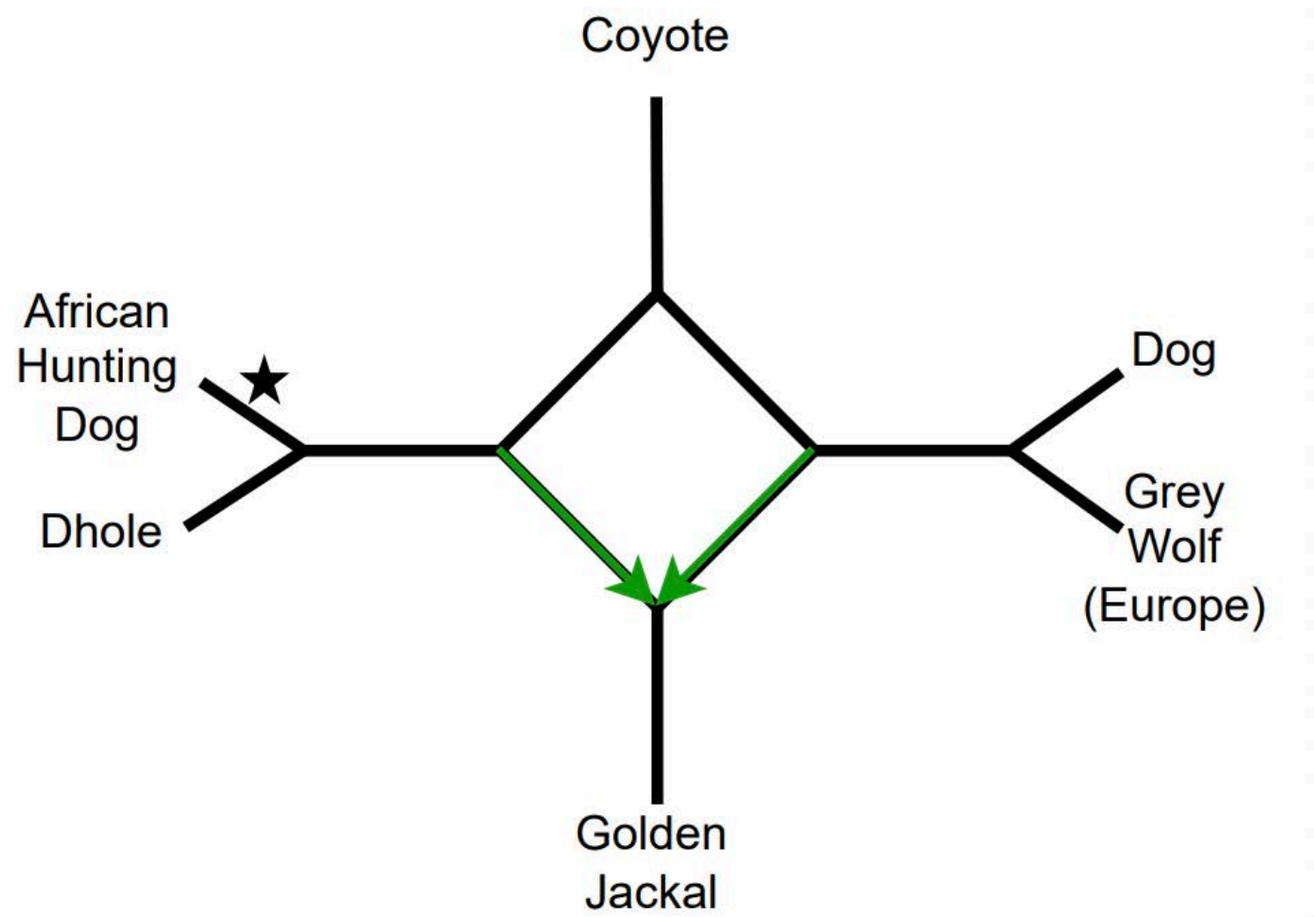
[Submitted on 30 Nov 2022]

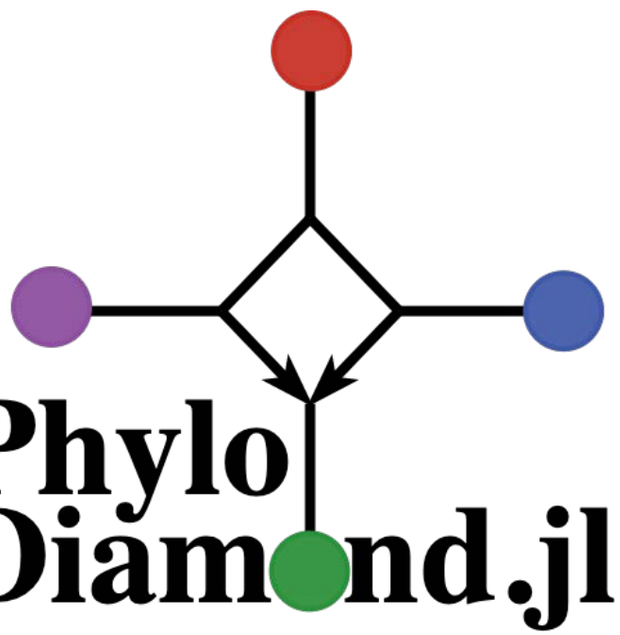
Ultrafast learning of 4-node hybridization cycles in phylogenetic networks using algebraic invariants

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)

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

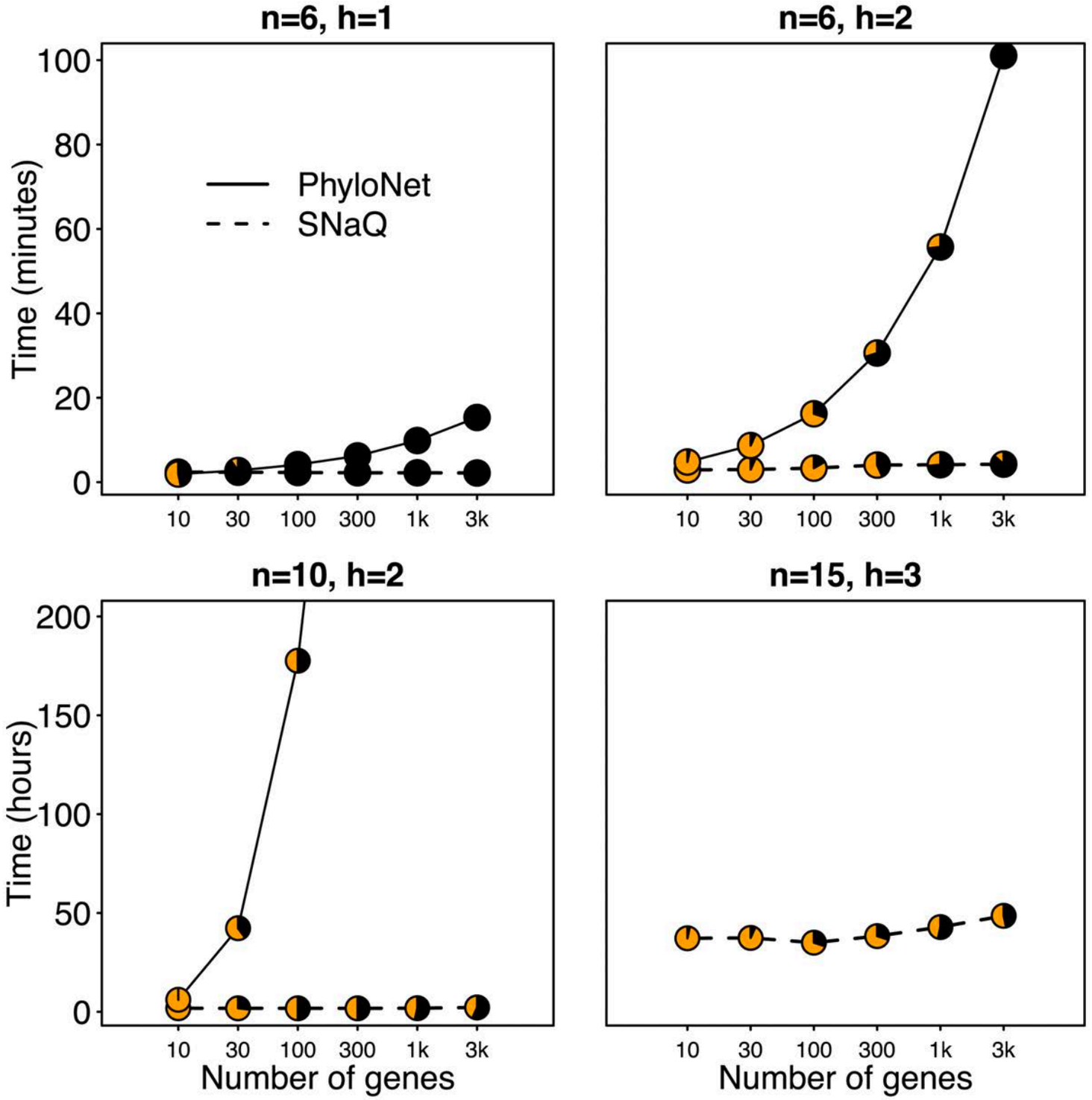
Hybrid detection methods

Hybrid detection + Network estimation

Network challenges

- Scalability
- Identifiability
- Network comparison

Scalability

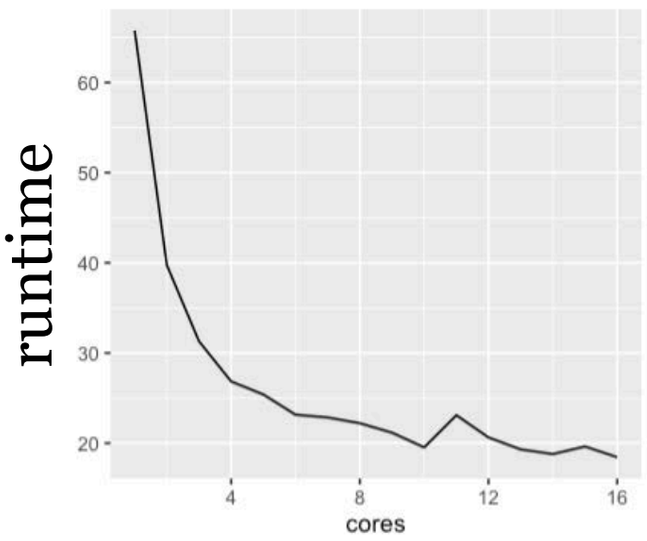


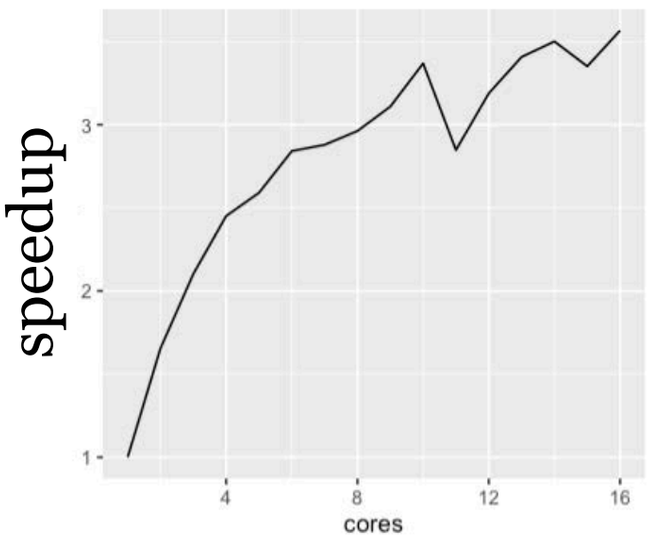
(Solís-Lemus, Ané, 2016, PLoS Genetics)

SNaQ is faster, but is it fast?

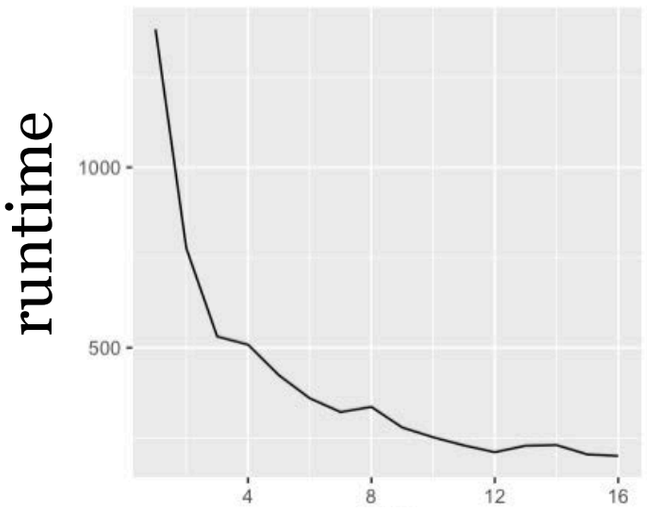
Tyler Chafin

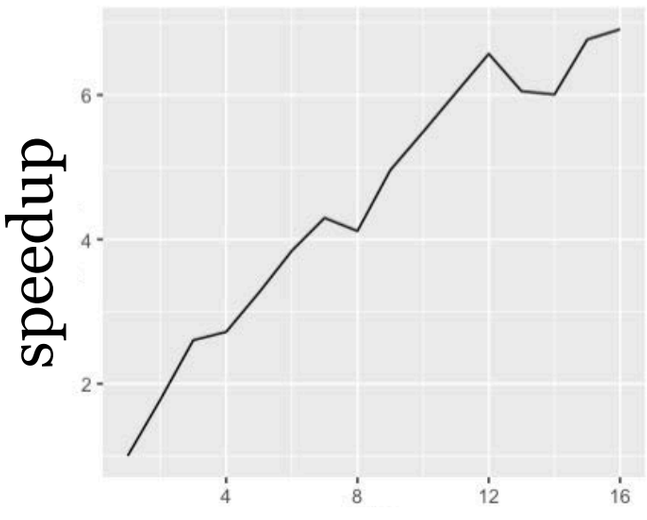
n=10





n=20

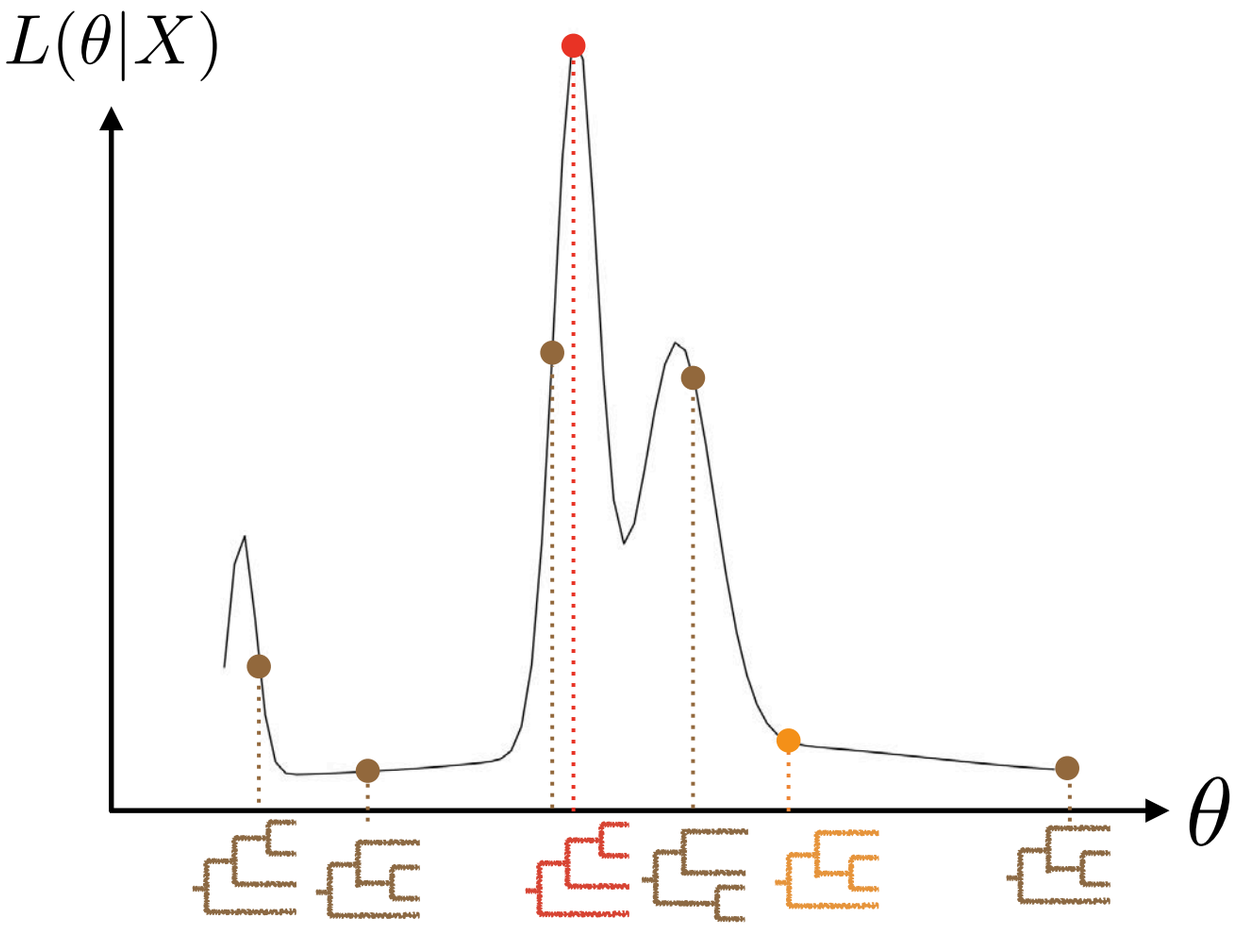


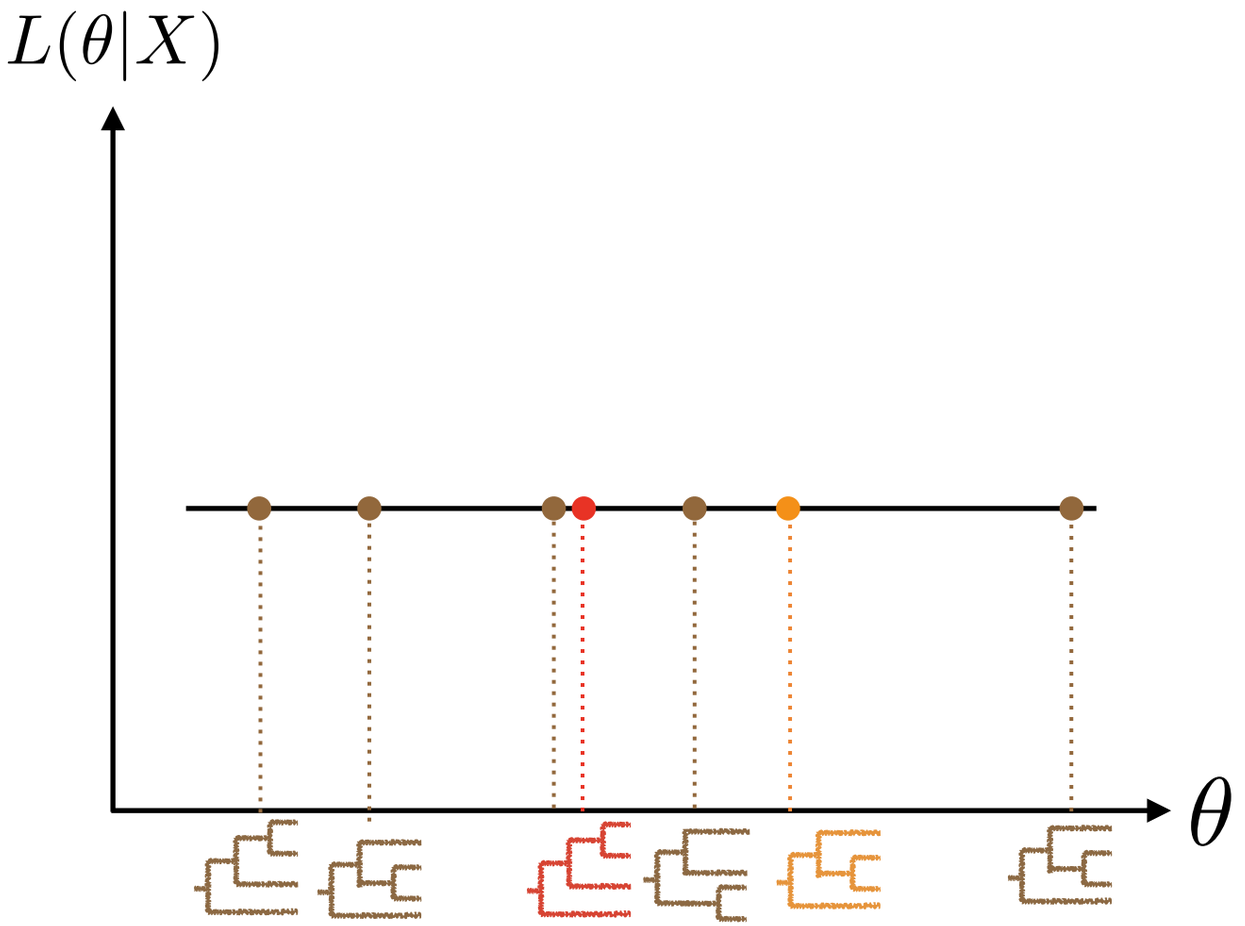


- Parallel quartet likelihood computation
- Weights to quartets: deviations from expected CFs
- Sampling of quartets without biasing inference

SNaQ 2.0 coming up soon!

Identifiability





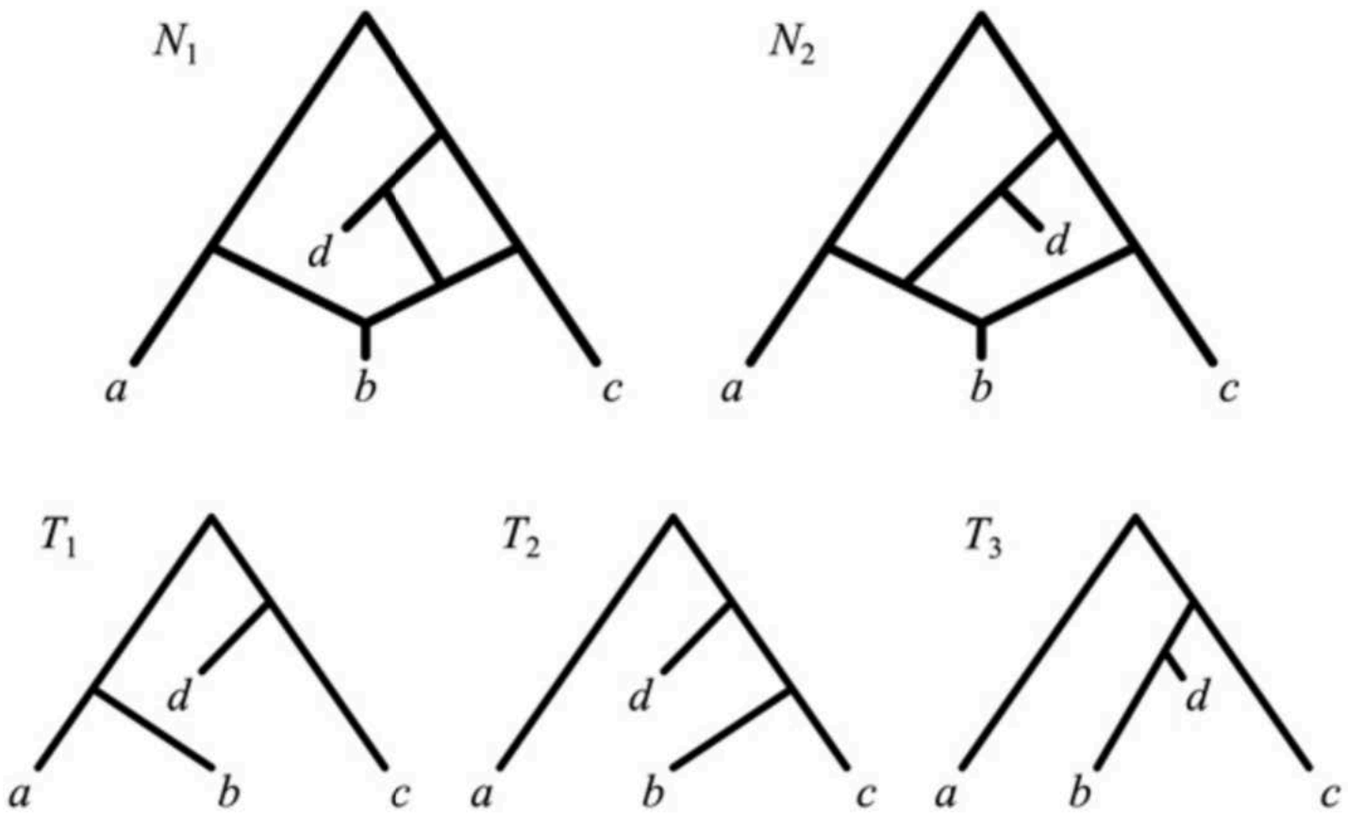
PLOS | COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

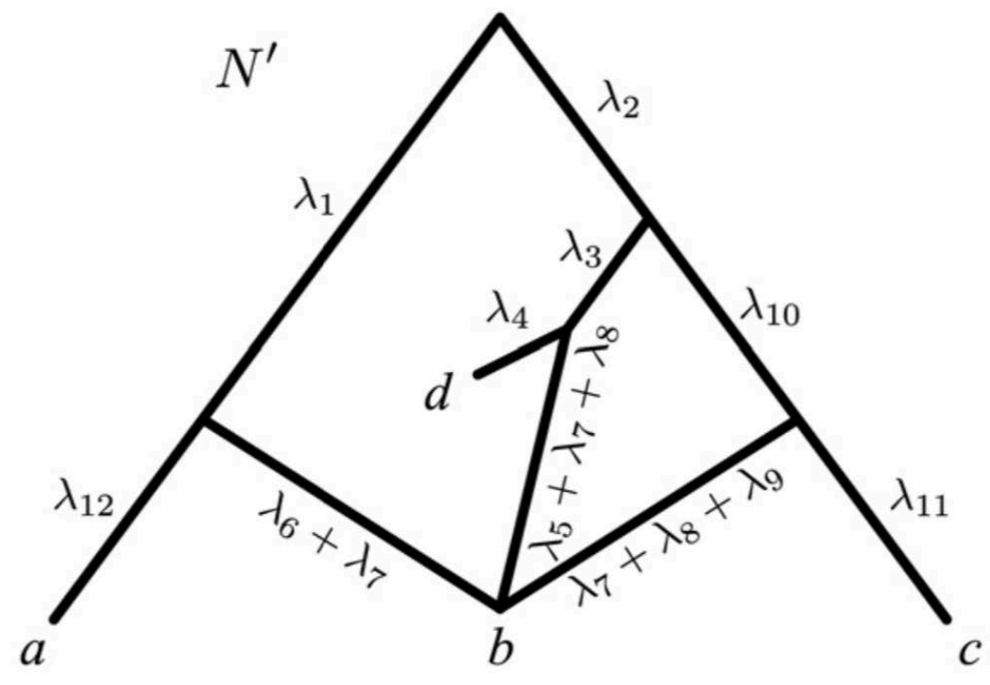
Reconstructible Phylogenetic Networks: Do Not Distinguish the Indistinguishable

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



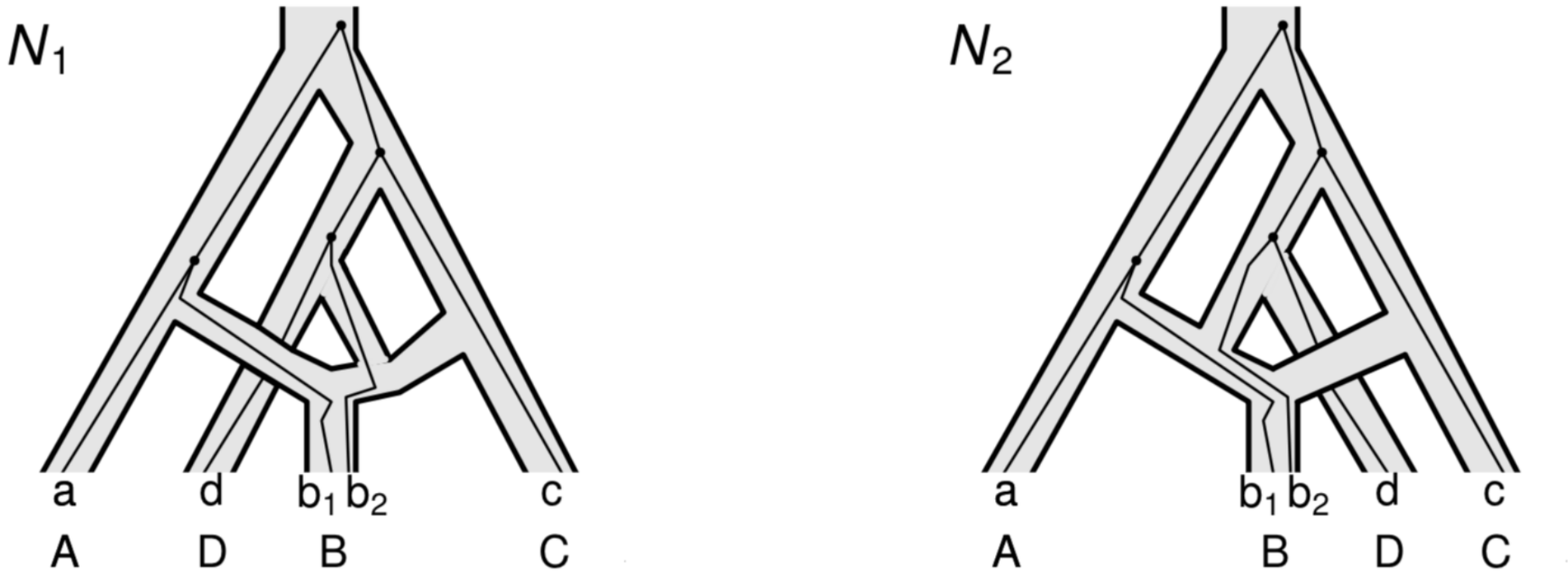
Undistinguishable with the "displayed trees" criterion



Solution: Canonical network ("unzipped")

Displayed Trees Do Not Determine Distinguishability Under the Network Multispecies Coalescent

Sha Zhu[1], James H. Degnan[2]



Distinguishable under the MSC

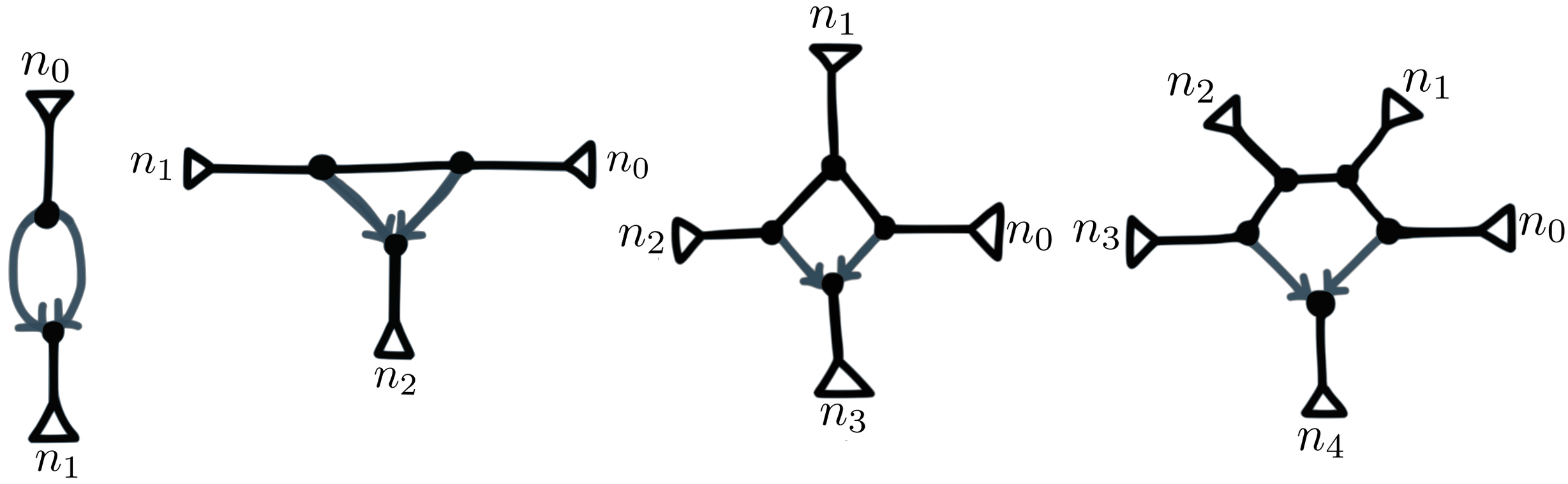
PLOS | GENETICS

RESEARCH ARTICLE

Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting

Claudia Solís-Lemus[1]*, Cécile Ané[1,2]

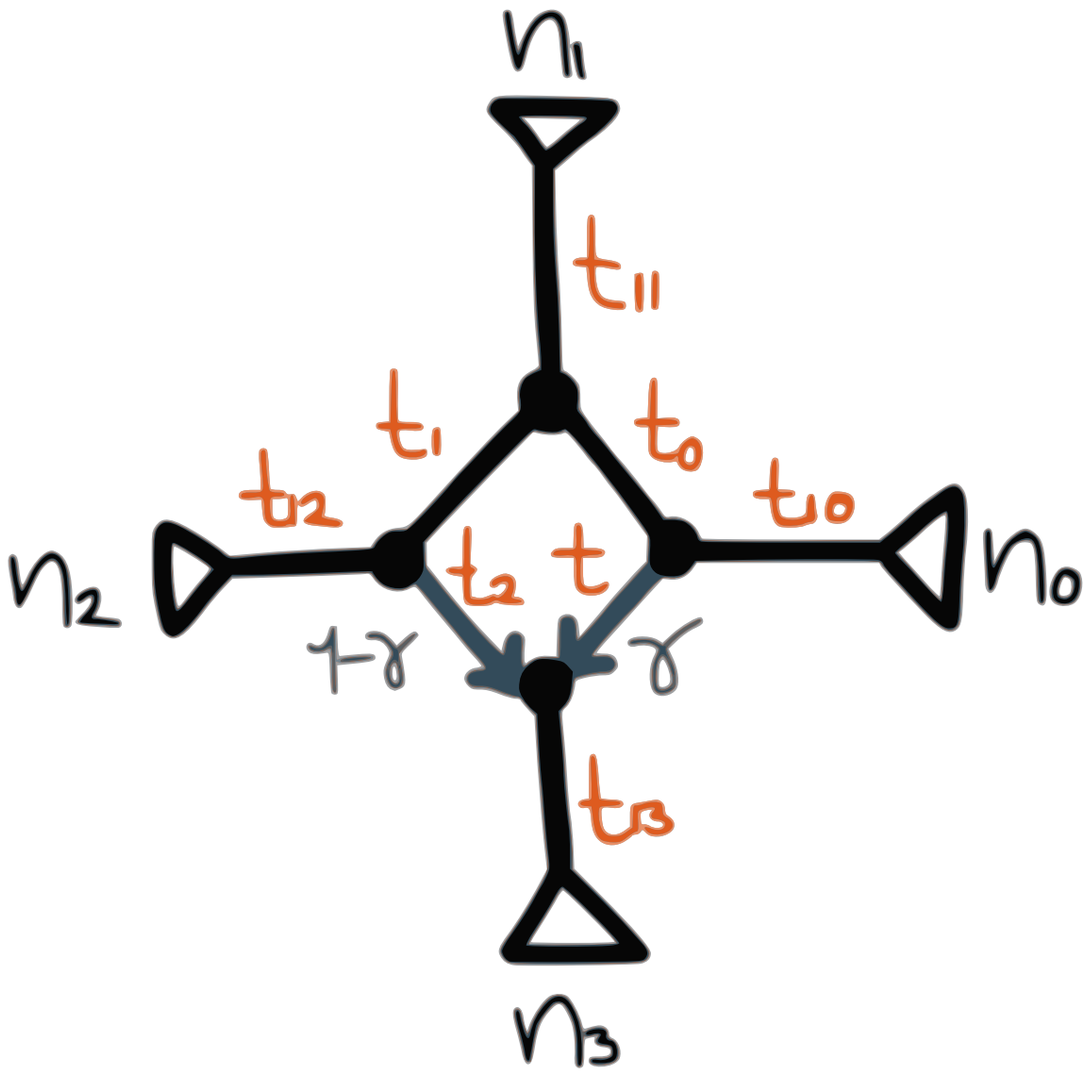
Can we detect the presence of hybridization in level-1 networks?

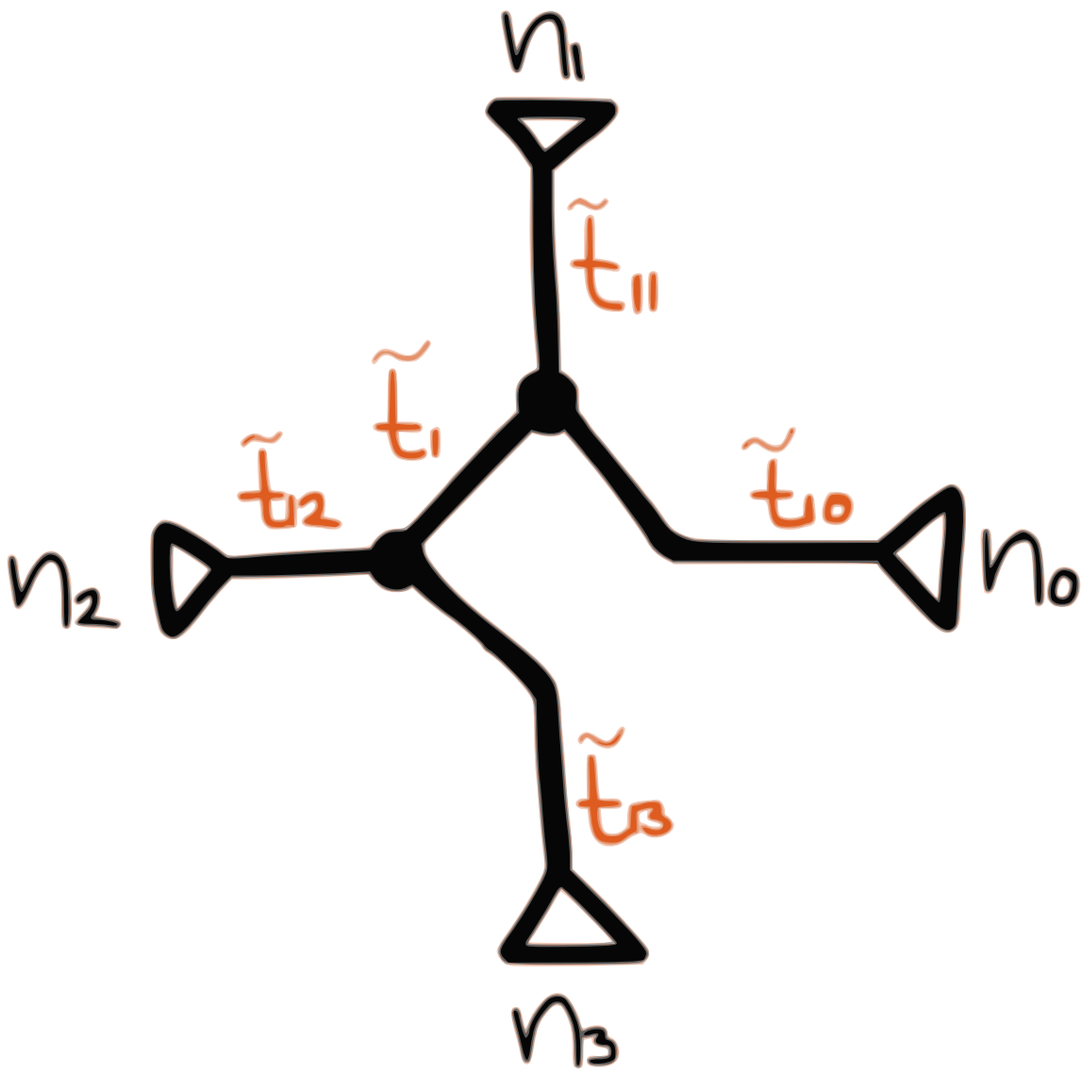


No **Yes** $(n_i, n_j \geq 2)$ **Yes** $(n_i \geq 2)$ **Yes**

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

Idea of proof of identifiability: hybridization





System of equations

$\{CF_{network}\}$

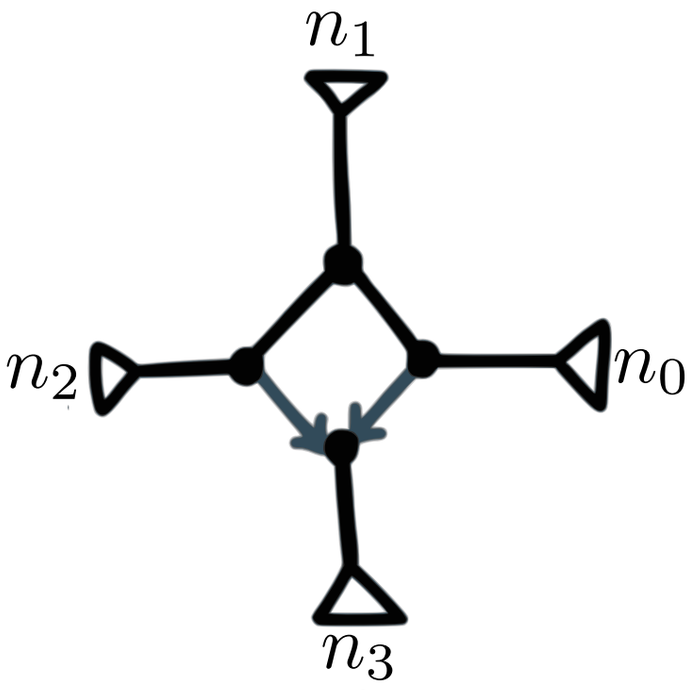
(Solís-Lemus & Ané, 2016;
Solís-Lemus et al, 2020)

System of equations

$\{CF_{tree}\}$

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

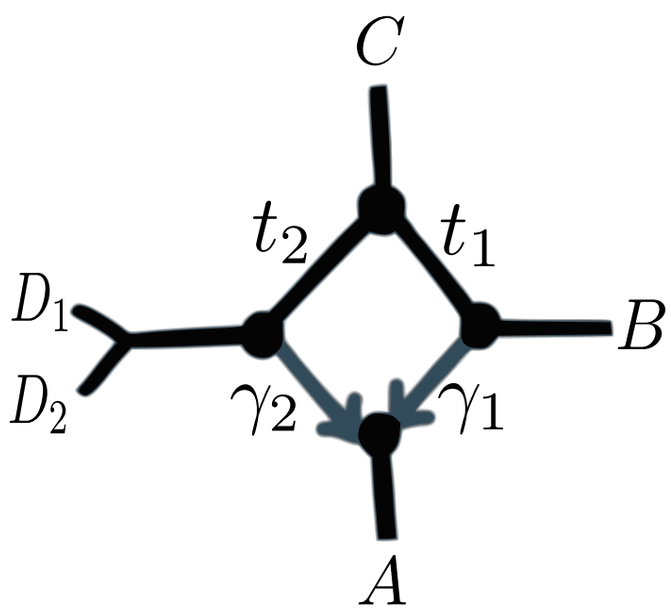
In theory

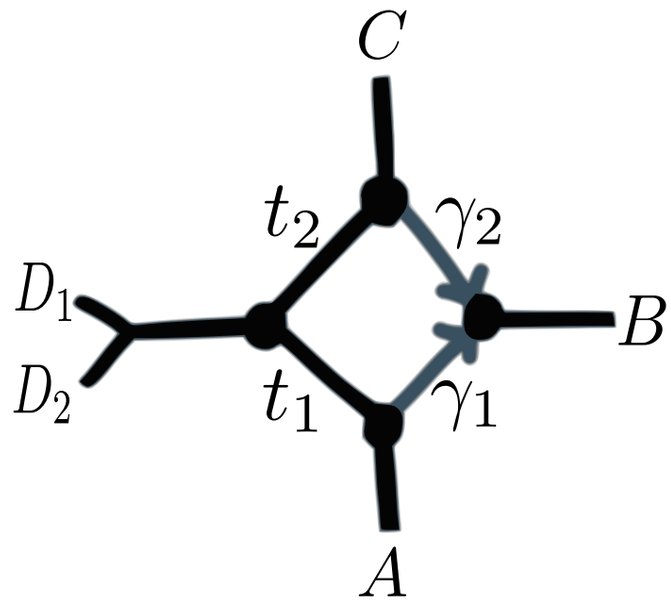


Yes

$(n_i \geq 2)$

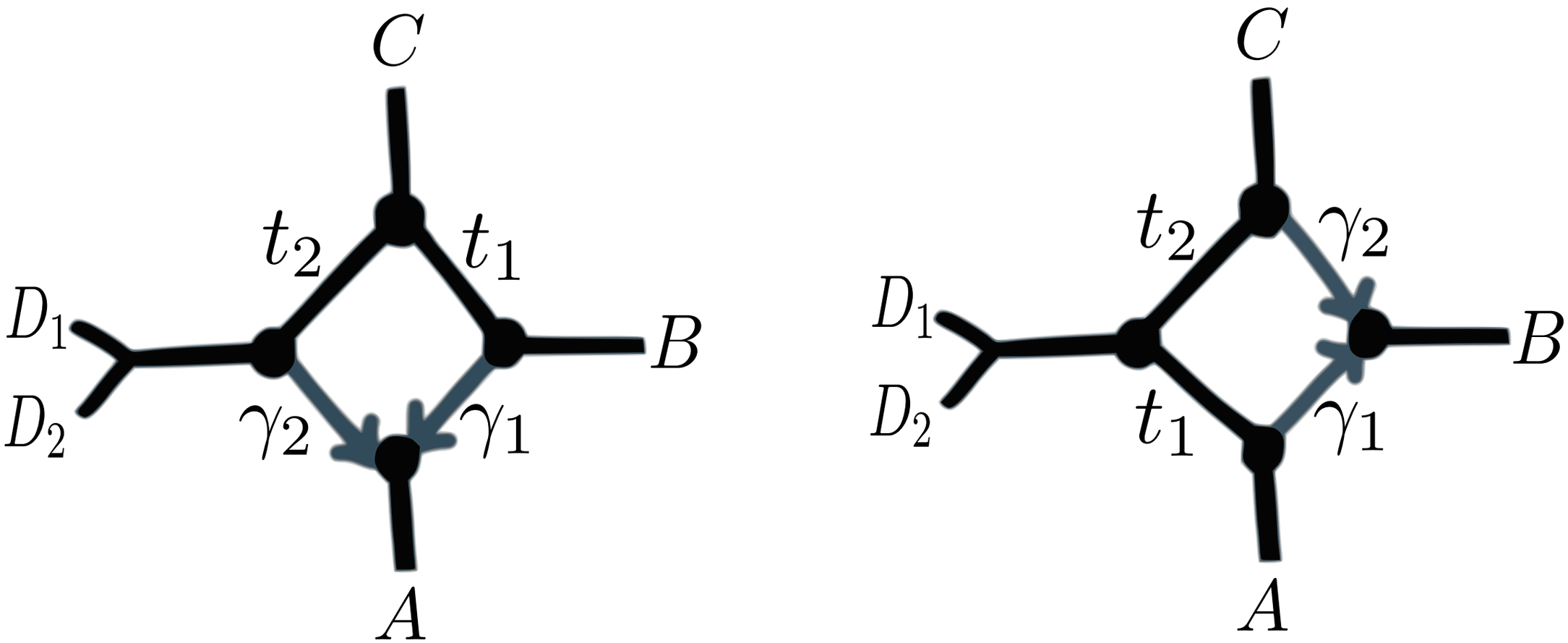
In practice





Sometimes

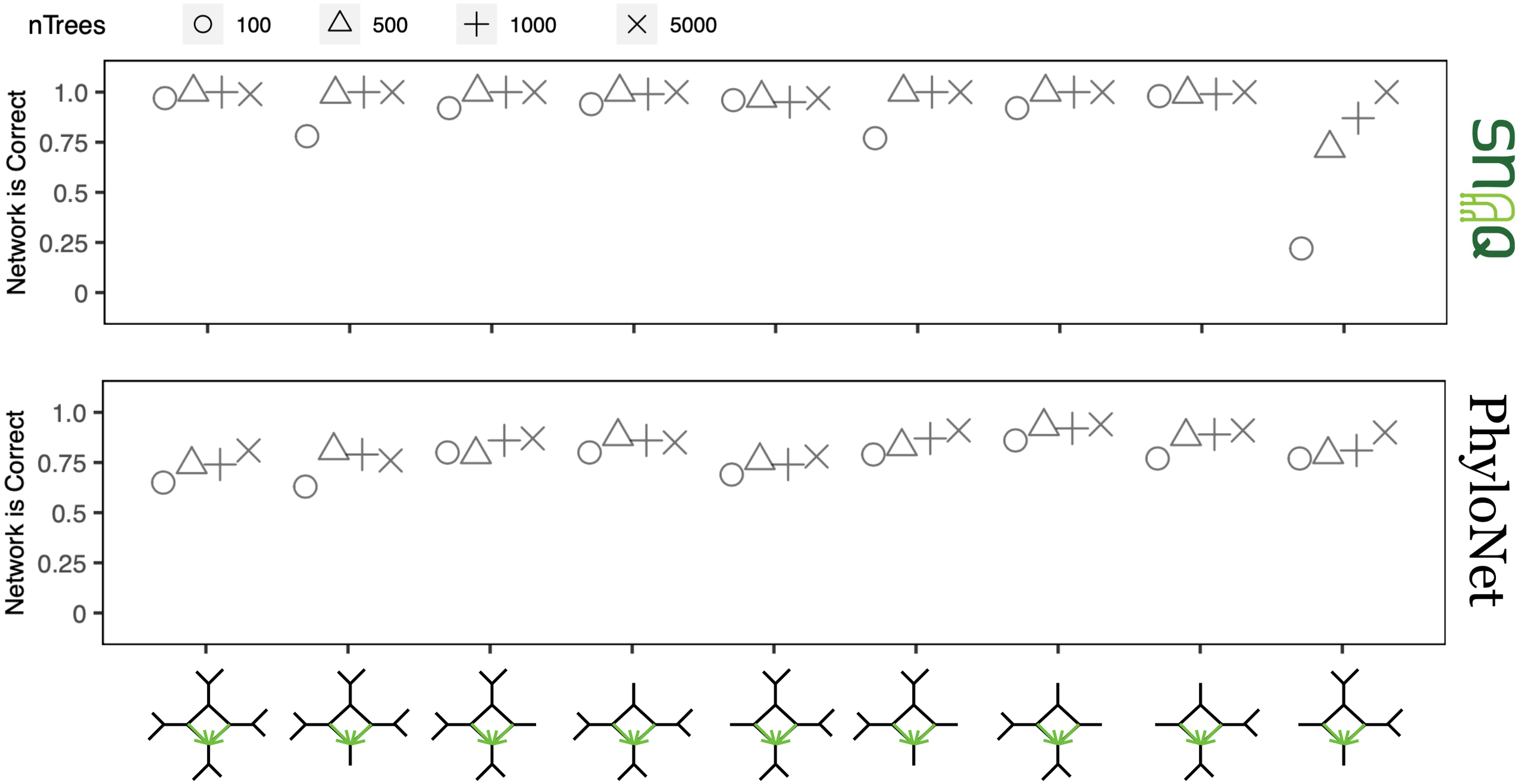
In practice: flat pseudolikelihood



(S.-L., Ané, 2016, PLoS Genetics)

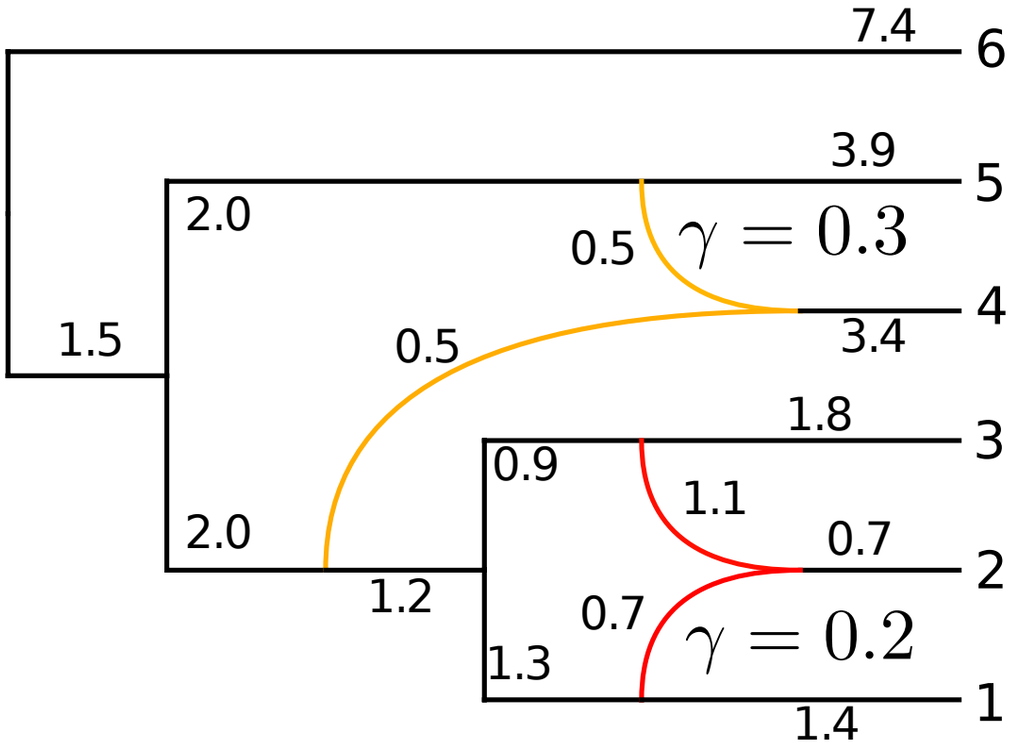
SNaQ Diamond identifiability

George Tiley

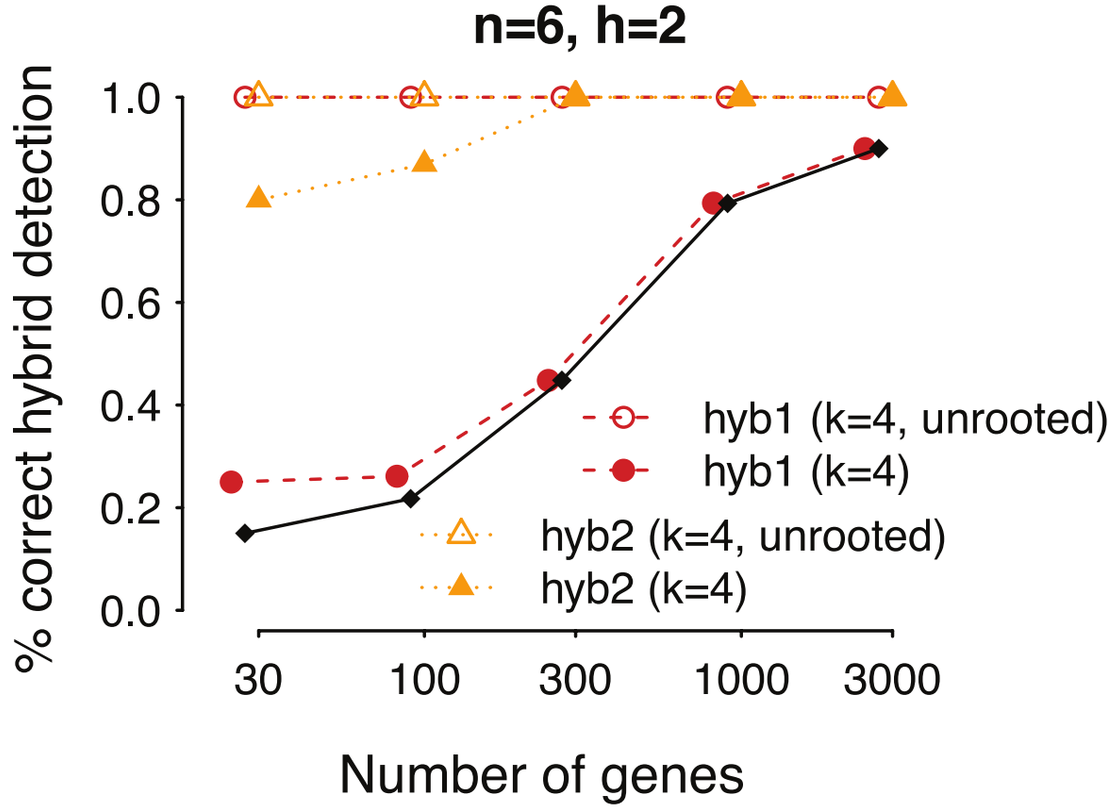


Identifiability matters: SNaQ performance

Good diamond

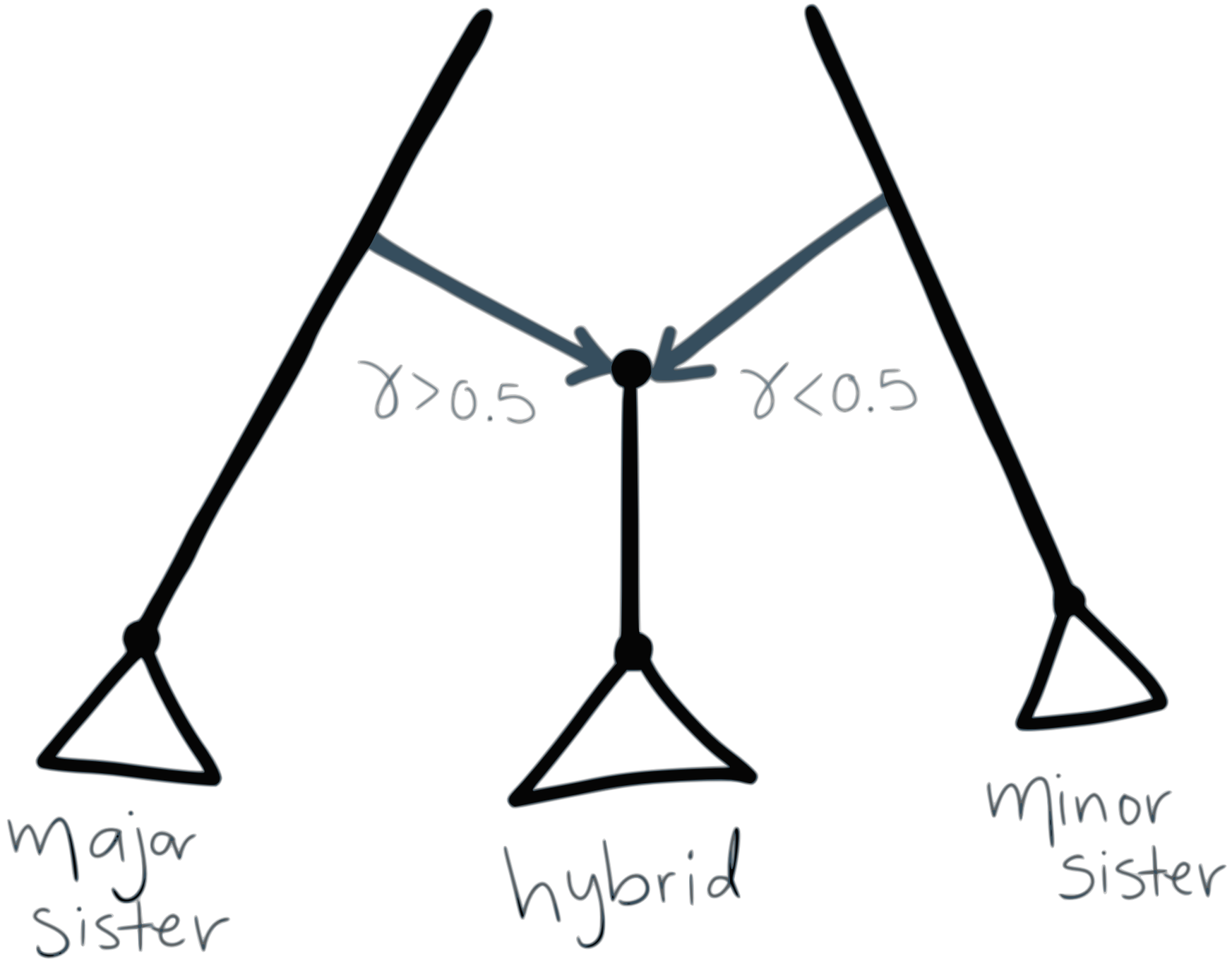


Bad diamond



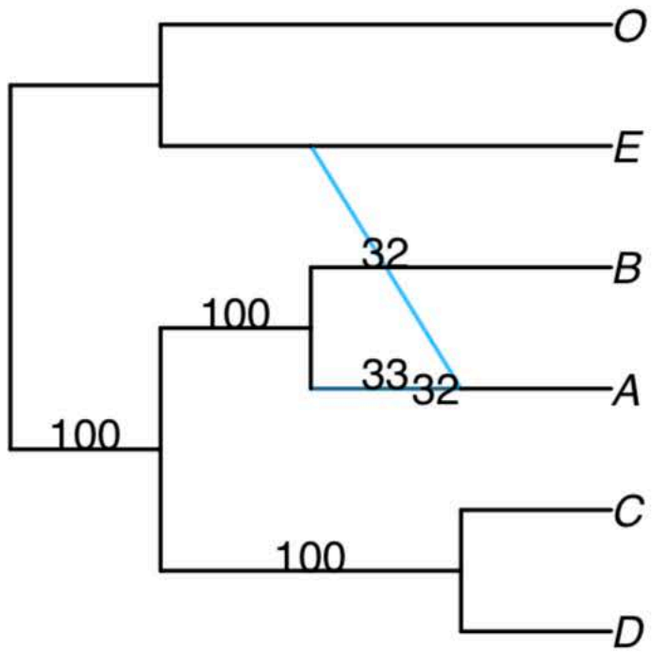
(S.-L., Ané, 2016, PLoS Genetics)

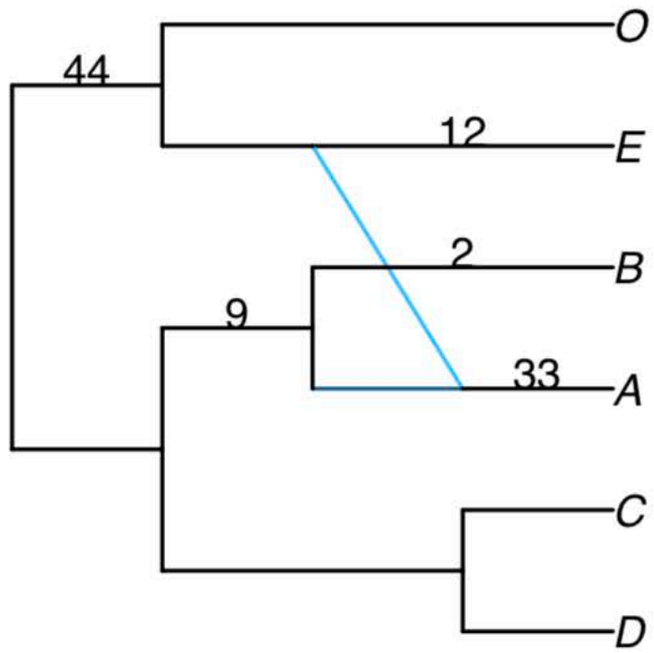
Network Comparison



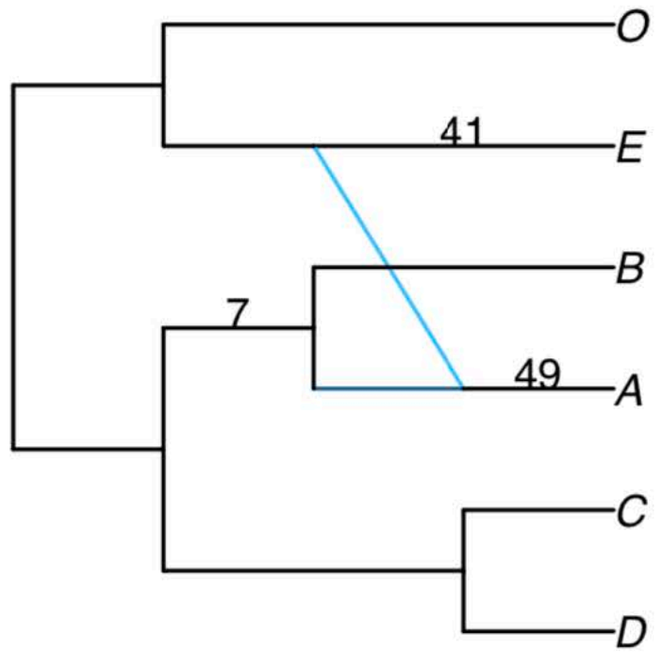
(S.-L. et al, 2017, MBE)

Network summary





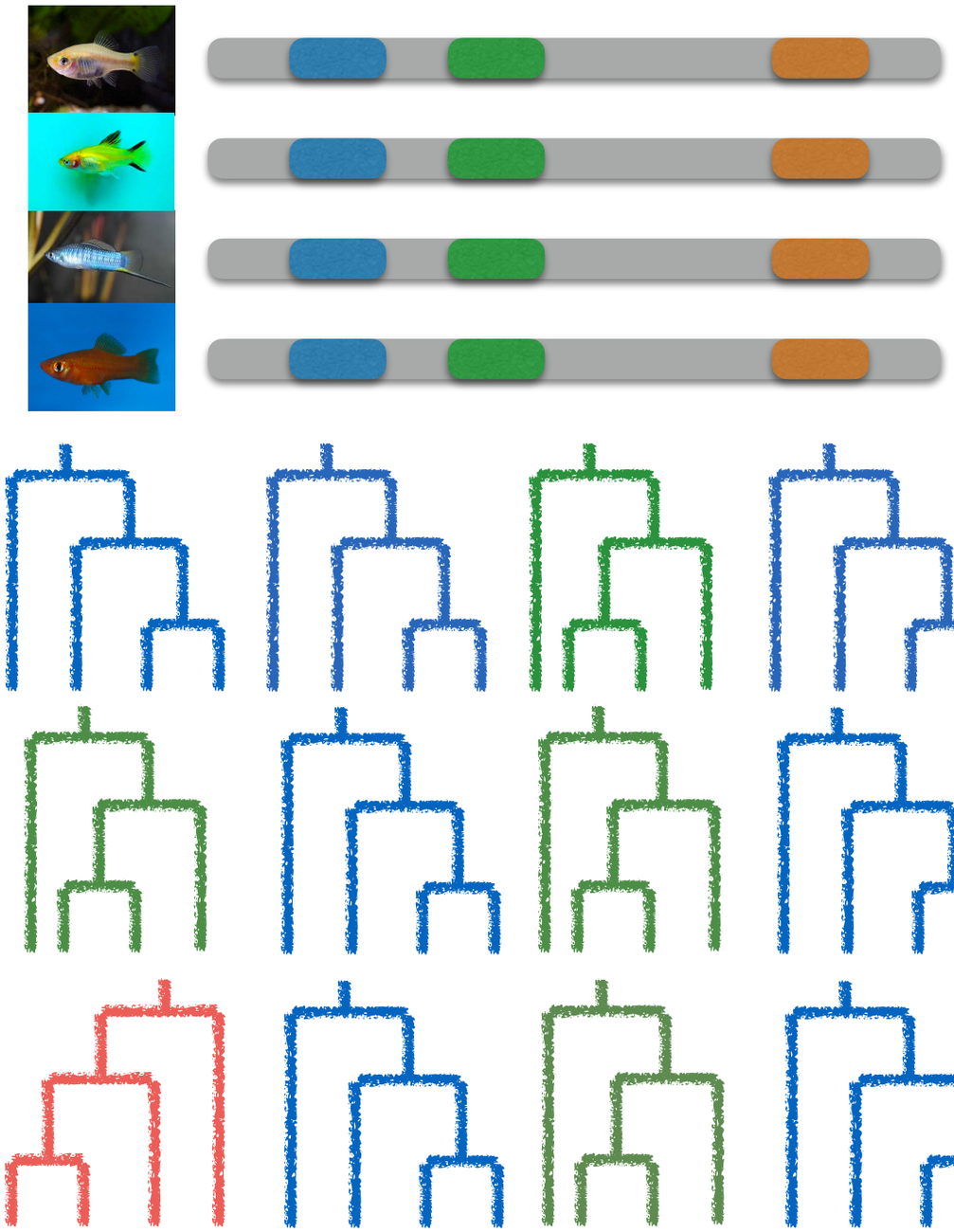
Hybrid clades



Minor sister clades

When?

Phylogenetic network



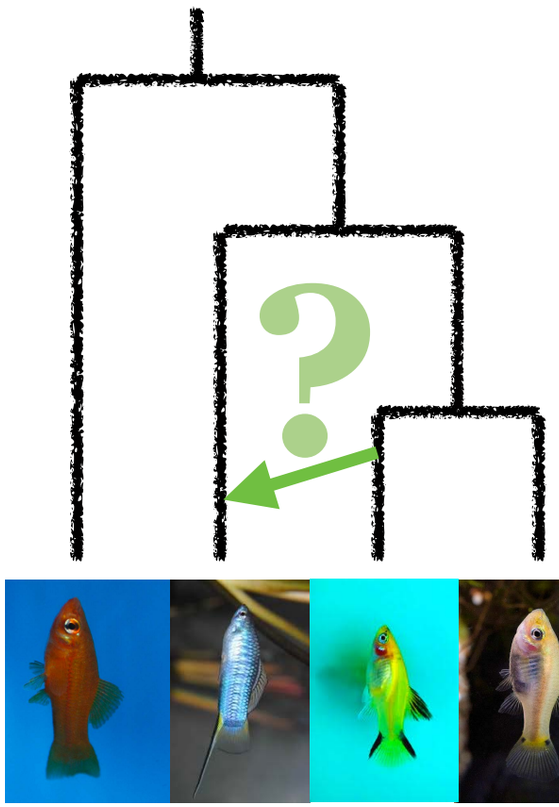
Data

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,…

SNaQ

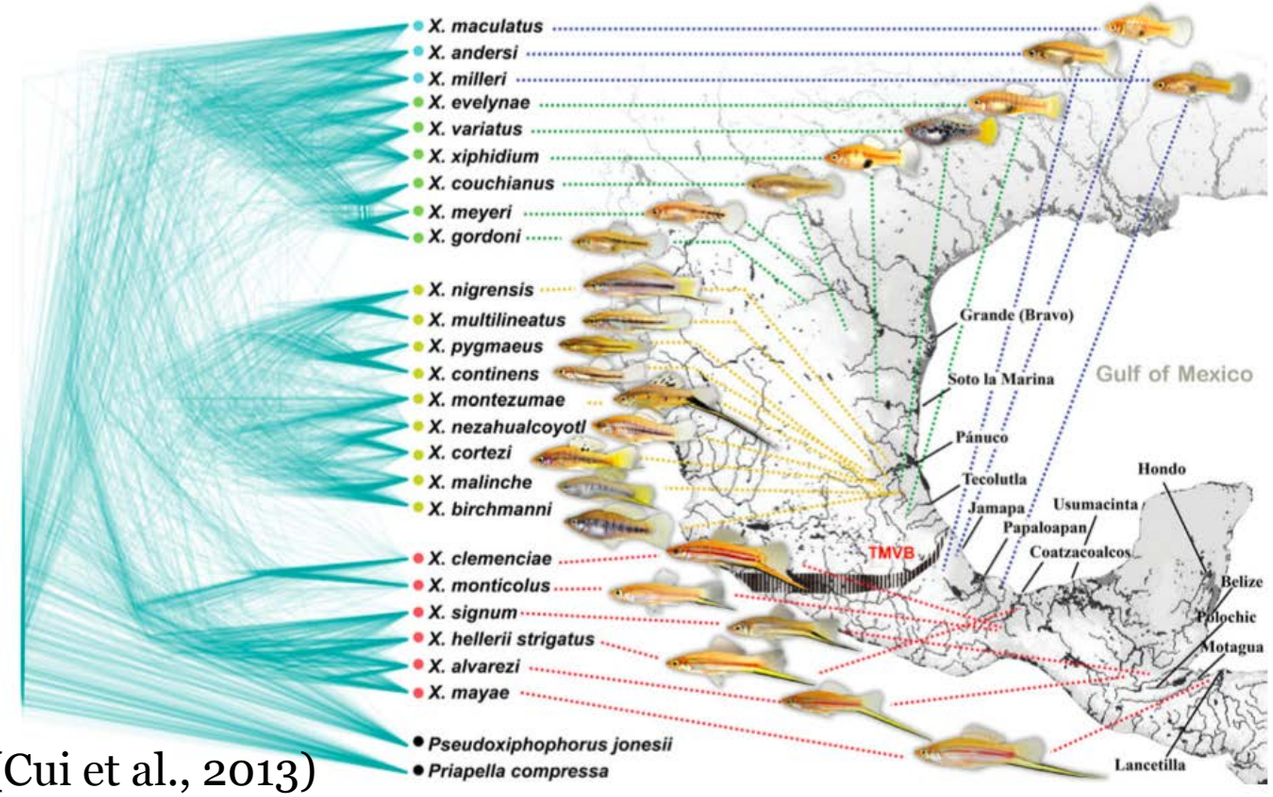
Practical advice

- 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,…

When to stop?
(Cai and Ané, 2020)

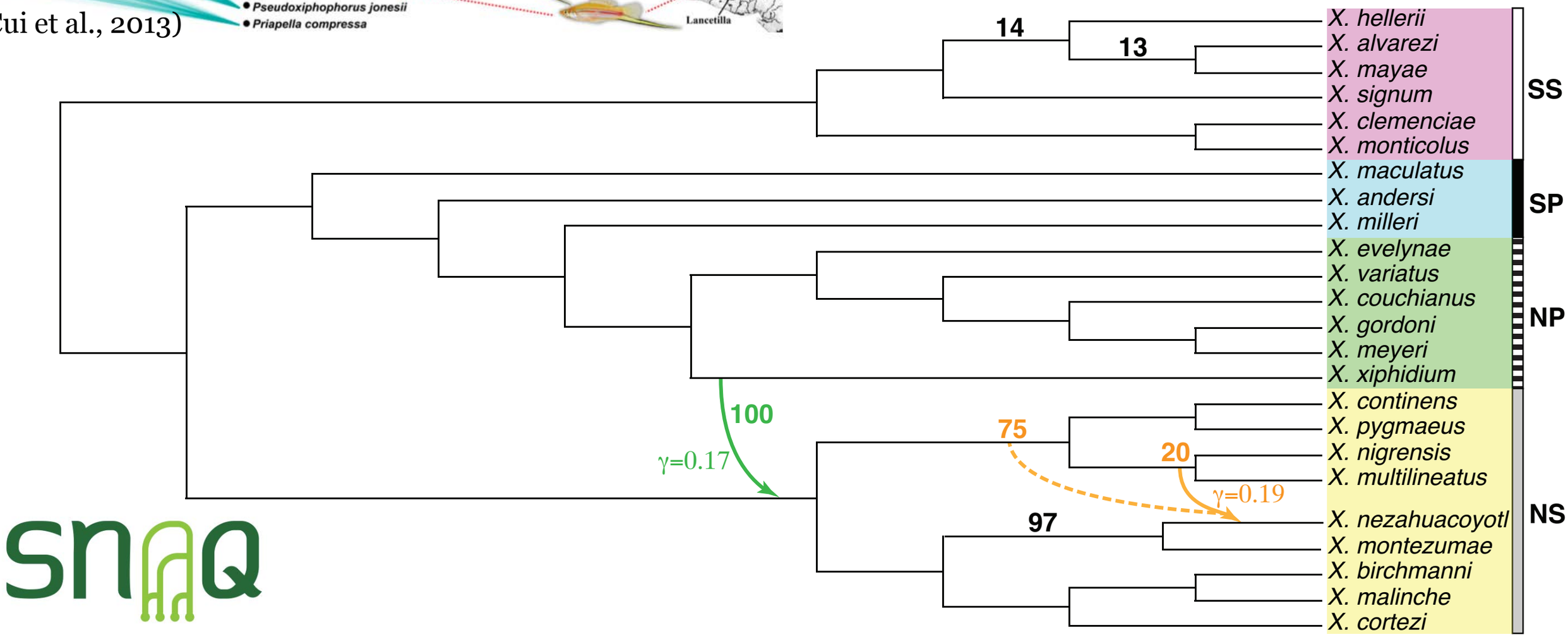
SNAQ

1183 genes, 24 swordtails and platyfish



(Cui et al., 2013)

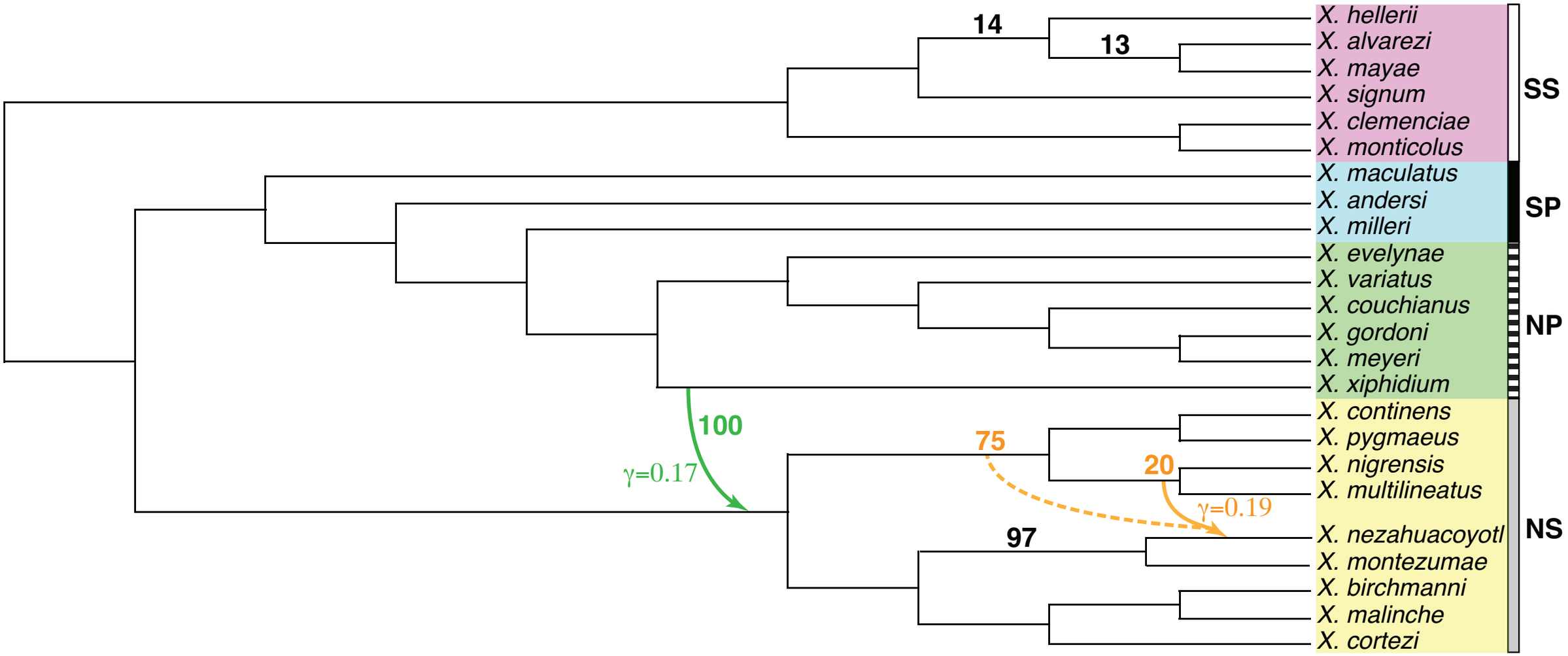
Xiphophorus fish data



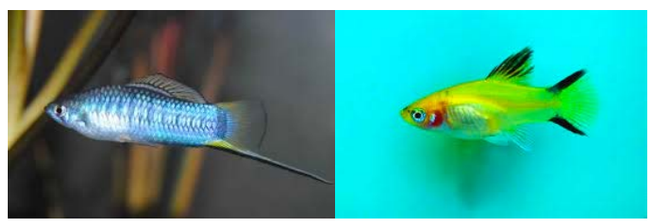
(Solís-Lemus, Ané, 2016, PLoS Genetics)

Part II

I have the network, now what?



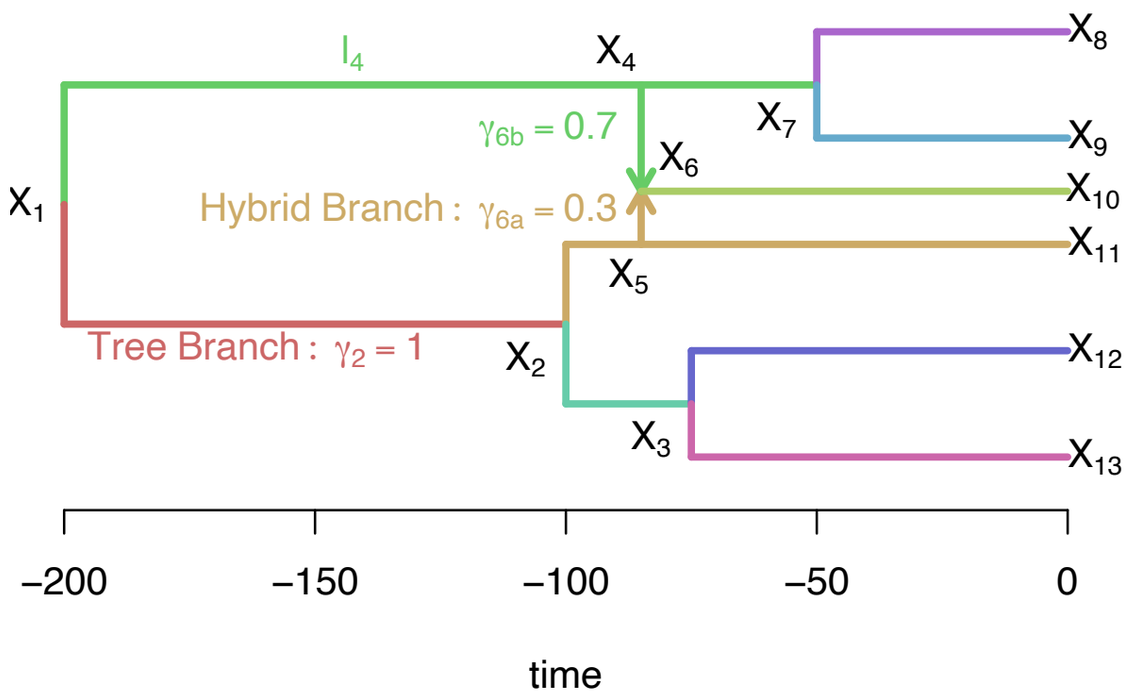
- Sword index
- Female preference

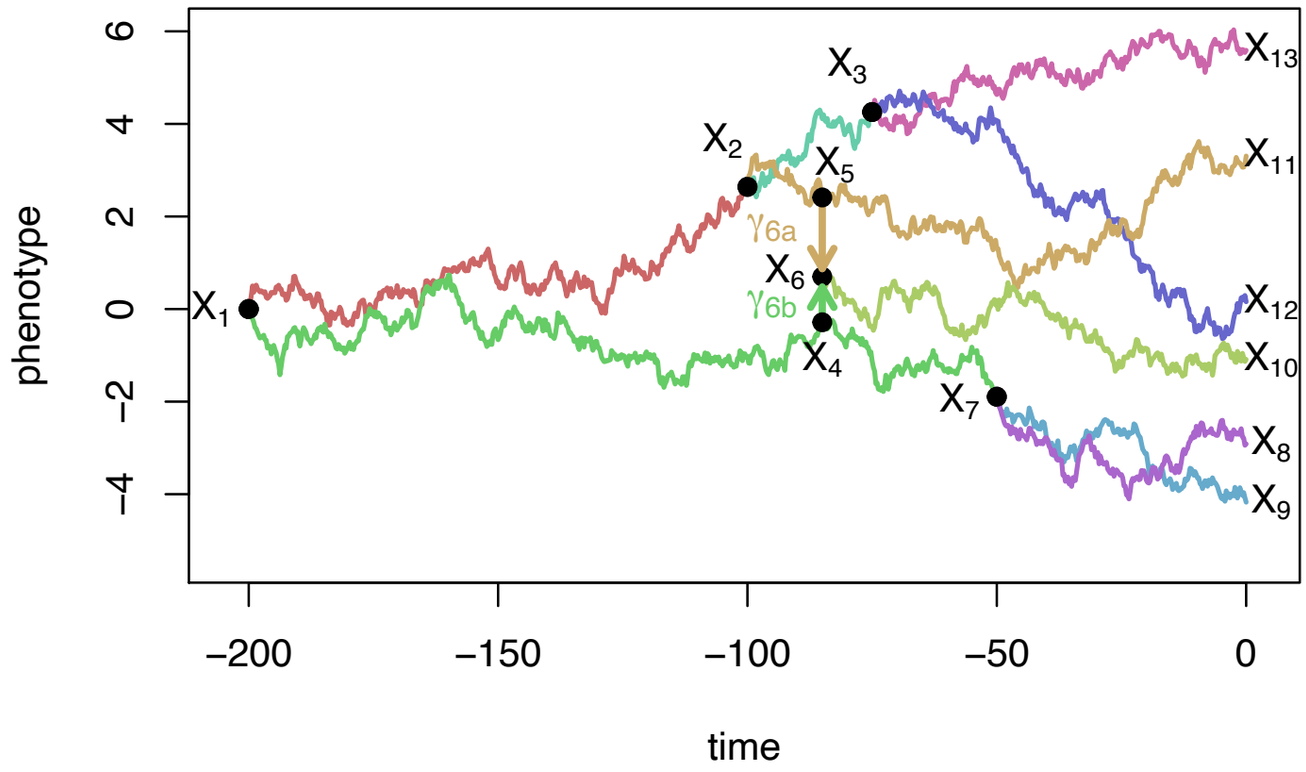


(Cui et al., 2013)

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

Trait models of evolution in networks





Brownian Motion + weighted average in hybrid

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

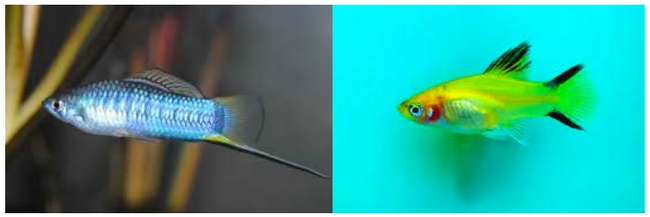
(Bastide et al, 2018, Syst Bio)

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

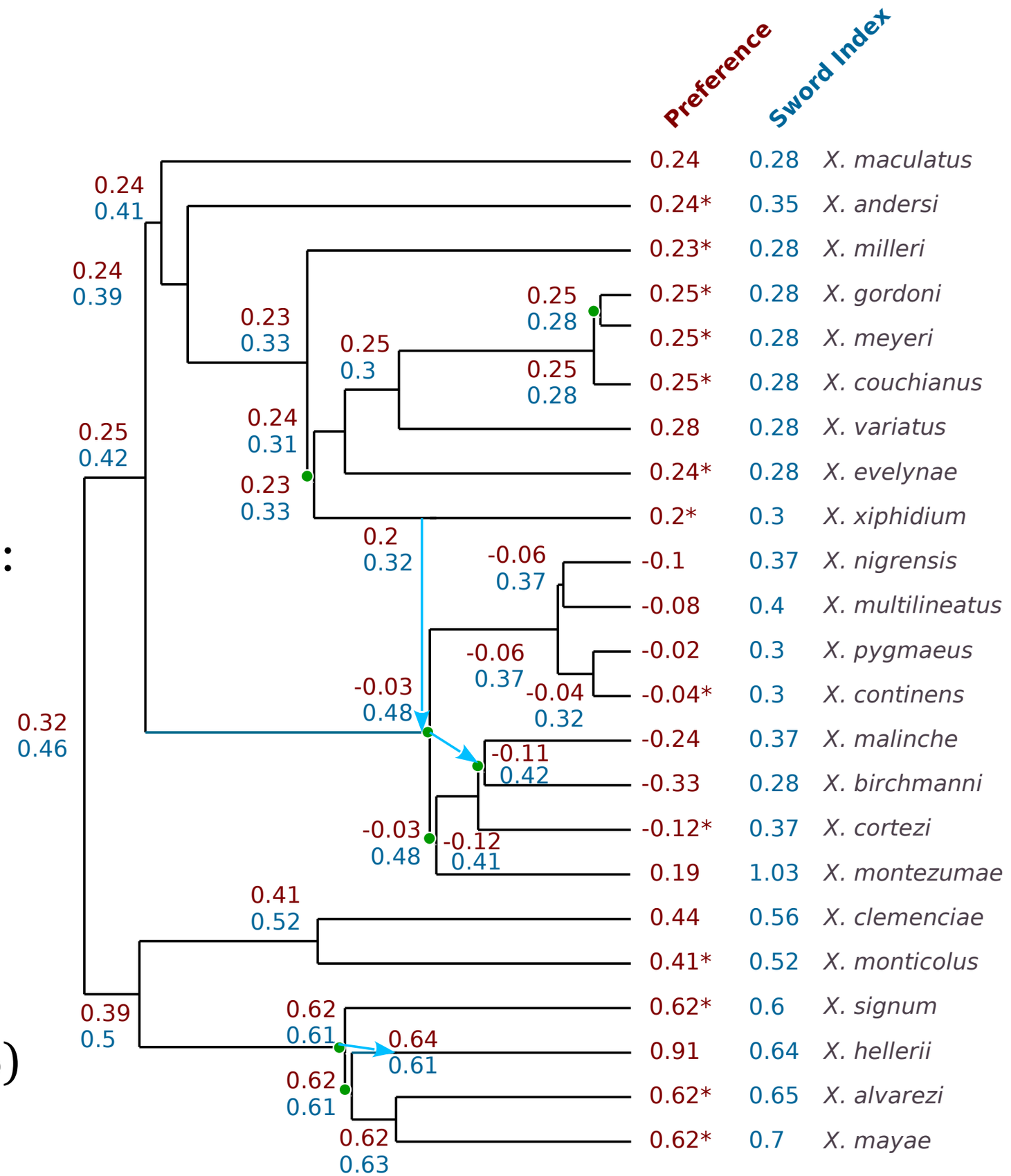
- Phylogenetic signal
- Ancestral reconstruction
- Phylogenetic regression
- Phylogenetic ANOVA

`www.github.com/CRSL4/PhyloNetworks`

- Sword index
- Female preference

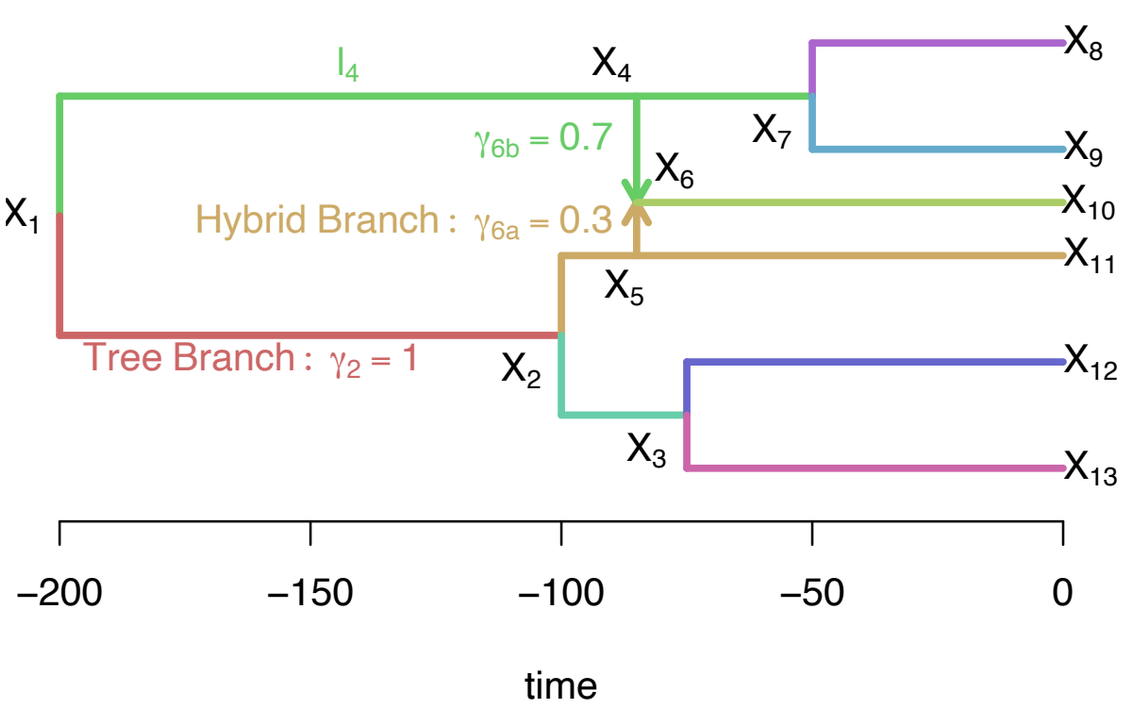


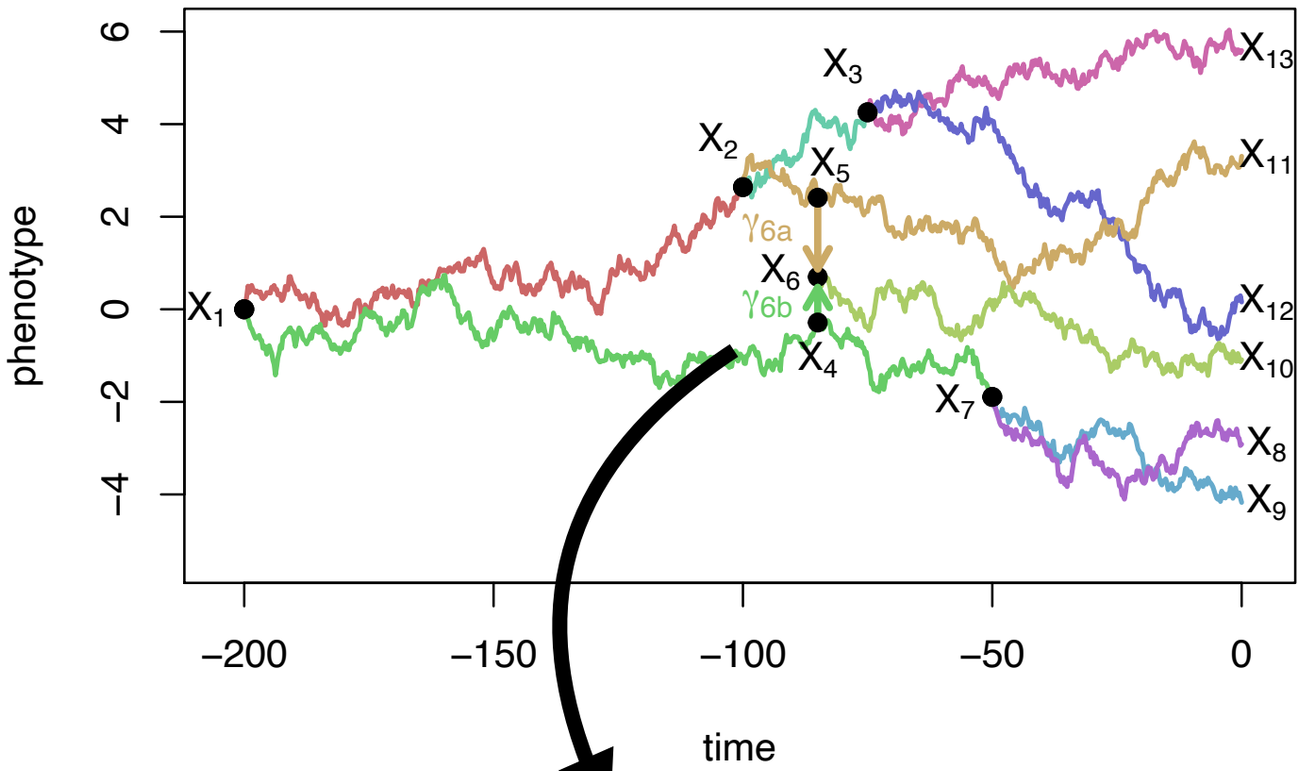
- **Ancestral reconstruction:** common ancestor likely had sword
- **Phylogenetic regression:** positive association between sword index and female preference
  but not significant (p = 0.106)



Test for transgressive evolution

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





$\Delta_h = 0$ No transgressive evolution

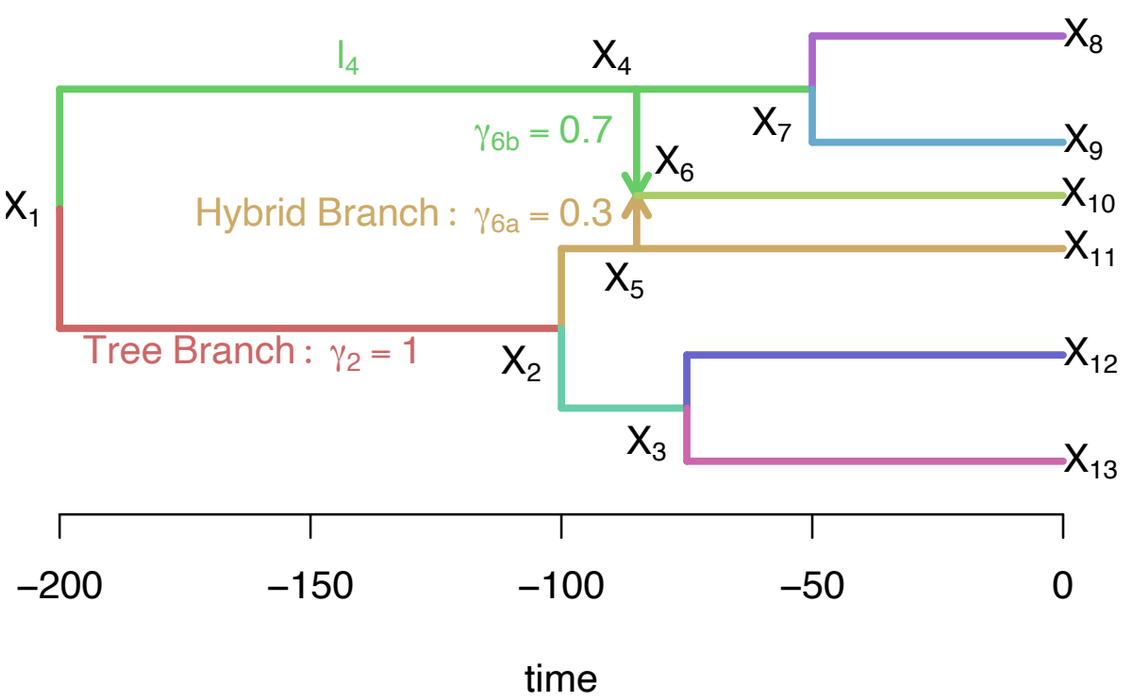
$\Delta_h = \Delta$ Single-effect transgressive evolution

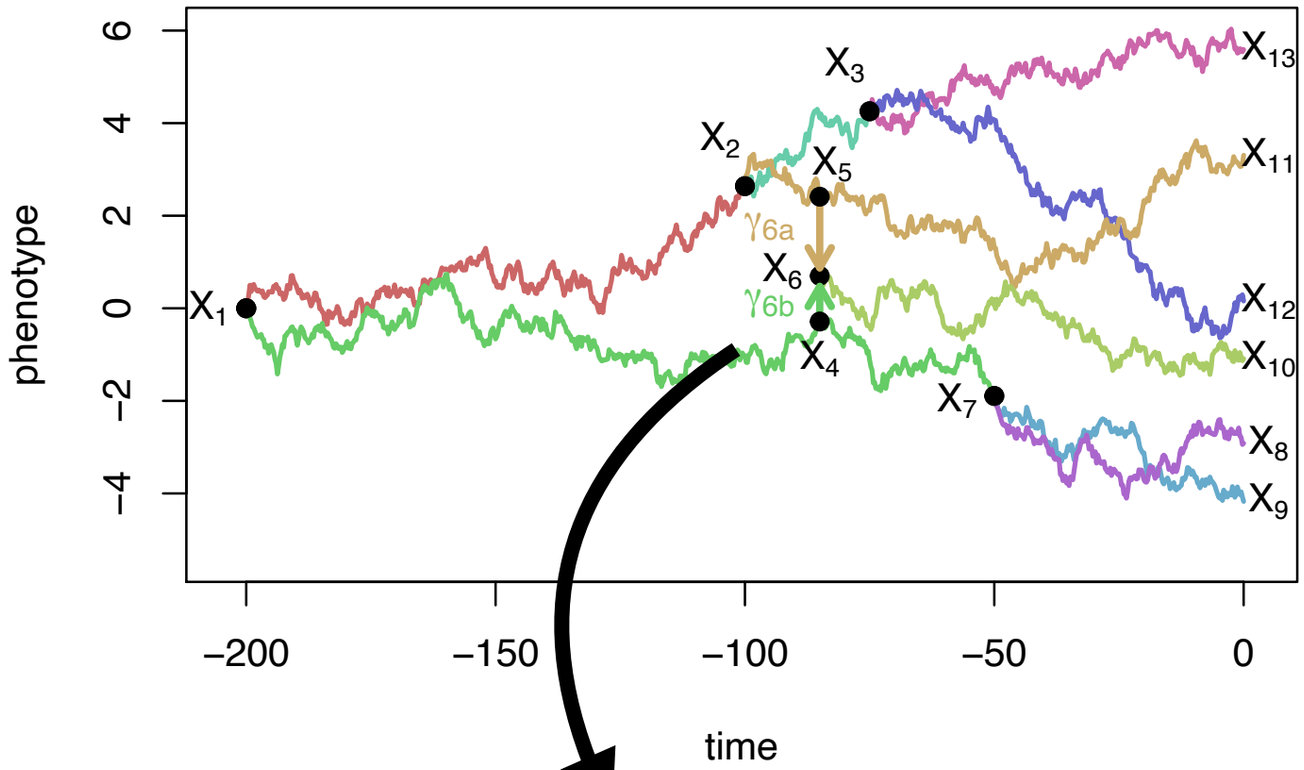
Δ_h Multi-effect transgressive evolution

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

Overview

Get your stickers!

PhyloNetworks is a Julia package with utilities to:

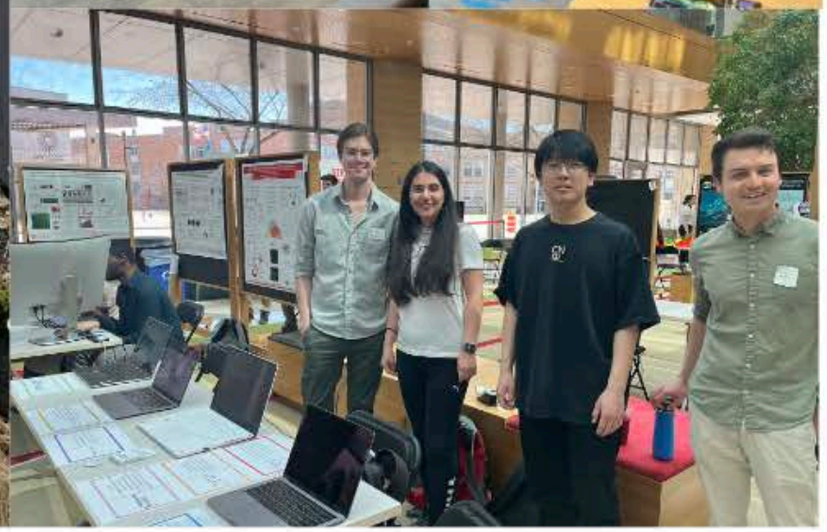
- read / write phylogenetic trees and networks, in (extended) Newick format. Networks are considered explicit: nodes represent ancestral species. They can be rooted or unrooted.
- manipulate networks: re-root, prune taxa, remove hybrid edges, extract the major tree from a network, extract displayed networks / 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

HT
CHTC

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

NSF

HOME GROWN
WISCONSIN
POTATOES

USDA

WID

Thank you!

Bayesian NetworkRegression.jl

CARlasso

snaq

phylnetworks

https://solislemuslab.github.io/ mstdn.social/@solislemuslab solislemuslab @thestatistician

Questions?