

from species trees to species networks

Cécile Ané

UW - Madison, Departments of Statistics and of Botany



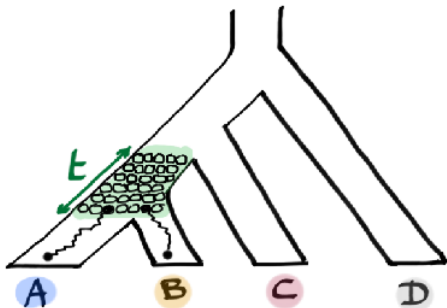
WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON

1. **impact** of introgression / hybridization?
2. is a **tree sufficient**, or do we need a network?
3. what is a **species network**?

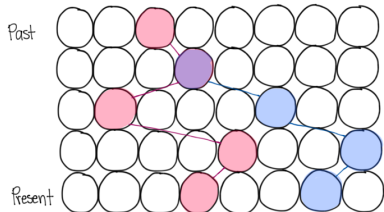
can we ignore incomplete lineage sorting?



Does **incomplete lineage sorting** impact tree reconstruction?



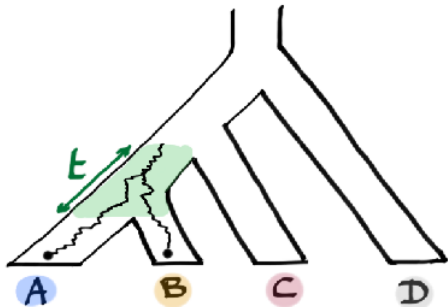
within 1 population:



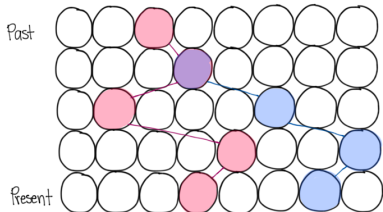
can we ignore incomplete lineage sorting?



Does **incomplete lineage sorting** impact tree reconstruction?



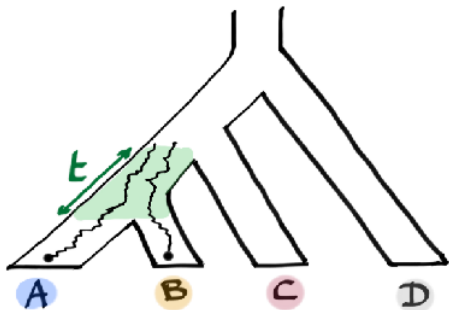
within 1 population:



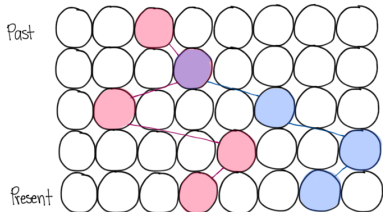
can we ignore incomplete lineage sorting?



Does **incomplete lineage sorting** impact tree reconstruction?



within 1 population:





Does **incomplete lineage sorting** impact tree reconstruction?

yes! concatenation is **not robust** to ILS.

(Kubatko & Degnan 2007)

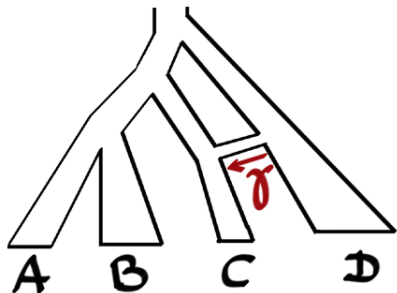
anomalous genes trees

coalescent methods: *BEAST, MP-EST, ASTRAL, SDVquartet, etc.

can we ignore introgression?



Does **gene flow / introgression** impact coalescent methods?



gene tree discordance



γ = inheritance, e.g. Neanderthals - modern humans: $\gamma \sim 2\%$

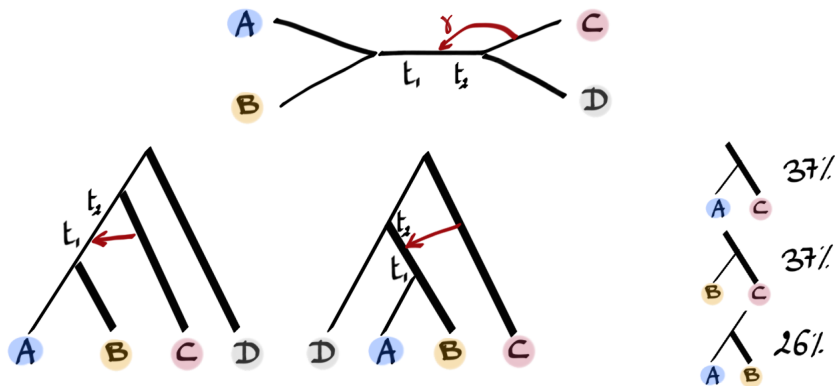


Does **gene flow** / **introgression** impact coalescent methods?

yes! some coalescent-based methods are **not robust** to gene flow.

(Solís-Lemus, Yang & Ané 2016)

anomalous unrooted gene trees: AuGT



under network model

(Solís-Lemus, Yang & Ané 2016)

under continuous gene flow between sister species (Long & Kubatko 2018)



frequency of
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

$t_1 = t_2 = 0.01$

ILS only: **no AuGT** on 4 taxa

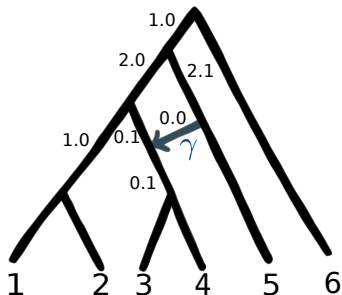
(Degnan 2013)

ILS + gene flow: **AuGT** on 4 taxa

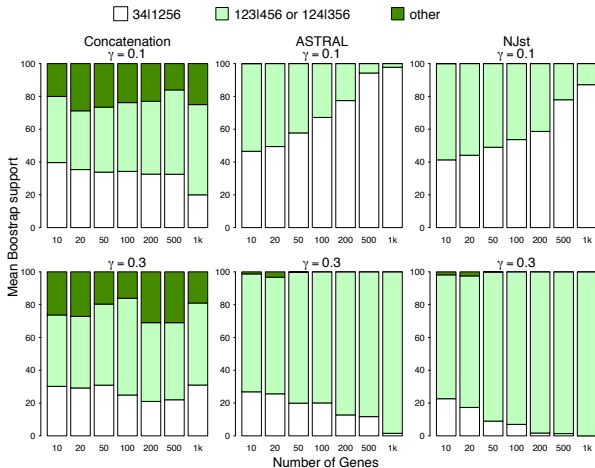
(Solís-Lemus, Yang & Ané 2016)

rooted gene trees, 3 taxa: same story

inconsistent methods: concatenation, ASTRAL, NJst



□ 3 & 4 sister
■ 3 & 4 not sister



>

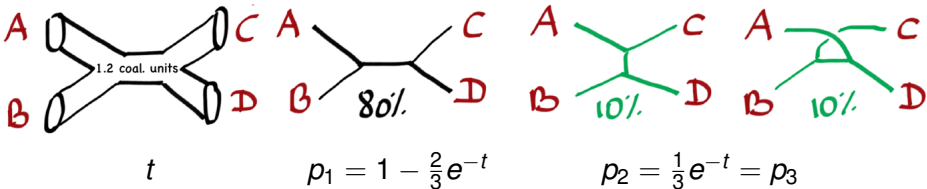
1. **impact** of introgression / hybridization?
2. is a **tree sufficient**, or do we need a network?
3. what is a **species network**?

does a tree fit the data well? or network needed?

TICR: goodness-of-fit test of ILS on a population tree

Stenz et al. (2015)

expectation from ILS: equal % genes (CF) with minor resolutions



- same idea as ABBA-BABA test on SNPs
- combine all 4-taxon sets in a single test

Green et al. (2010)

Durand et al. (2011)

data: quartet concordance factors

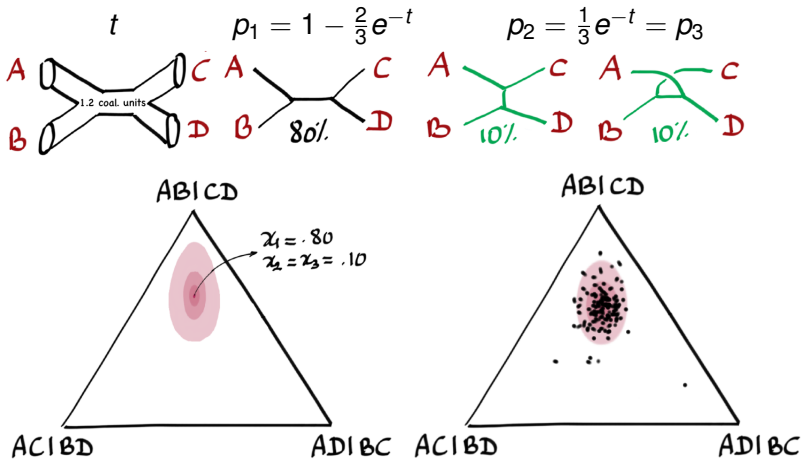


4-taxon subset				proportion of genes with		
1	2	3	4	12 34	13 24	14 23
				x_1	x_2	x_3
A.gre	A.dig	A.gran	A.za	0.38	0.30	0.32
A.gre	A.dig	A.gran	A.mad	0.42	0.28	0.30
	⋮			⋮		
A.gran	A.za	A.per	A.mad	0.25	0.35	0.40

Stenz et al. (2015): 3,595 loci; 30 taxa so 27,405 four-taxon sets

x_1, x_2, x_3 : % genes for 3 quartet trees, one 4-taxon set

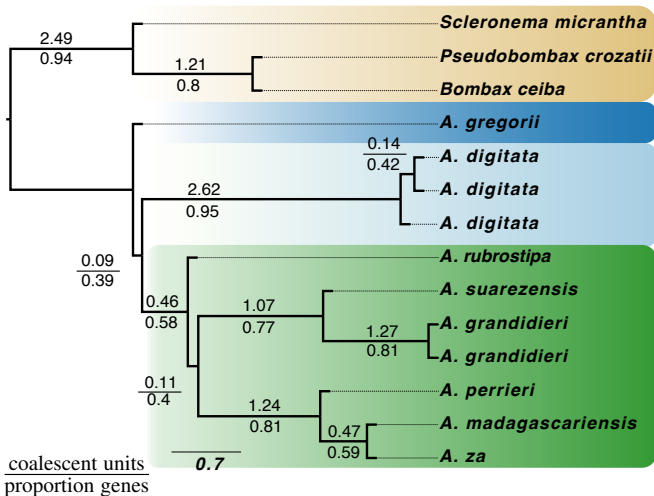
~ Dirichlet, precision α , centered at p_1, p_2, p_3 from the tree:



- p-value for each 4-taxon set
- overall test: based on proportion of outlier 4-taxon sets

baobabs (*Adansonia*): tree with ILS rejected ($p=0.04$)

14 individuals, 282 orthologous genes (targeted sequence capture)



A. rubrostipa



A. grandidieri



1. **impact** of introgression / hybridization?
2. is a **tree sufficient**, or do we need a network?
3. what is a **species network**?



Network coalescent:

- branch lengths: coalescent units for **ILS**
- network topology: extra edges for **gene flow**, hybridization or HGT
- inheritance γ , $1 - \gamma$ on hybridization edges

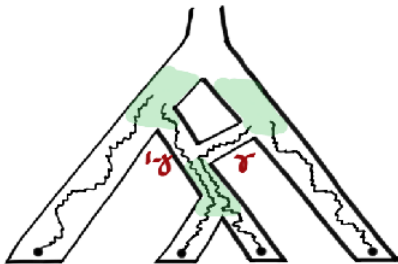
Meng & Kubatko (2009), Yu Degnan & Nakhleh (2012)



Network coalescent:

- branch lengths: coalescent units for **ILS**
- network topology: extra edges for **gene flow**, hybridization or HGT
- inheritance γ , $1 - \gamma$ on hybridization edges

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



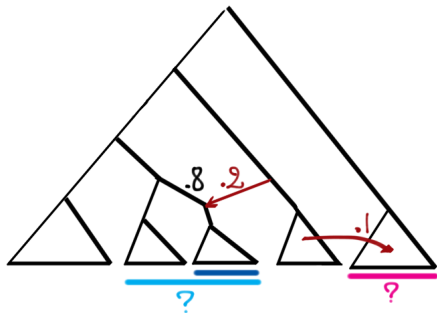
Network coalescent:

- branch lengths: coalescent units for **ILS**
- network topology: extra edges for **gene flow**, hybridization or HGT
- inheritance γ , $1 - \gamma$ on hybridization edges

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

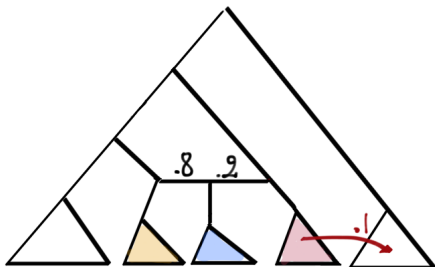
- blurred "sister" relationship, half-sibs
- clade concept?

classification more difficult

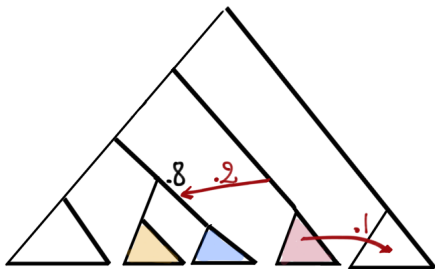


- "major" tree concept: drop each minor hybrid edge ($\gamma < 0.5$)
meaning of species tree?

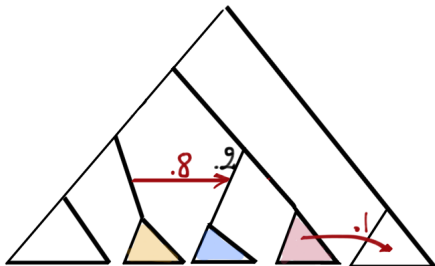
- visual artifacts: can mislead interpretation



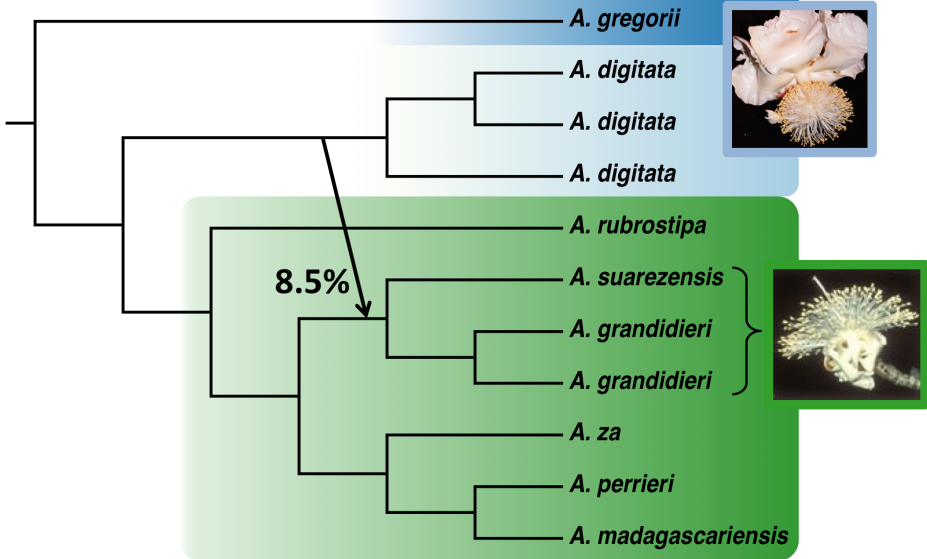
- visual artifacts: can mislead interpretation



- visual artifacts: can mislead interpretation



baobabs: 1 reticulation event



Fast algorithms and heuristics for phylogenomics under ILS and hybridization

Yun Yu*, Nikola Ristic*, Luay Nakhleh*

(2013, BMC Bioinformatics)

Maximum likelihood inference of reticulate evolutionary histories

Yun Yu^{a,1}, Jianrong Dong^a, Kevin J. Liu^{a,b}, and Luay Nakhleh^{a,b,1}

(2014, PNAS)

Inferring Phylogenetic Networks with Maximum Pseudolikelihood under Incomplete Lineage Sorting

Claudia Solís-Lemus^{1*}, Cécile Ané^{1,2}

(2016, PLOS Genetics)

identifiability: what can we (not) learn from data?

Bayesian Inference of Reticulate Phylogenies under the Multispecies Network Coalescent

Syst. Biol. 67(3):439–457, 2018

© The Author(s) 2017. Published by Oxford University Press, on behalf of the Society for Systematic Biology

For Permissions, please email: journals.permissions@oup.com

DOI:10.1093/sysbio/syx085

Advance Access publication October 27, 2017

Dingqiao Wen^{1*}, Yun Yu¹, Luay Nakhleh^{1,2*}

(2016, PLOS Genetics)

Coestimating Reticulate Phylogenies and Gene Trees from Multilocus Sequence Data

DINGQIAO WEN¹ AND LUAY NAKHLEH^{1,2,*}

Bayesian Inference of Species Networks from Multilocus Sequence Data

Chi Zhang^{*,1,2,3} Huw A. Ogilvie^{,4,5} Alexei J. Drummond^{,5,6} and Tanja Stadler^{*,1,2} (2018, MBE)

Bayesian inference of phylogenetic networks from bi-allelic genetic markers

Jiafan Zhu¹, Dingqiao Wen¹, Yun Yu¹, Heidi M. Meudt², Luay Nakhleh^{1,3*} (2018)



Complex task!

- PhyloNet
- PhyloNetworks
- BEAST2

requirements:

- not too many species!
- multiple loci: the more the better
- or SNPs (biallelic)

PhyloNetworks.jl

[Home](#)

Manual ^

[Installation](#)[Input Data for SNaQ](#)[TICR pipeline](#)

PhyloNetworks.jl

PhyloNetworks is a [Julia](#) package for the manipulation, visualization, inference of phylogenetic networks, and their use for trait evolution.

Table of contents

[How to get help](#)[Manual Outline](#)[Library Outline](#)[Index](#)[Functions](#)

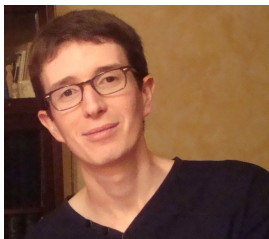
PhyloNetworks package:

- SNaQ: gene trees or quartet CFs → species network
bootSNaQ: bootstrap gene trees → bootstrap networks
bootstrap support: for tree edges, gene flow recipient, donor
- trait evolution on networks: continuous response
- plot, root, re-root networks
extract the major tree, extract all displayed trees

Acknowledgements



Claudia Solís-Lemus



Paul Bastide

Mengyao Yang
John Malloy
John Spaw
Doug Bates
Sarah Friedrich



baobabs:

David Baum

Nisa Karimi

Jonathan Wendel, Joe Gallagher,
Corrinne Grover, Noah Stenz



DEB 0949121

DEB 1354793

