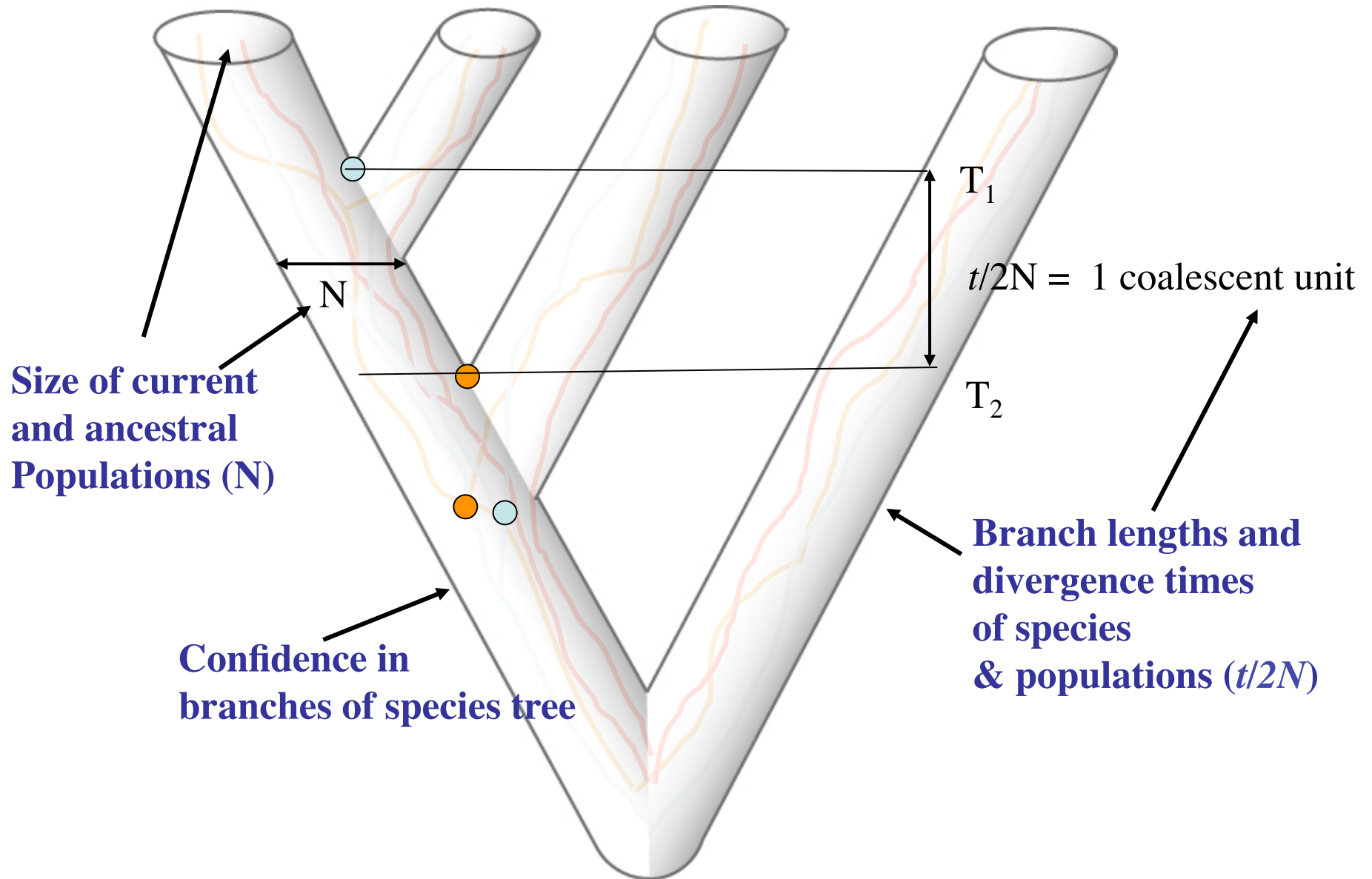
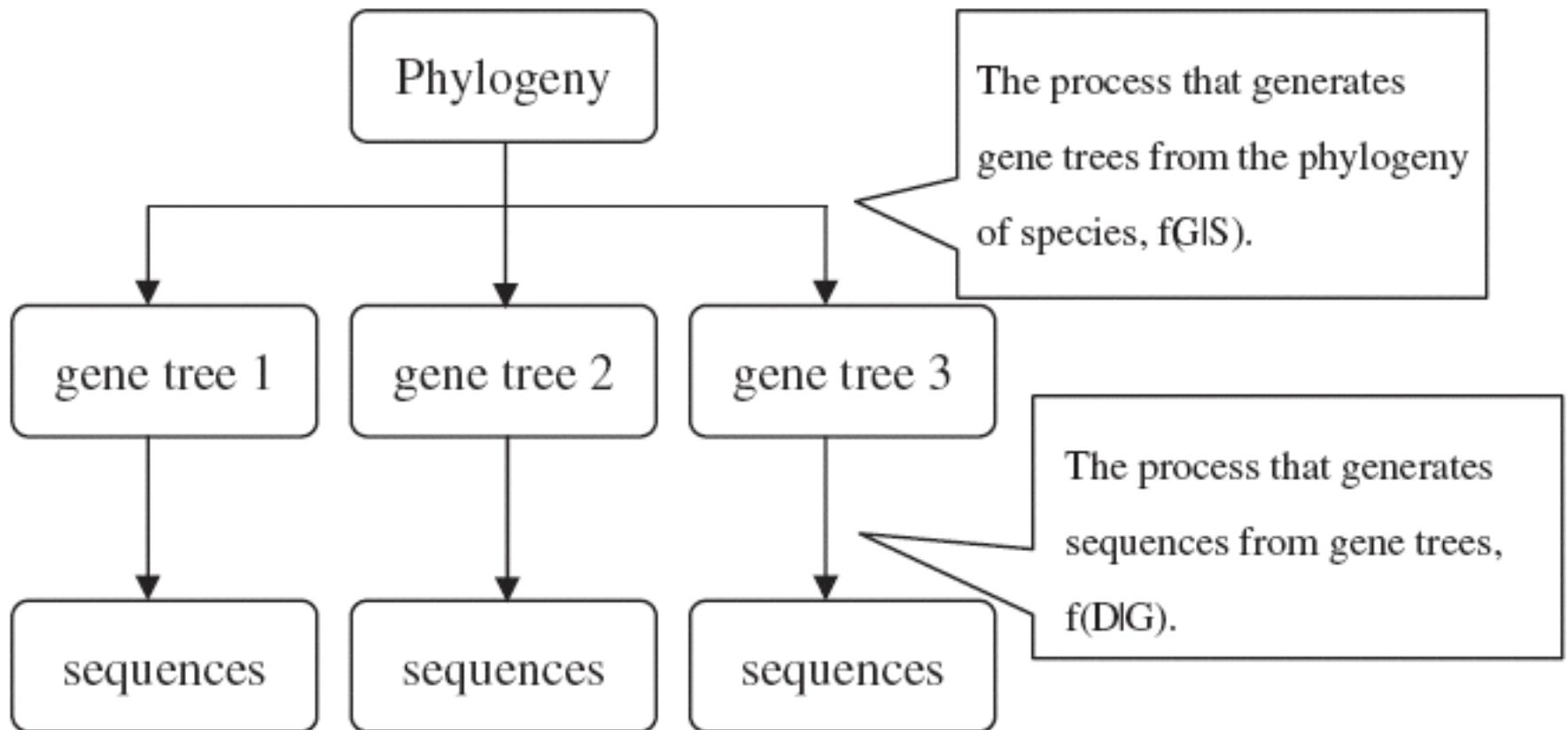


Anatomy of a species tree



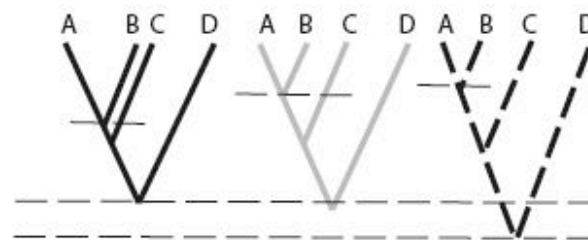
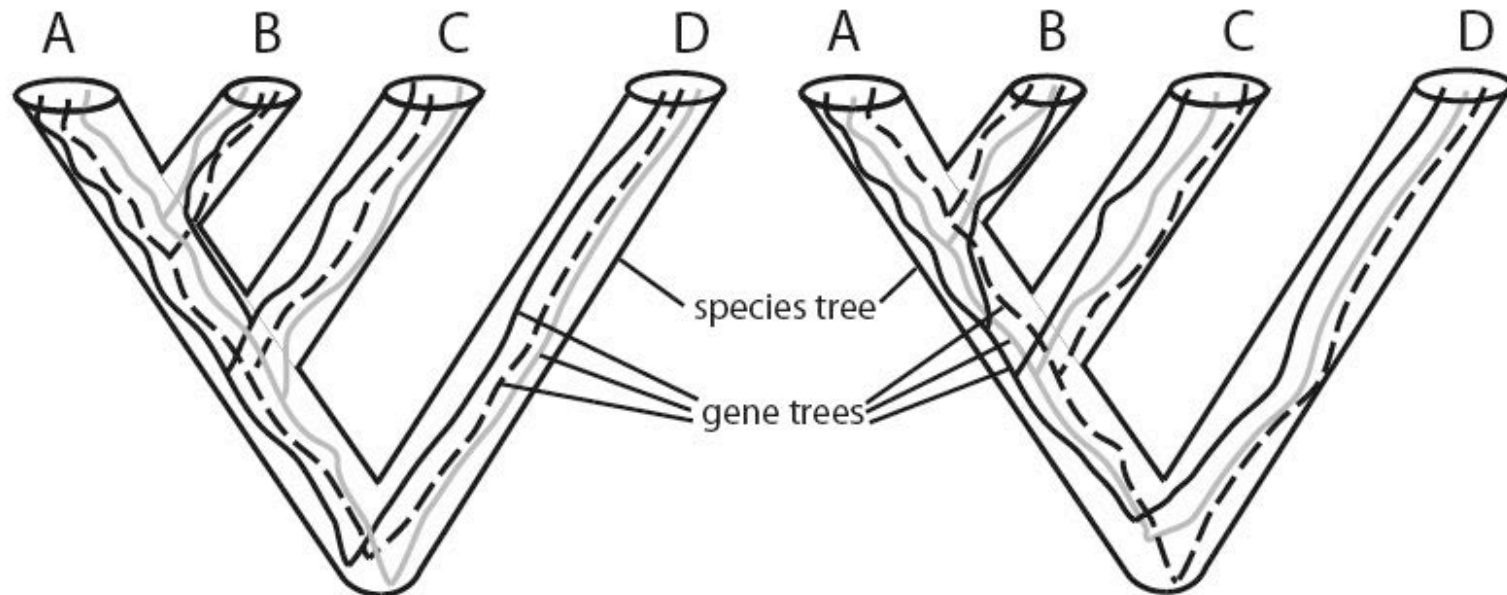
Hierarchical nature of phylogeny



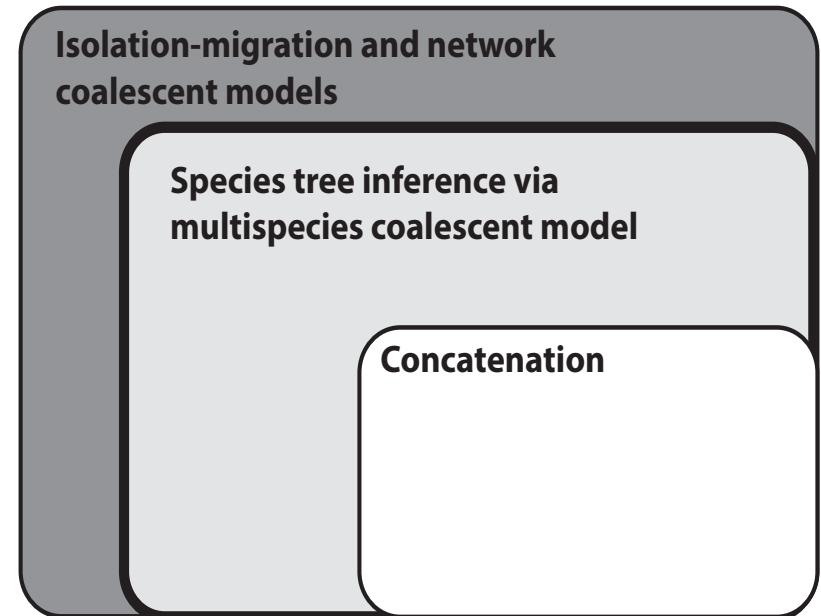
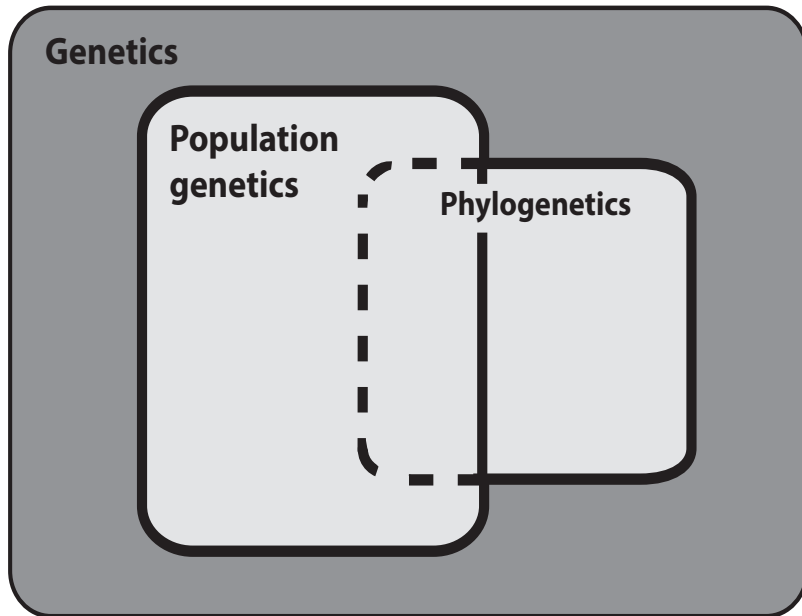
Deep coalescence vs. branch length heterogeneity

Deep coalescence

Branch length heterogeneity



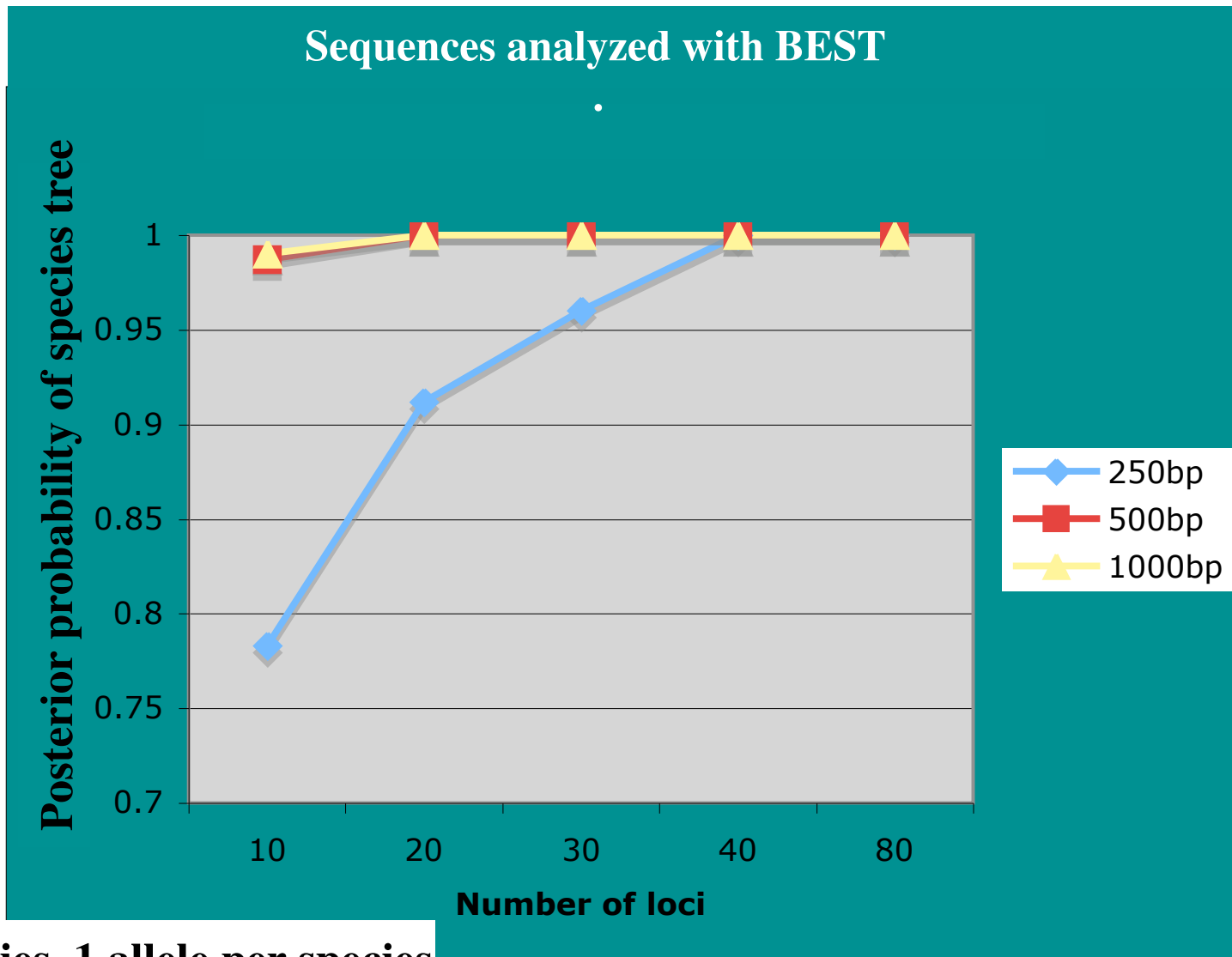
The multispecies coalescent model in context



Key assumption of the MSC:
conditional independence of loci
(mediated by recombination)
***not* presence of incomplete lineage sorting**

Do species tree methods require discordance of gene trees?

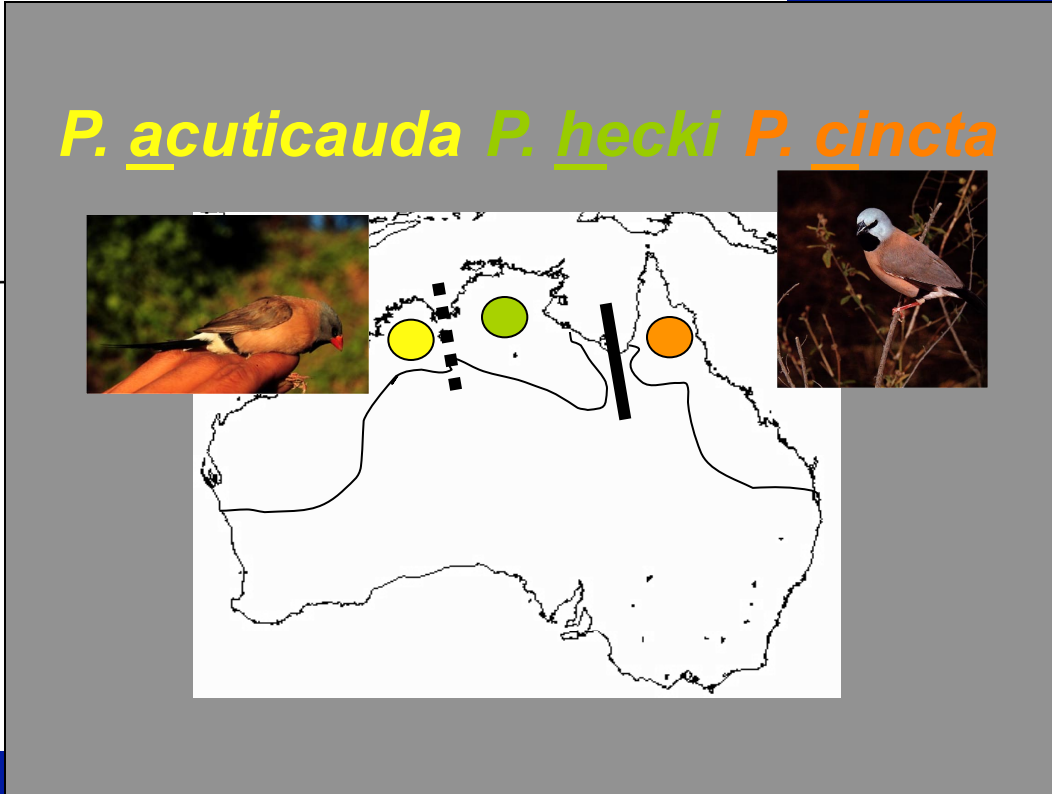
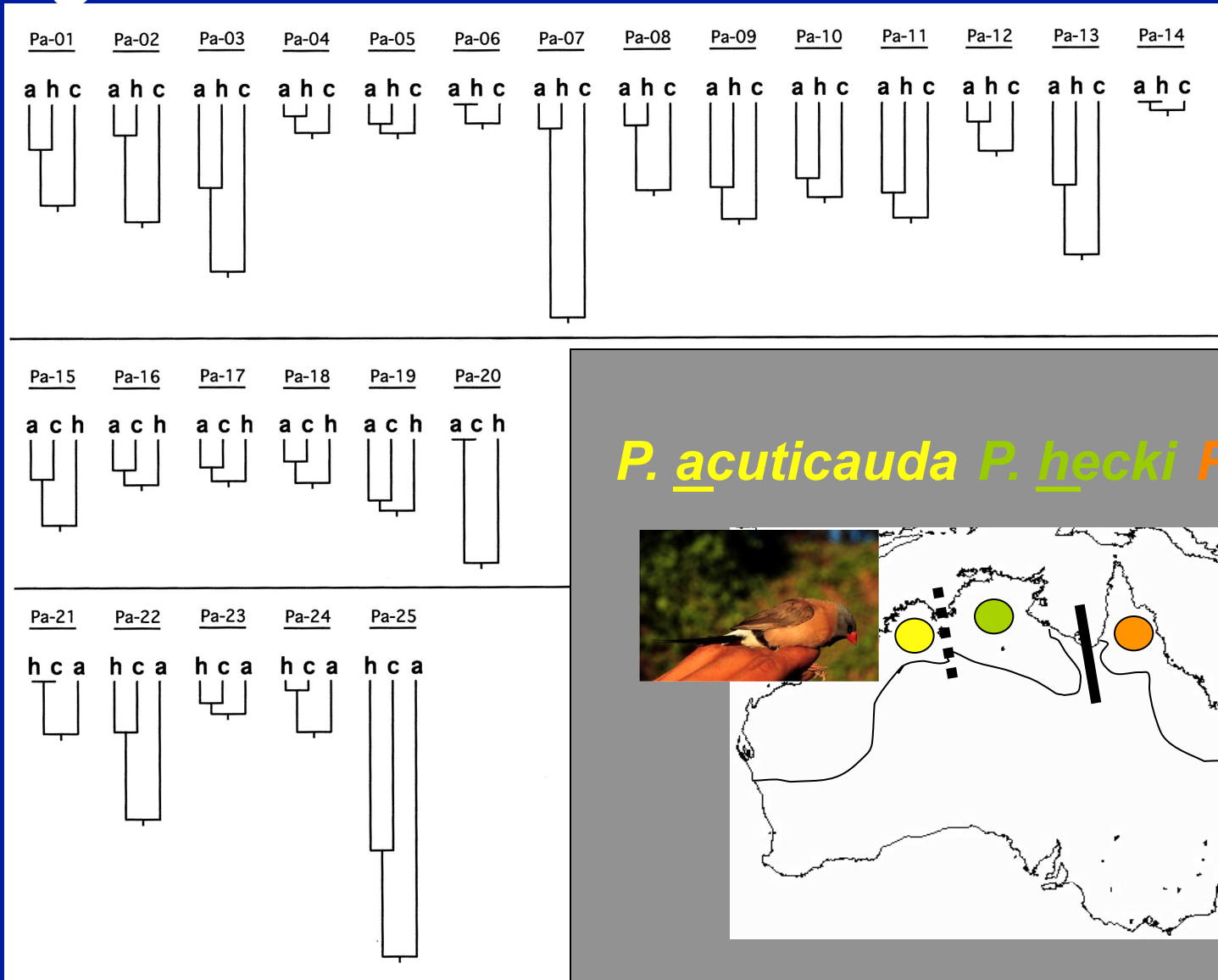
Simulation: all gene trees are topologically congruent with the species tree



8 species, 1 allele per species

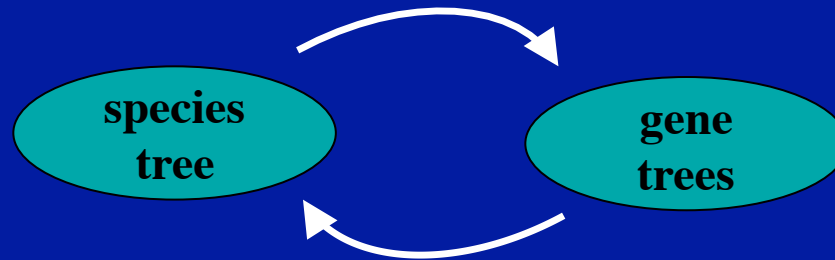
Castillo et al. 2010

30 gene trees from Australian finches



Species trees from gene trees

Bayesian Estimation of Species Trees = The “BEST” method



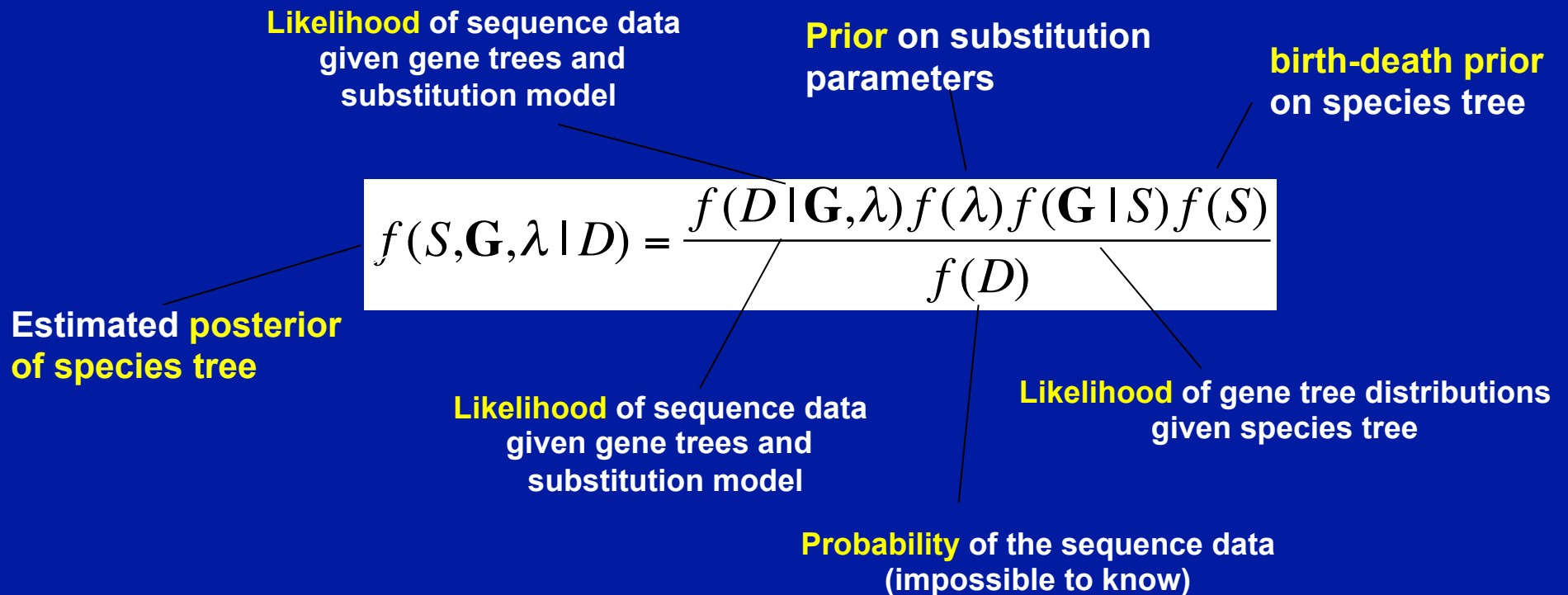
1. Define an approximate species tree as a prior for gene trees
2. Estimate **posterior of species tree $f(S|D)$** using a **birth-death prior** on species tree and **likelihoods of gene tree vectors** from step 1 using coalescent theory.

Liu and Pearl. 2007. Species trees from gene trees: reconstructing posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* 56 (May): 504 - 514.
www.stat.osu.edu/~dkp/BEST

Bayesian hierarchical model

(Liu and Pearl)

Bayesian Estimation of Species Trees = The “BEST” method (ha ha!)



Australo-Papuan Fairy Wrens - Maluridae



Splendid Fairywren (*Malurus*)



Superb Fairywren (*Malurus*)

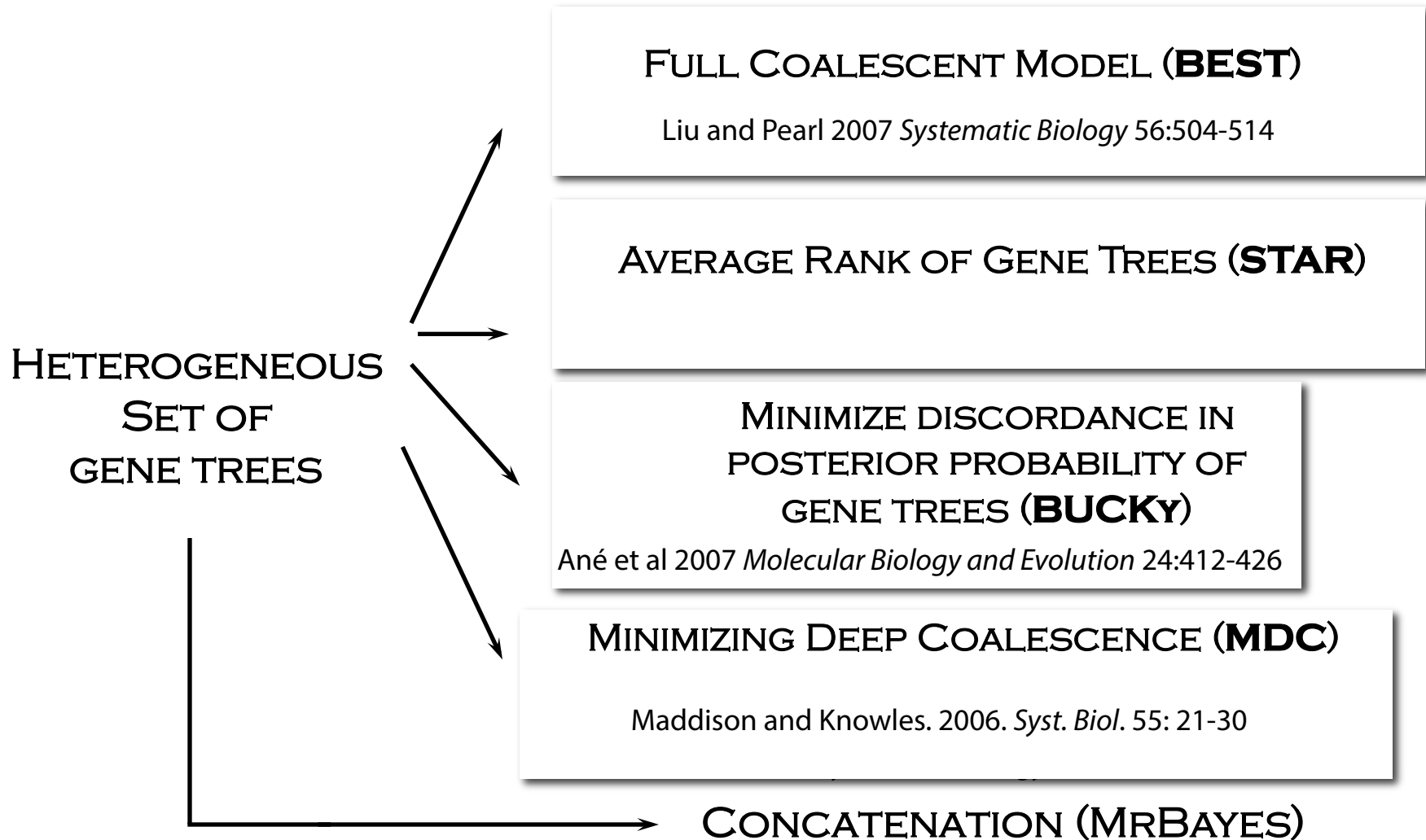


Striated Grasswren (*Amytornis*)



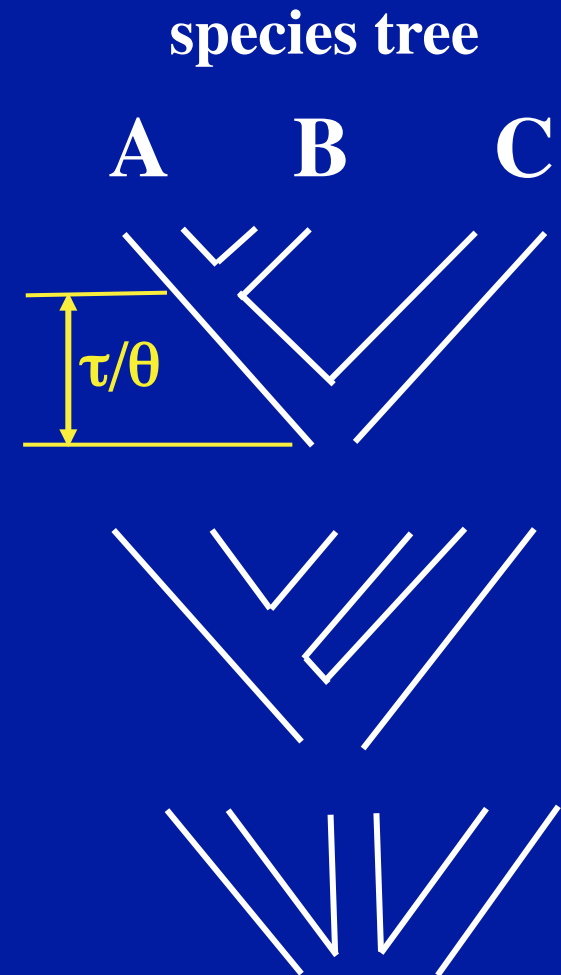
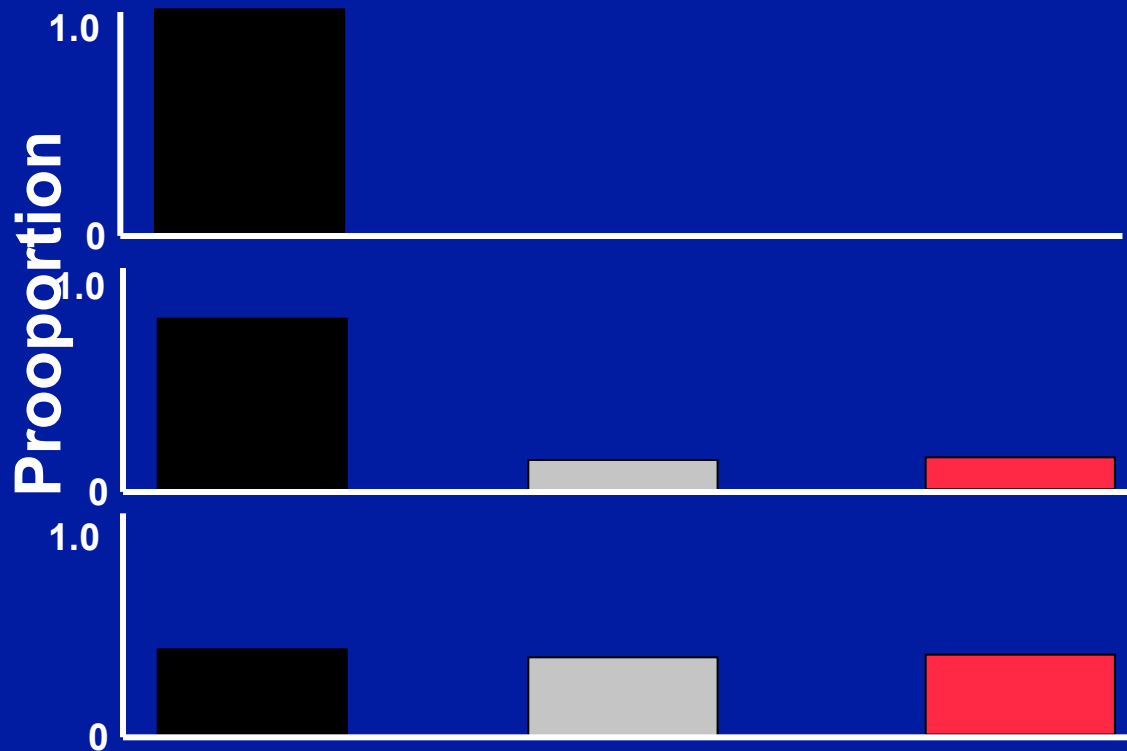
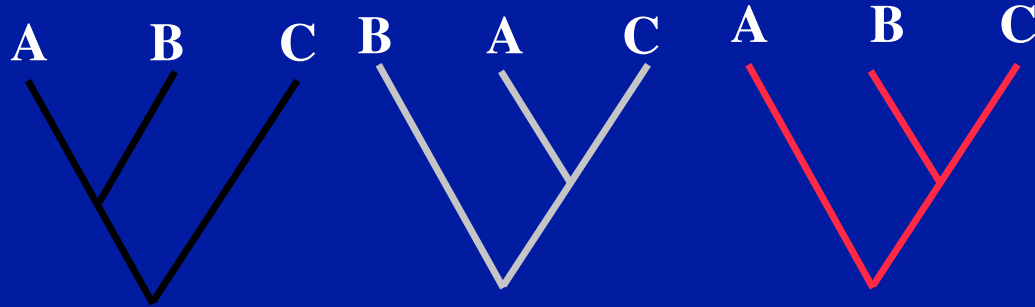
Southern Emu-wren (*Stipiturus*)

Exploring incomplete lineage sorting through species tree methods

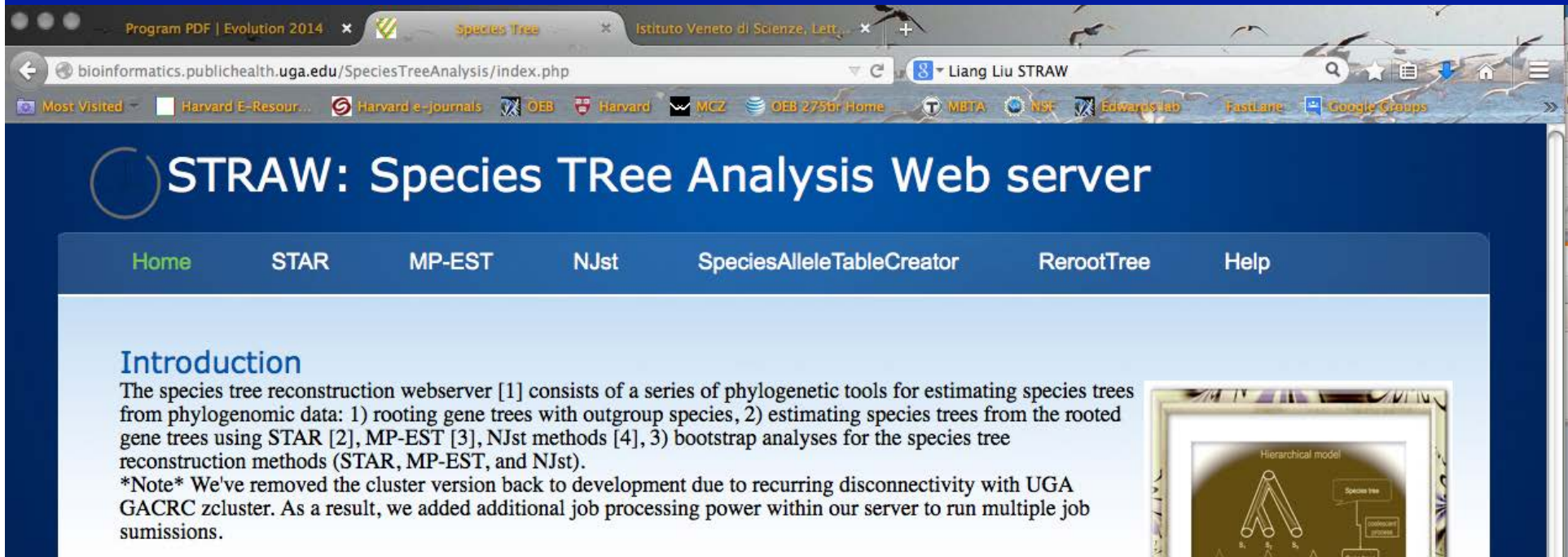


Maximum (pseudo) likelihood method for species trees

Rooted gene tree triplets (topologies only)



Species Tree Web Server: STRAW



The screenshot shows a web browser window with the URL bioinformatics.publhealth.uga.edu/SpeciesTreeAnalysis/index.php. The page title is "STRAW: Species TRee Analysis Web server". The navigation menu includes "Home", "STAR", "MP-EST", "NJst", "SpeciesAlleleTableCreator", "RerootTree", and "Help". The "Introduction" section describes the webserver as a series of phylogenetic tools for estimating species trees from phylogenomic data, listing methods like STAR, MP-EST, and NJst. A note mentions the removal of a cluster version due to connectivity issues. A diagram titled "Hierarchical model" is partially visible on the right.


STRAW: Species TRee Analysis Web server

[Home](#) [STAR](#) [MP-EST](#) [NJst](#) [SpeciesAlleleTableCreator](#) [RerootTree](#) [Help](#)

Introduction

The species tree reconstruction webserver [1] consists of a series of phylogenetic tools for estimating species trees from phylogenomic data: 1) rooting gene trees with outgroup species, 2) estimating species trees from the rooted gene trees using STAR [2], MP-EST [3], NJst methods [4], 3) bootstrap analyses for the species tree reconstruction methods (STAR, MP-EST, and NJst).

Note We've removed the cluster version back to development due to recurring disconnectivity with UGA GACRC zcluster. As a result, we added additional job processing power within our server to run multiple job submissions.



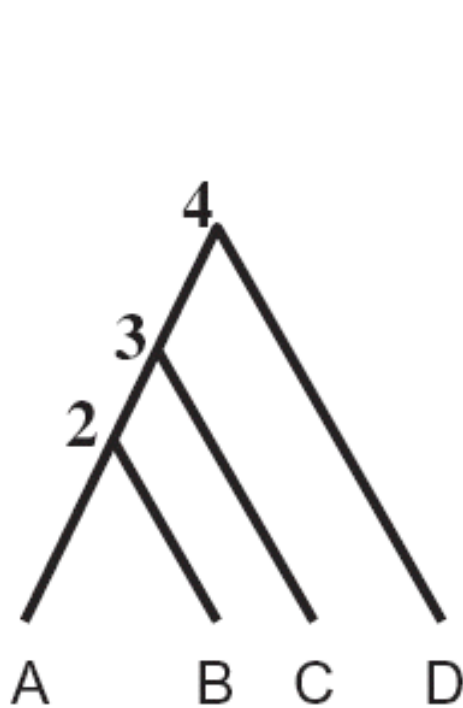
The diagram shows a hierarchical model with a tree structure and a flowchart. The tree has three main branches labeled S_1 , S_2 , and S_3 . The flowchart includes boxes for "Species tree", "Outgroup species", and "Gene tree".

<http://bioinformatics.publhealth.uga.edu/SpeciesTreeAnalysis/index.php>

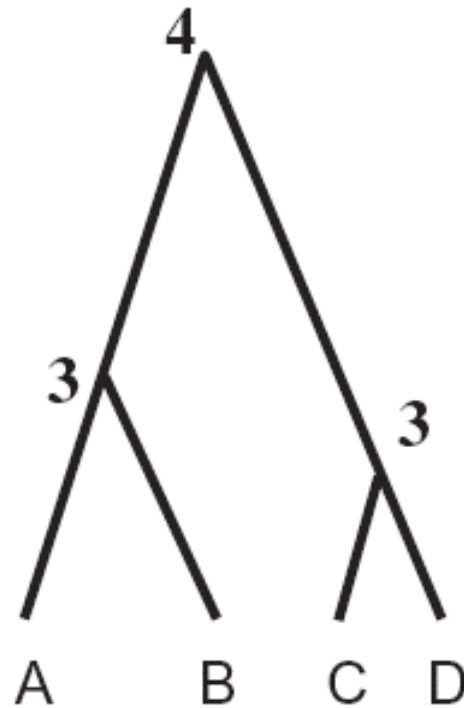
Summarizing variation in gene trees

Compare taxon pairs via

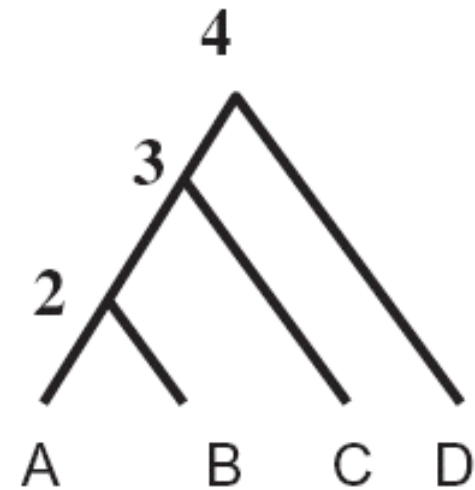
- 1) Ranks 2) coalescence times 3) minimum divergences



gene tree 1

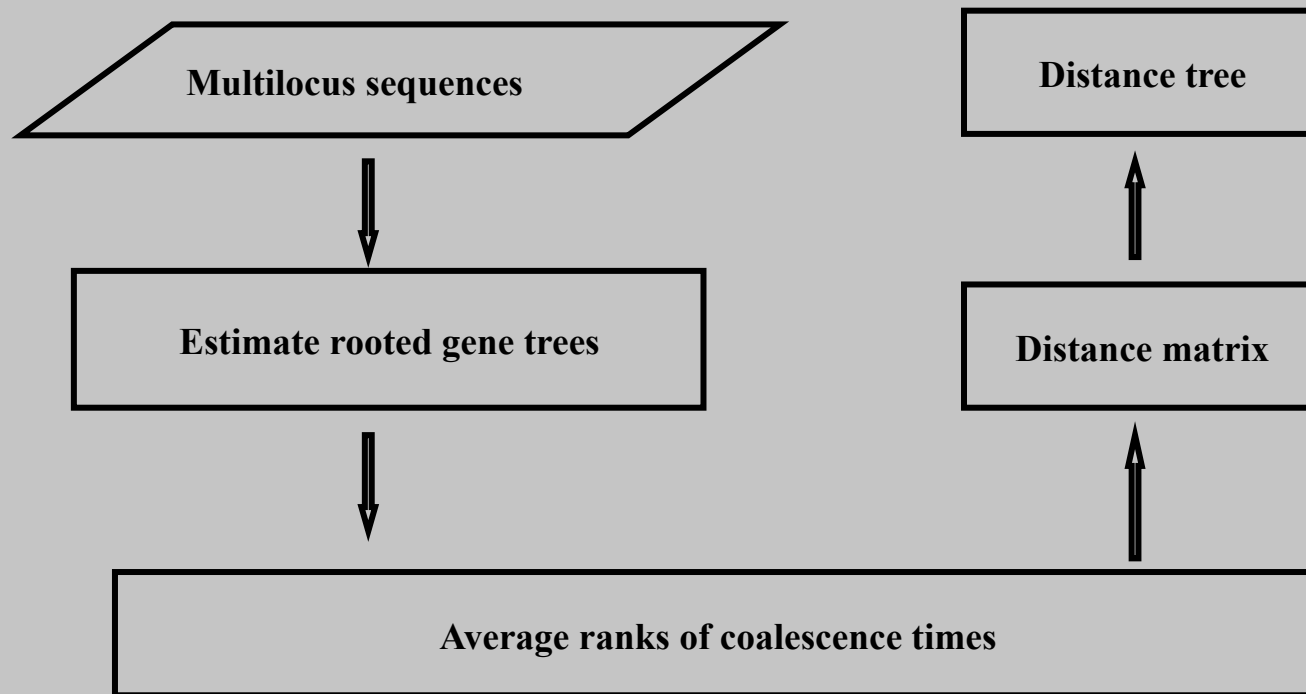


gene tree 2



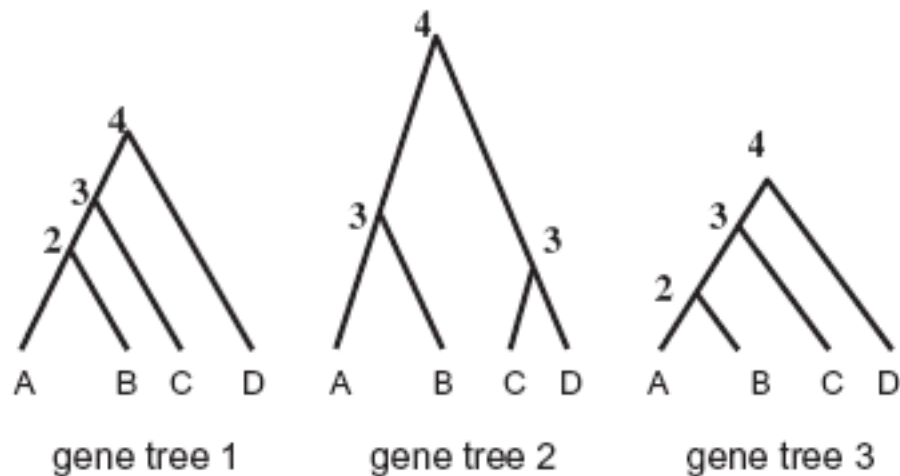
gene tree 3

Species Trees from Average Ranks of Coalescence Times (STAR)



Calculating a STAR tree

(a)

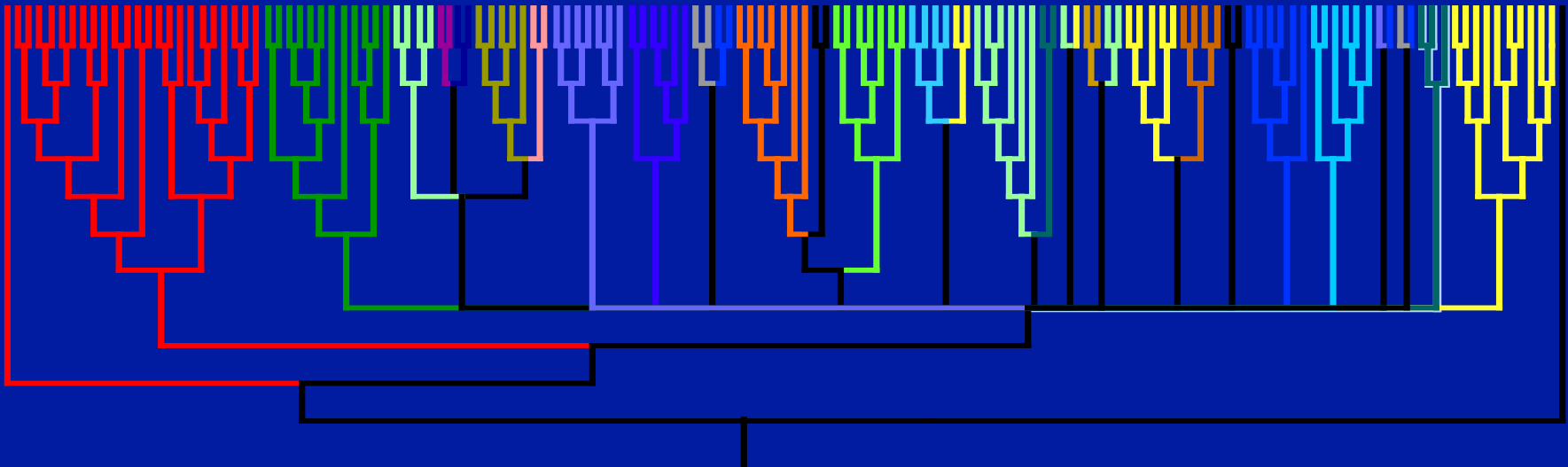


(c)

	A	B	C	D
A	--	4.66	6.66	8
B	4.66	--	6.66	8
C	6.66	6.66	--	7.34
D	8	8	7.34	--



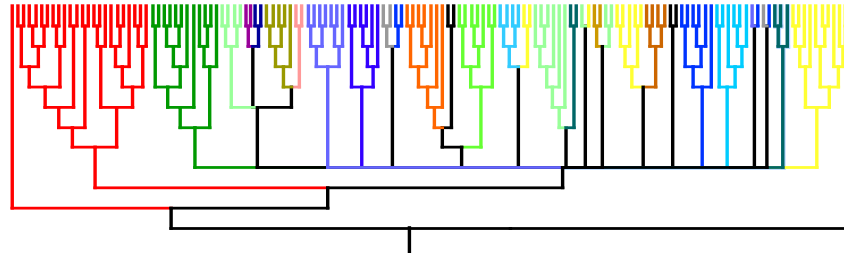
PHYBASE: Constructing, manipulating and evaluating species trees



Liu & Yu. 2010. Bioinformatics 26: 962-963

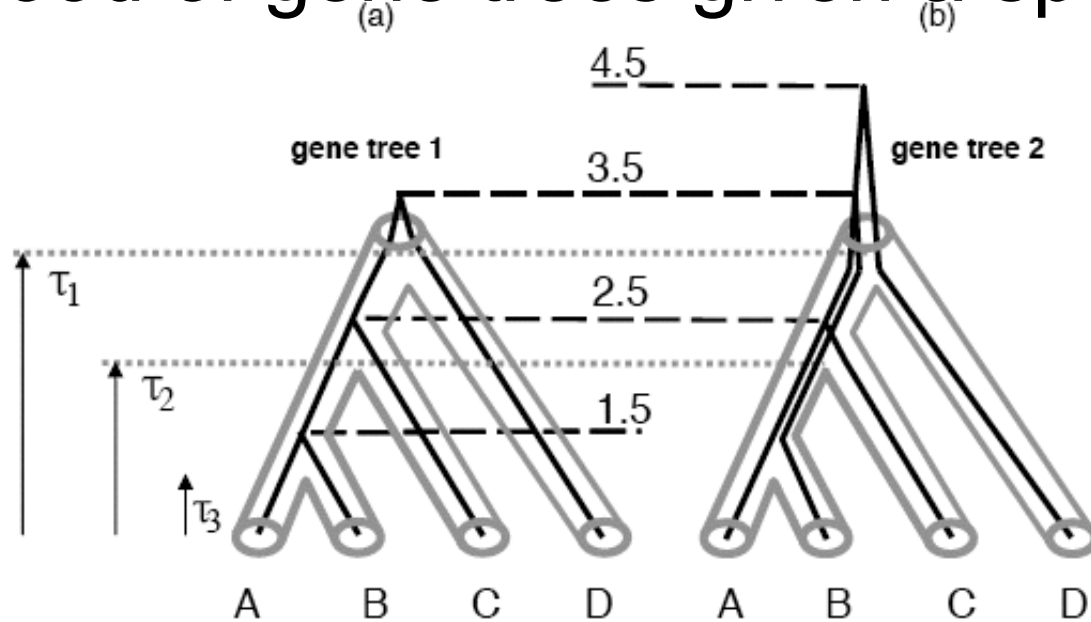
<http://code.google.com/p/phybase/>

Uses of phybase

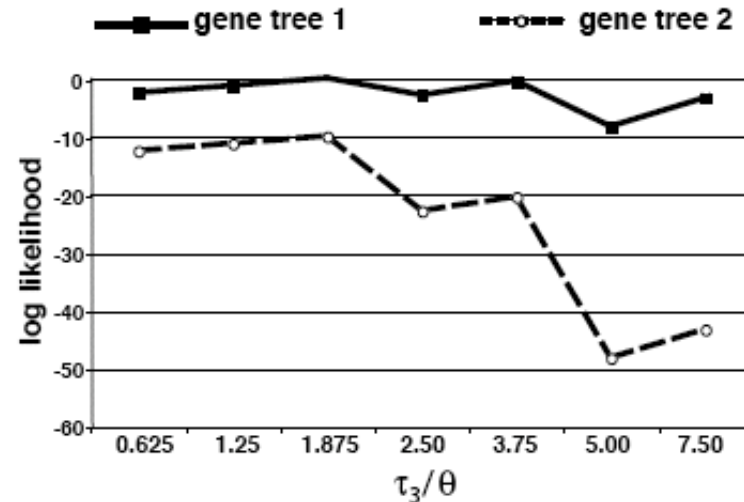


- Conduct multilocus bootstrap
- Estimate STAR, STEAC and Maximum Trees
- Calculate likelihood of species trees given gene trees
- Simulate gene trees (ultrametric and variable rates) and DNA sequences

Likelihood of gene trees given a species tree



(c)



Multilocus bootstrap

Sample loci at random with replacement ...

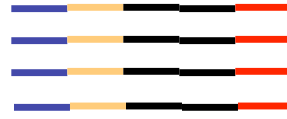
Pseudomatrix 1



Pseudomatrix 2

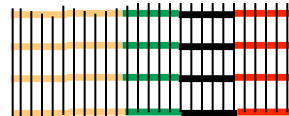
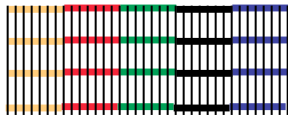


Pseudomatrix 3



•••••

Then sample sites within loci with replacement, just as in the normal bootstrap....



•••••

Seo, T.-K. 2008. Calculating Bootstrap Probabilities of Phylogeny Using Multilocus Sequence Data. *Mol. Biol. Evol.* 25:960-971.

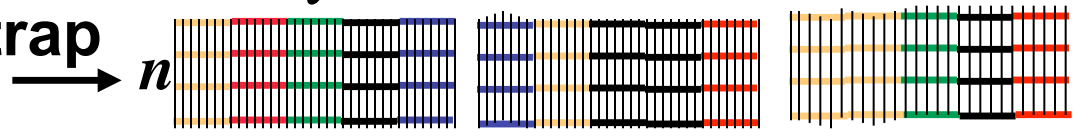
Pipeline for STAR/STEAC trees in phybase

n species, l loci for each replicate
 l

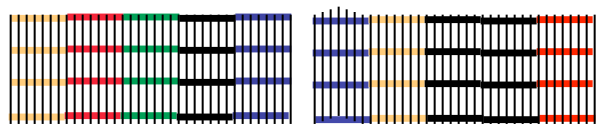
Phybase

Multilocus bootstrap

(b replicates)



phyml



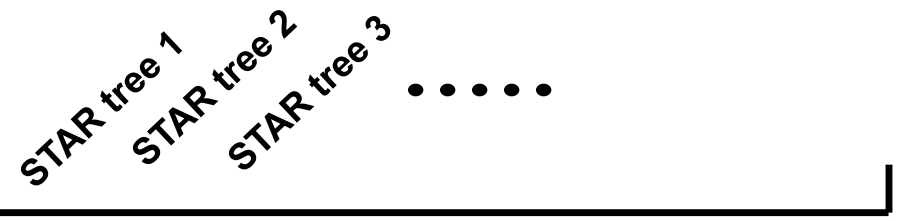
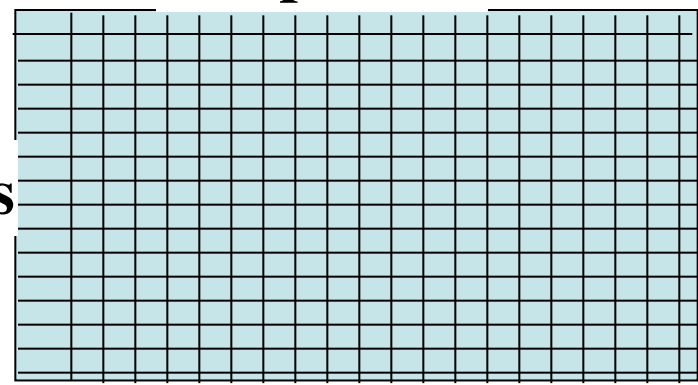
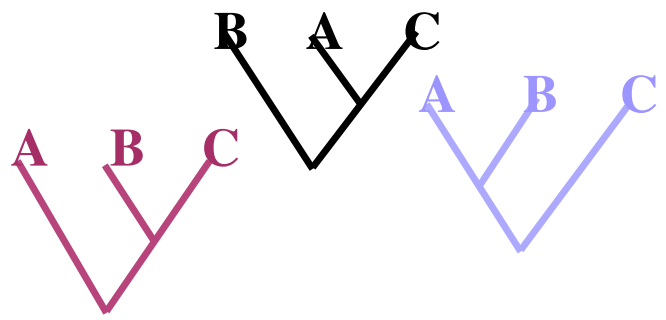
Make gene trees for all $b \cdot l$ single gene matrices

$\leftarrow b$ replicates \rightarrow

Phybase

Make matrix of gene trees

n gene trees

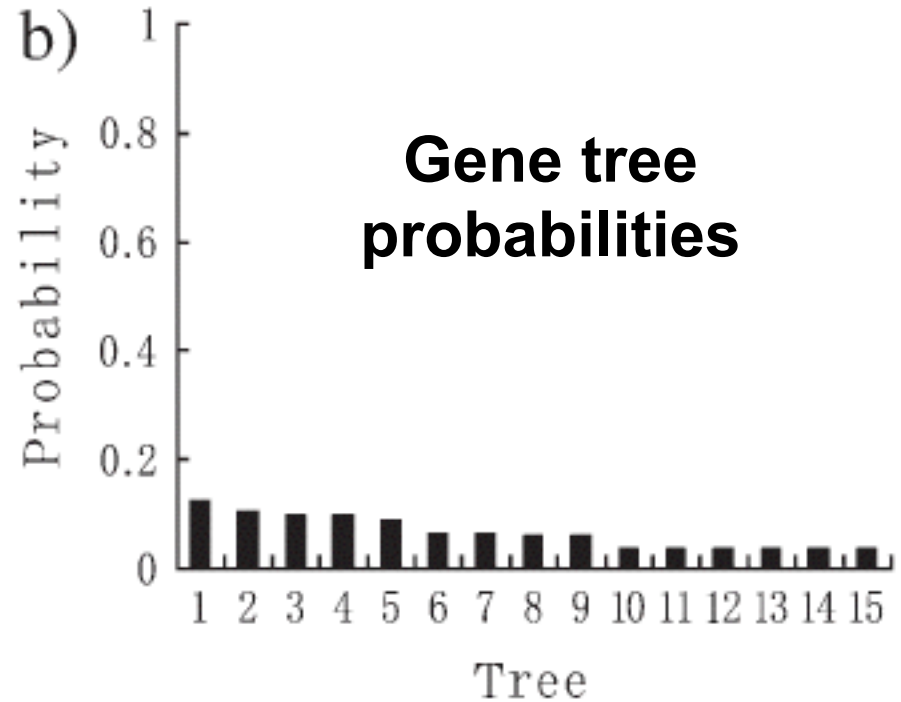
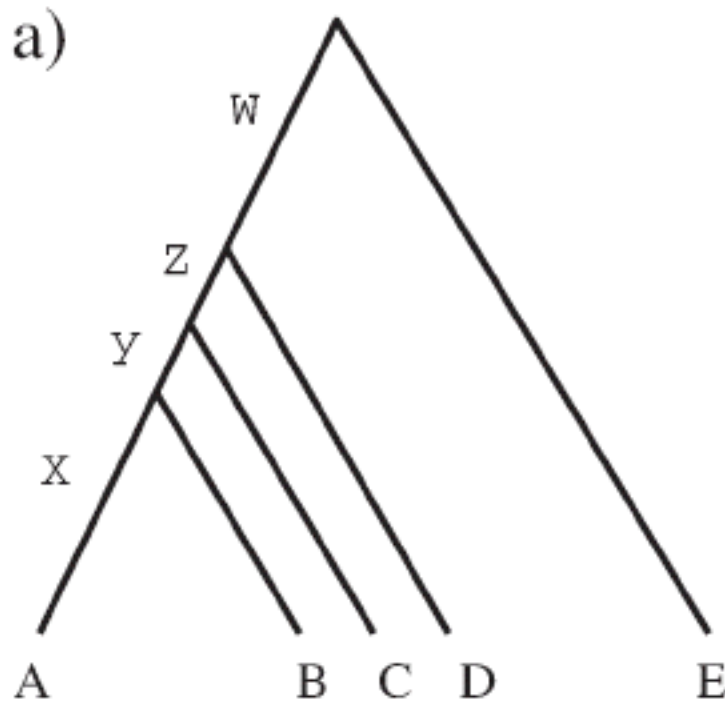


Phybase or other program

Consensus STAR tree

The anomaly zone

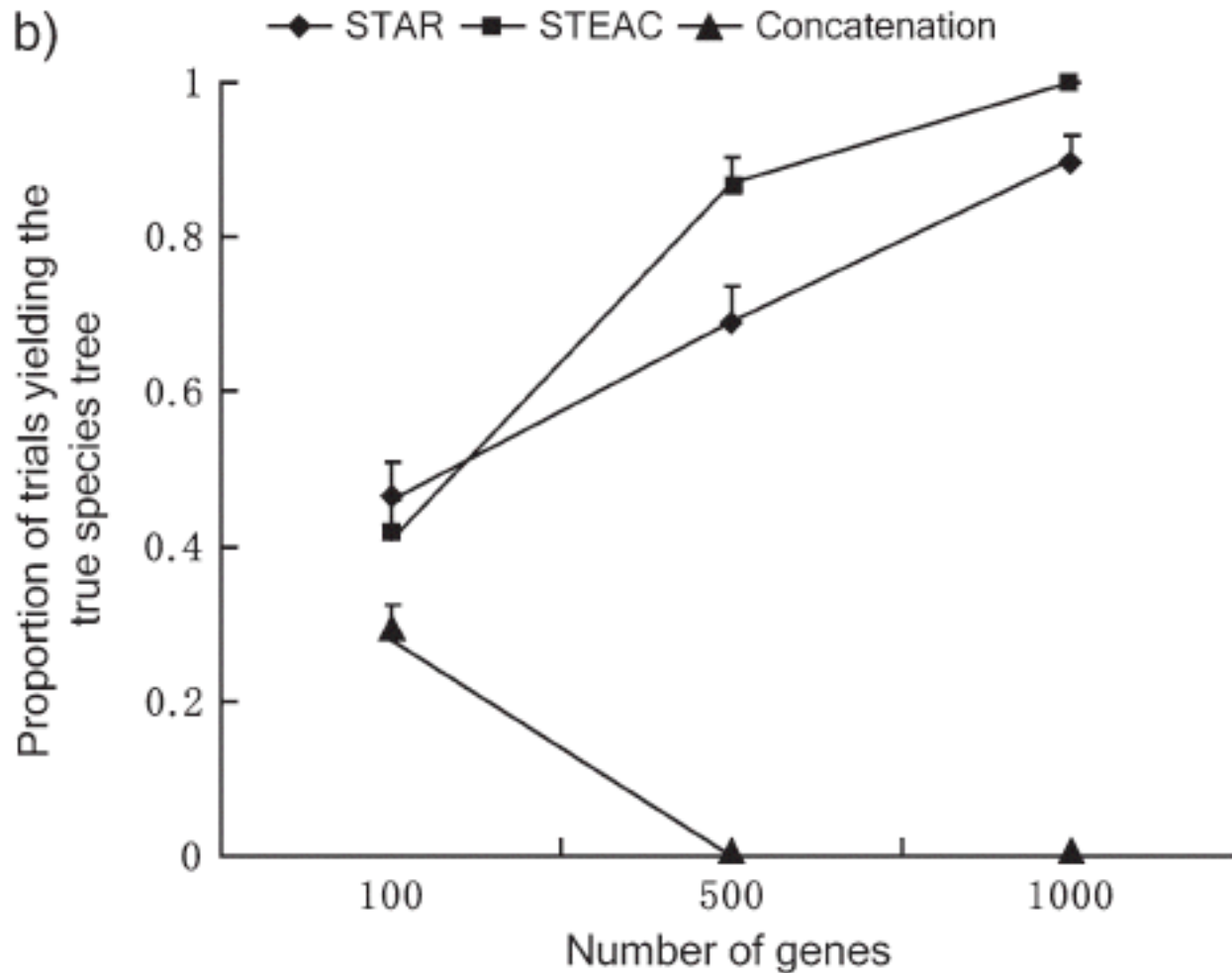
Species tree



Gene trees

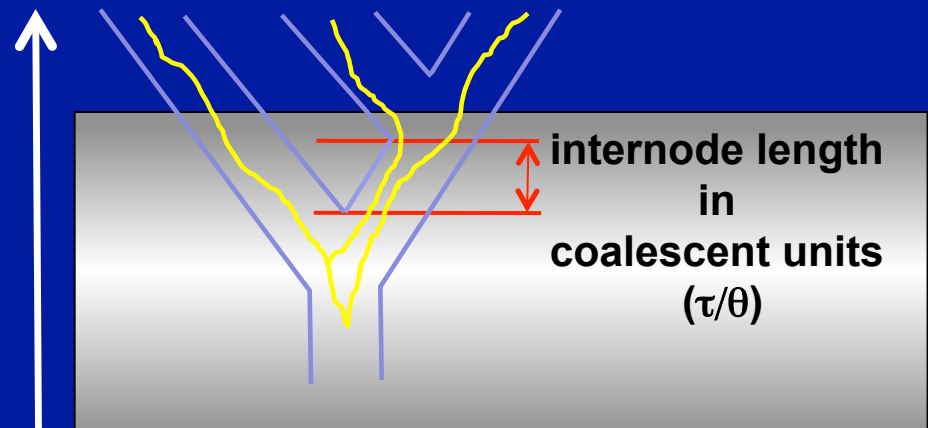


Phylogenetic analysis in the anomaly zone



Liu, L., et al. 2009. *Syst Biol* 58:468-477.

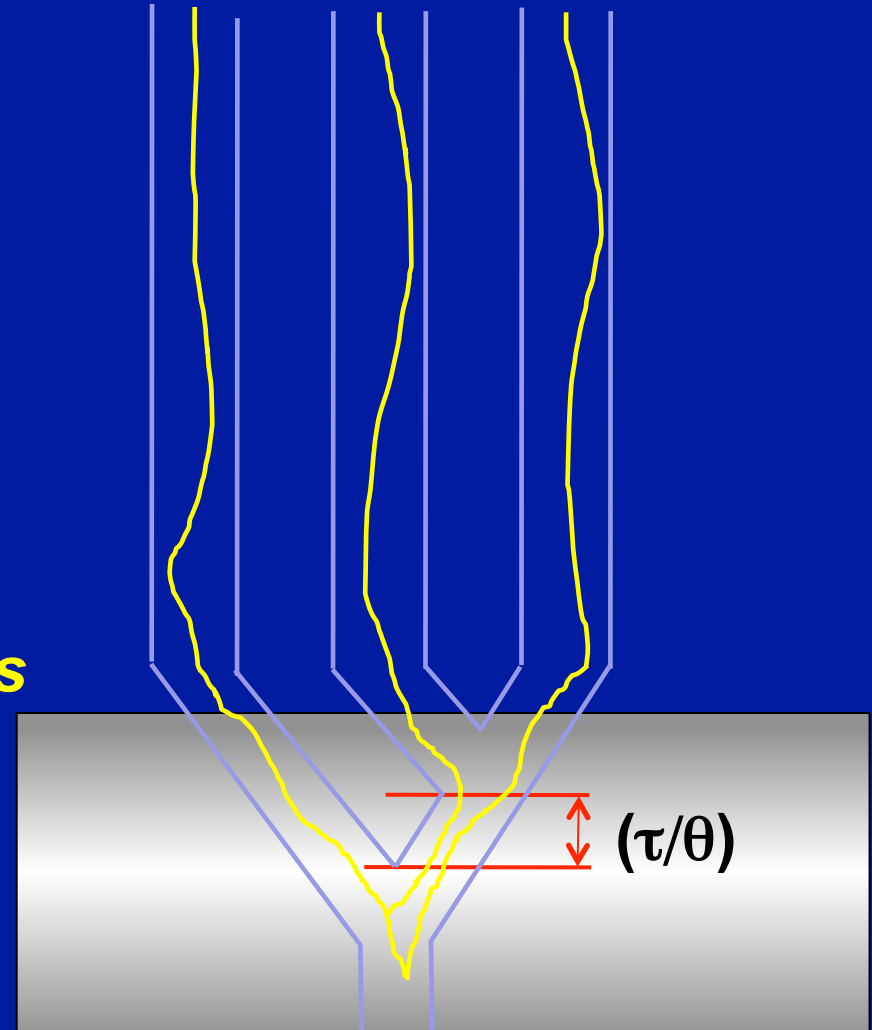
The multispecies coalescent applies to ancient as well as recent divergences



Pleistocene

***Probability of incongruence is the same in both cases!!
Think: "short", not "recent"***

**Absolute time
in the past**

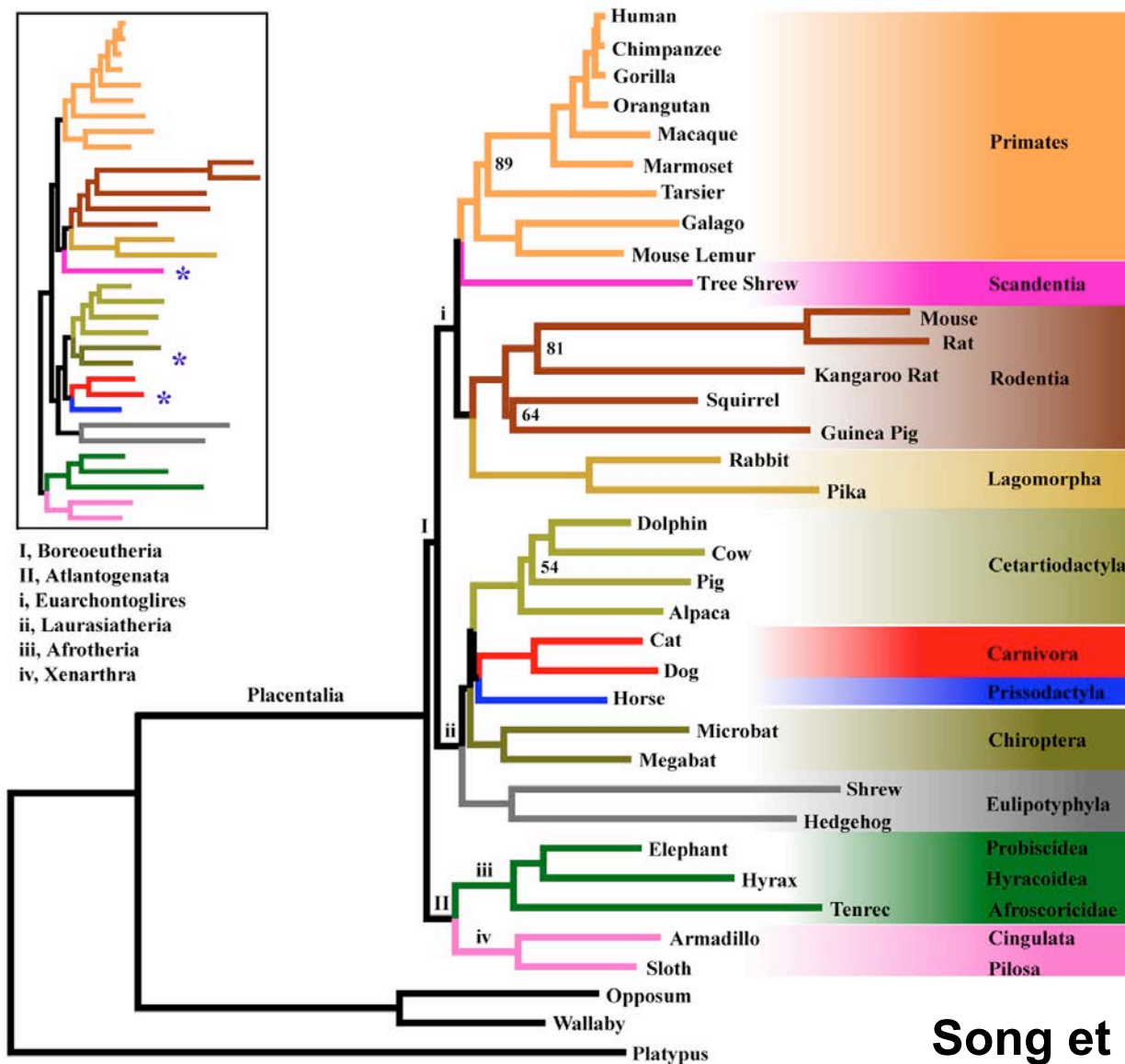


K/T boundary

Mammal data set

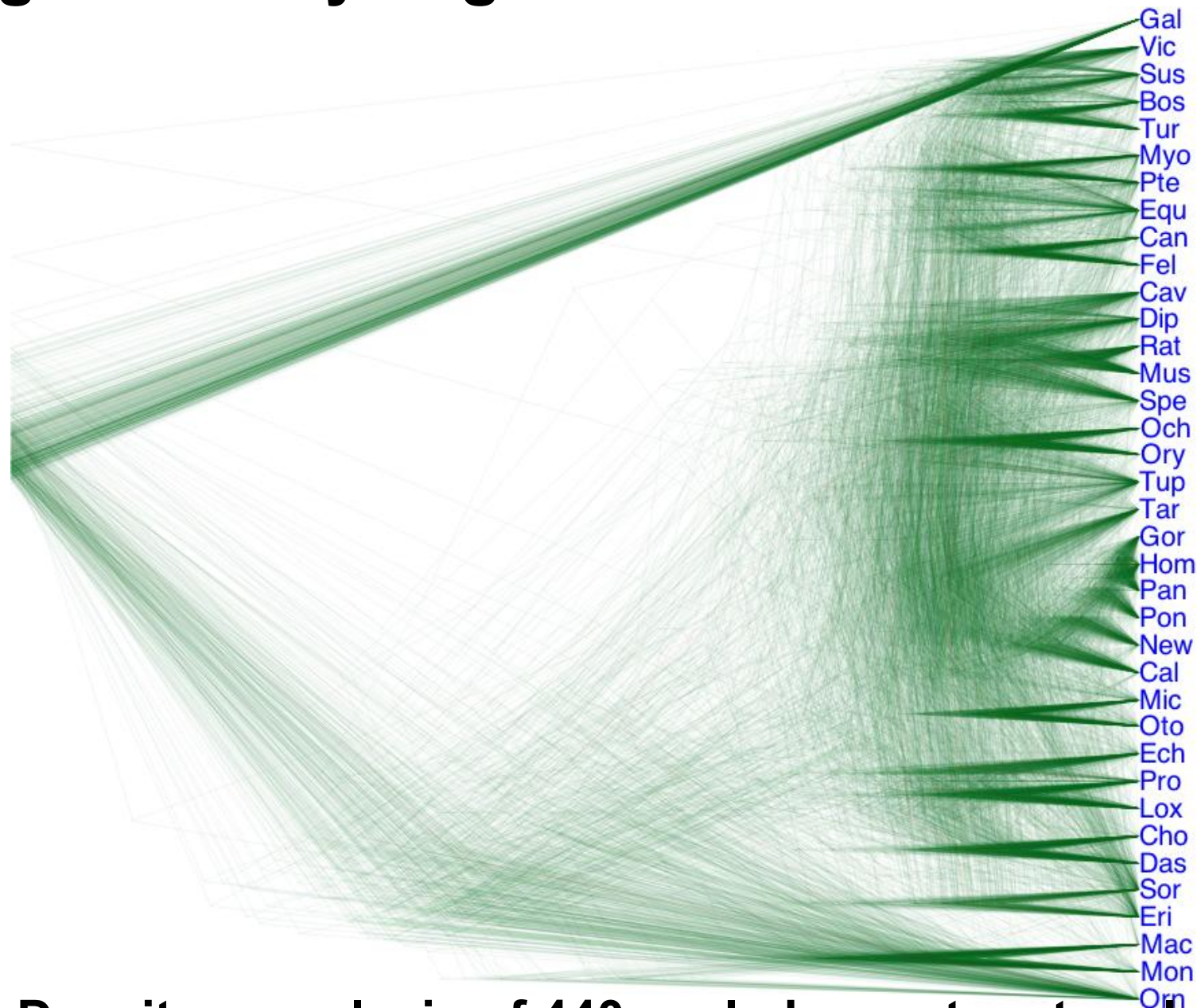
- 37 ingroup taxa (placental mammals)
- 758 genes from Orthomam database
- 447 genes passed filtering
- Total ~1.38 Mb per species, avg. ~3.1 kb/gene
- Gene trees made via RaxML and best-fit substitution model
- MP-EST and STAR method
- ML and Bayesian supermatrix analysis

Conflict between concatenation and species tree approaches



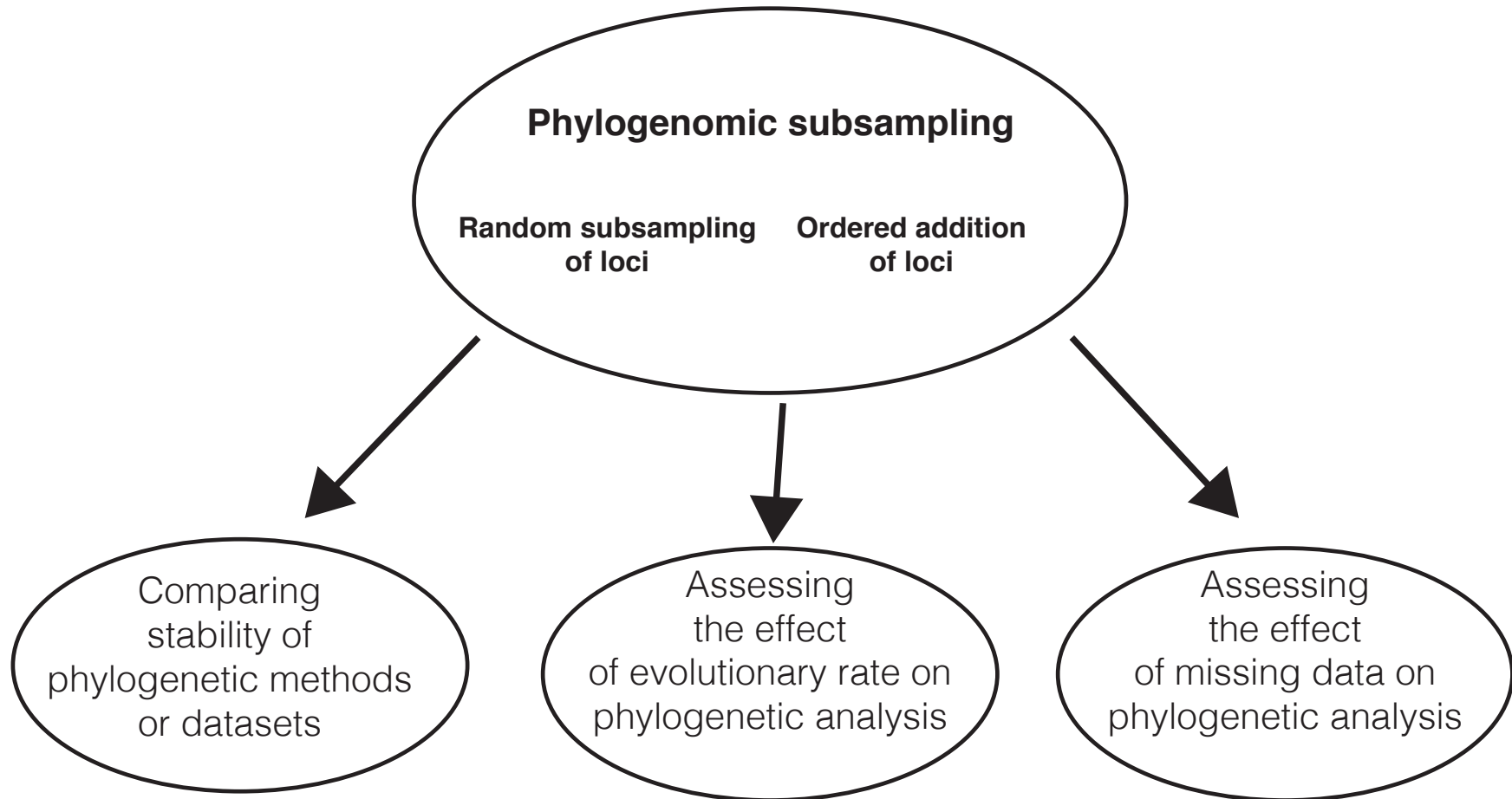
Song et al. 2012. PNAS

Large diversity of gene trees in mammal data set

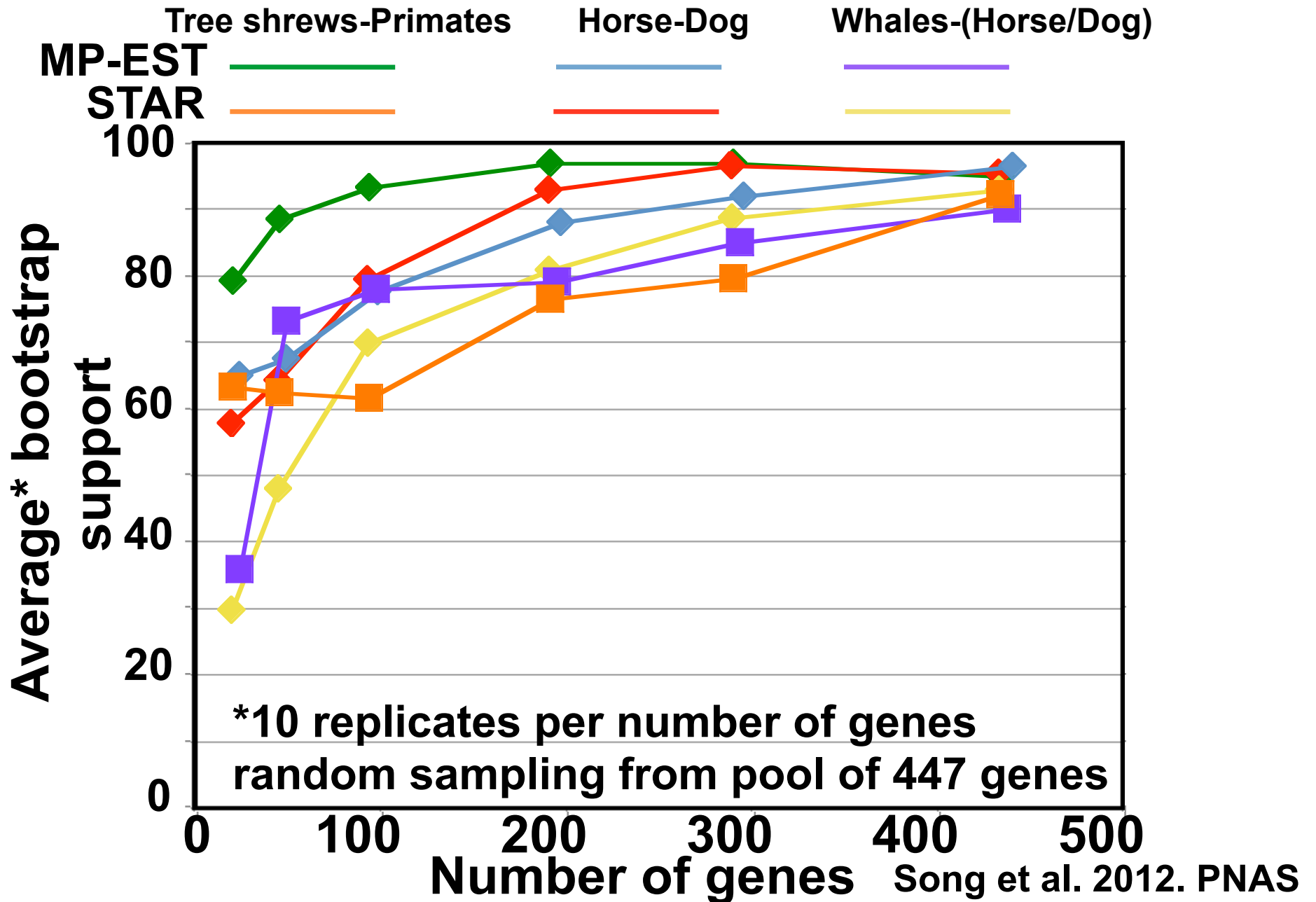


Densitree analysis of 440 scaled gene tree topologies

Phylogenomic subsampling



More genes = more support



Concatenation: more genes = strong but conflicting support for clades

- Tree shrews-Primates
- Whales-(Horse/Dog)
- < 90%, < 0.9
- Whales-Bats
- Tree shrews-Rodents
- Bats-(Horse/Dog)

coalescence | MP-EST
 concatenation | STAR
 Bayesian
 ML

coalescence | MP-EST
 concatenation | STAR
 Bayesian
 ML

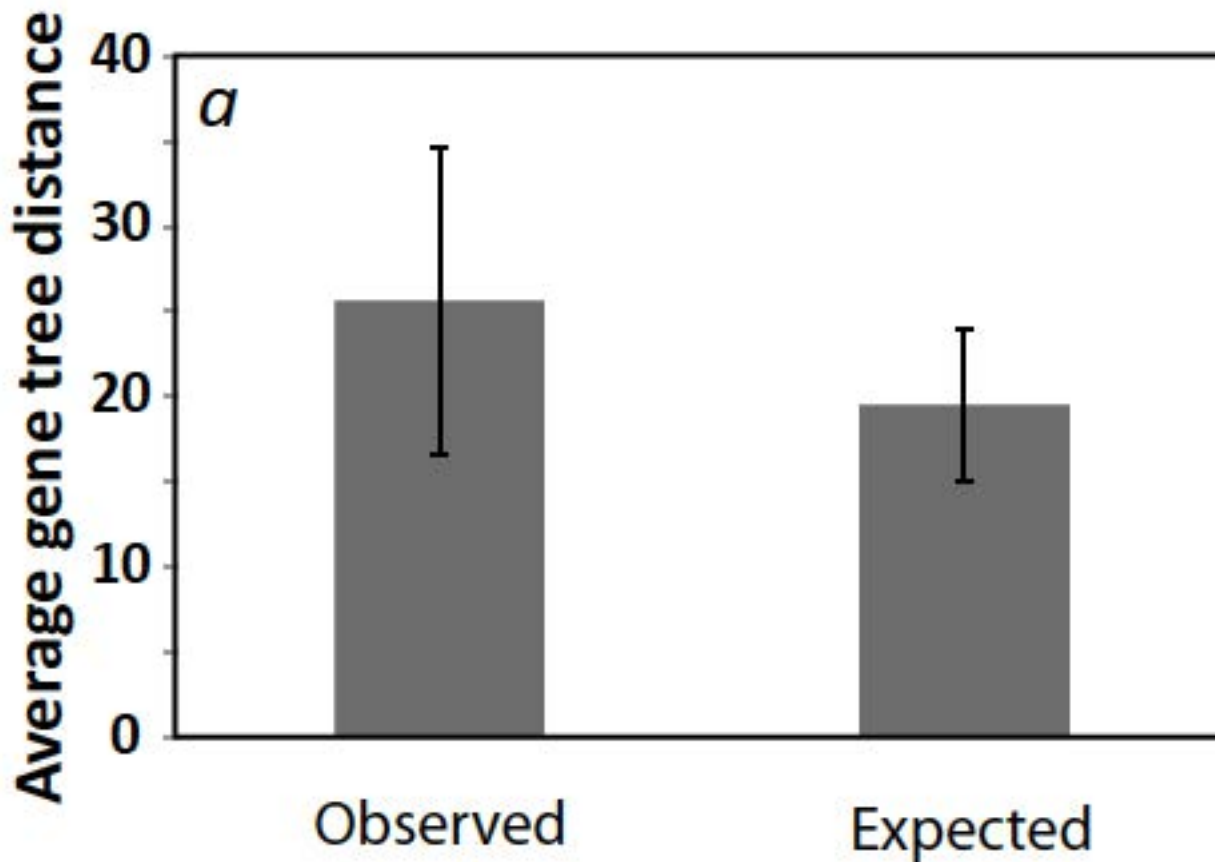
25 50 100 200 300 447 27 21

 Number of genes Number of taxa

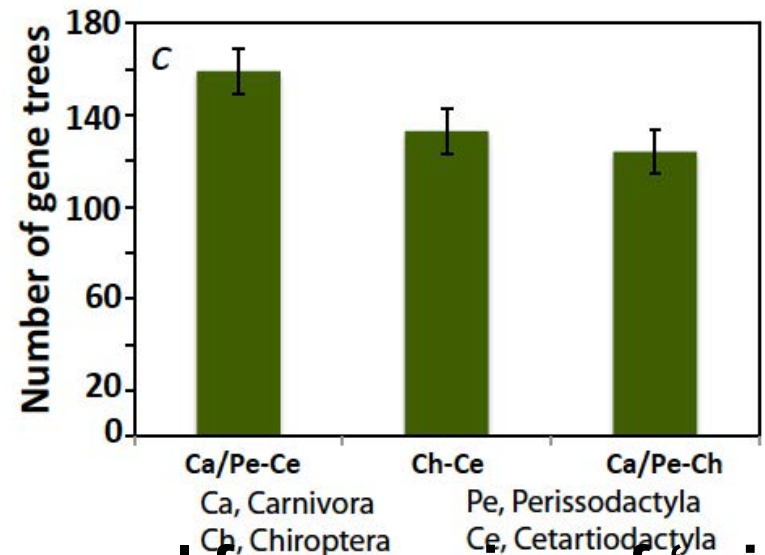
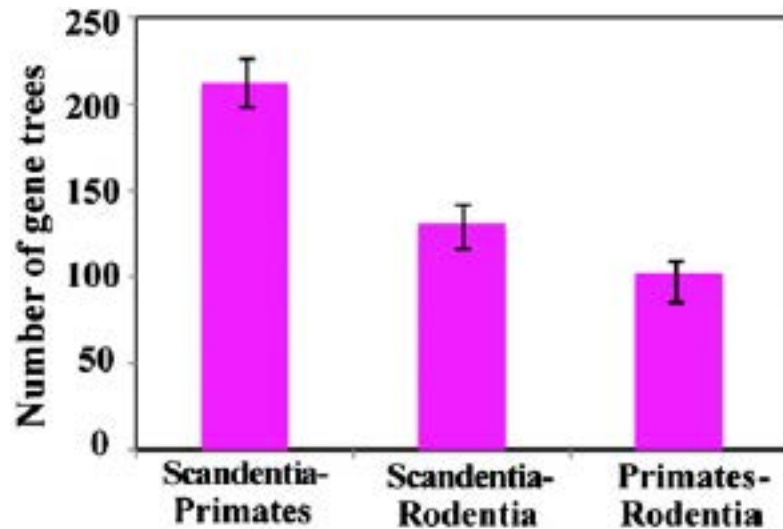
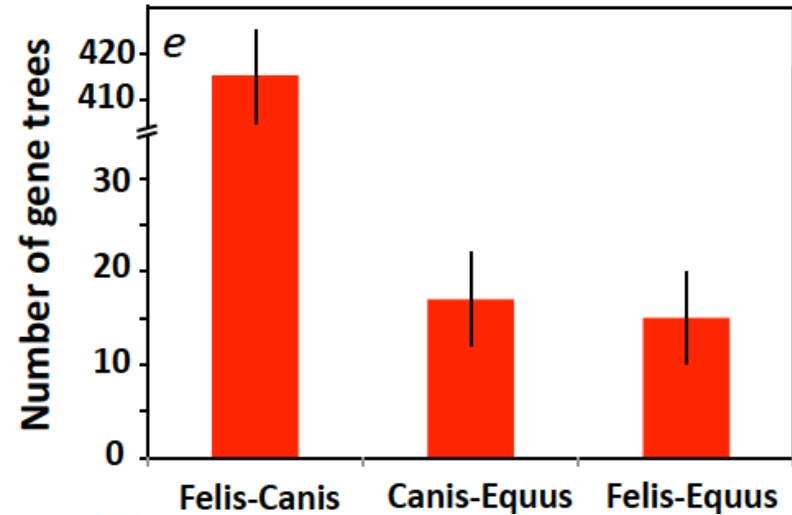
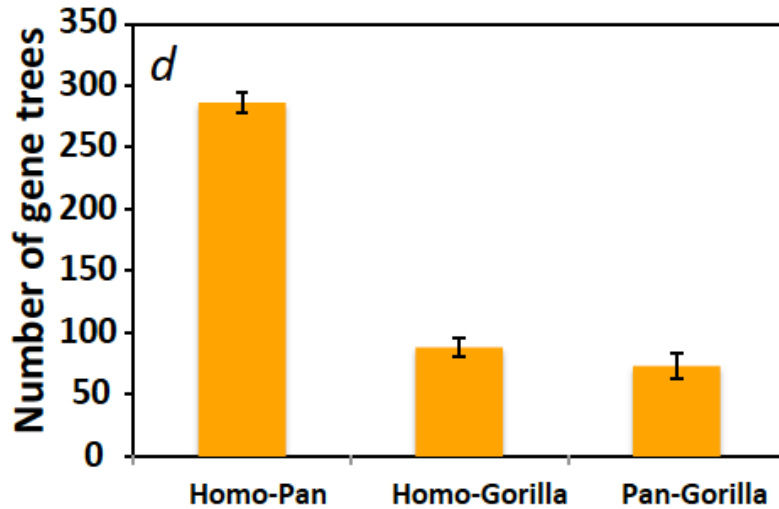
10 replicates (gene sets) per box, random sampling
 from pool of 447 genes

Song et al. 2012. PNAS

Multispecies coalescent model explains ~77% of gene tree variation



Gene tree distributions consistent with multispecies coalescent model*



*Model predicts equal frequencies of 'minority gene trees'

Misconceptions about “species tree” methods

- Species tree methods require discordance among gene trees – **FALSE**
- Species tree methods don't acknowledge gene tree estimation error – **FALSE**
- Species tree methods attribute all gene tree heterogeneity to incomplete lineage sorting – **TRUE and FALSE**