

An integrative model of species evolution accounting for fossilization and coalescence

Huw A. Ogilvie

Department of Computer Science
Rice University, Houston TX

16 August 2019, Taming the BEAST eh!

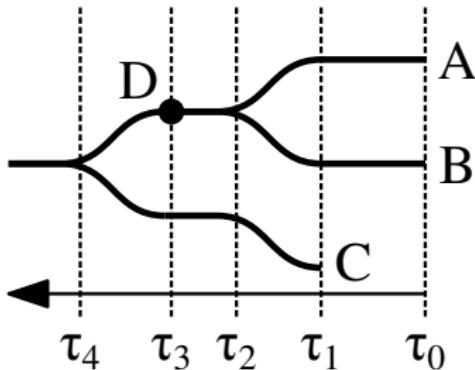
Introduction

“FBD–MSC”

A new model, the **fossilized birth–death multispecies coalescent**,
for inference of **absolute time trees**.

Time trees

What are time trees?

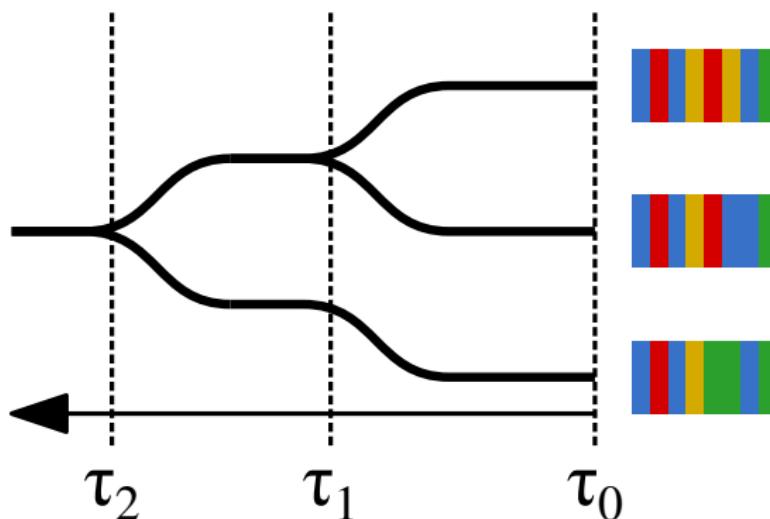


Observations can be:

- A, B in the present
- C, D from the past
- C from a lineage unsampled or extinct in the present
- D a direct ancestor (a “sampled ancestor”)

Time trees of genes

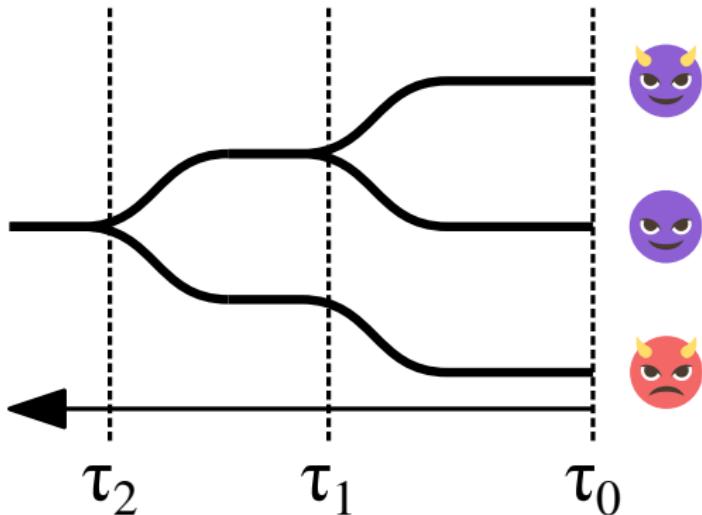
Time trees of genes represent a coalescent history



Observed sequences must share a common ancestor (the sequences are “homologous”)

Time trees of species

Time trees of species represent the history of speciation



Incomplete lineage sorting 1

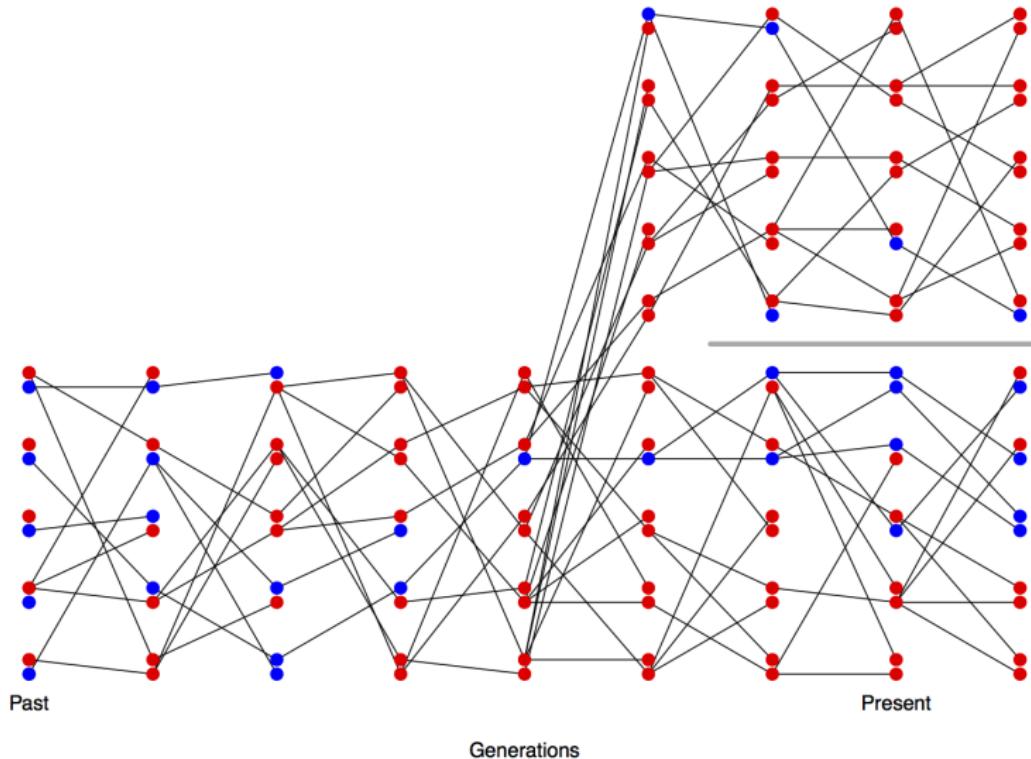


Figure from Graham Coop

Incomplete lineage sorting 2

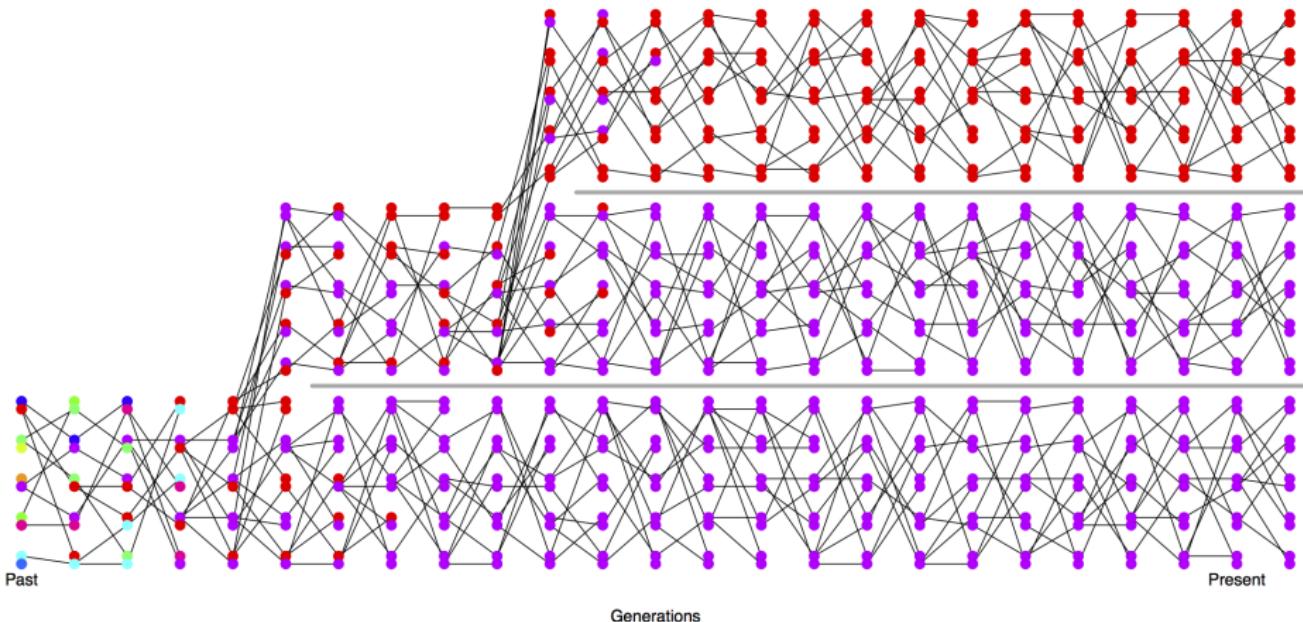
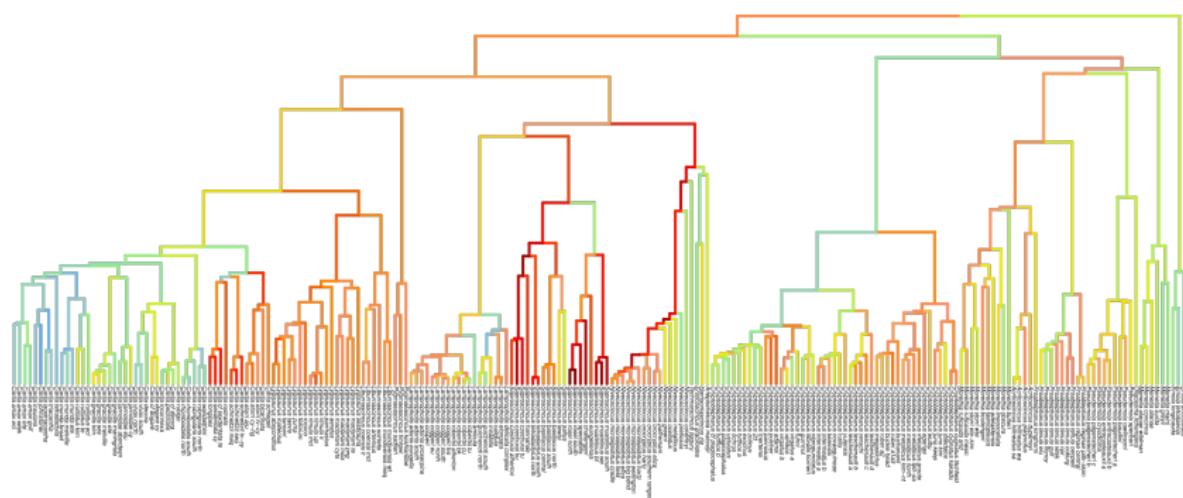


Figure from Graham Coop

Relative time trees 1

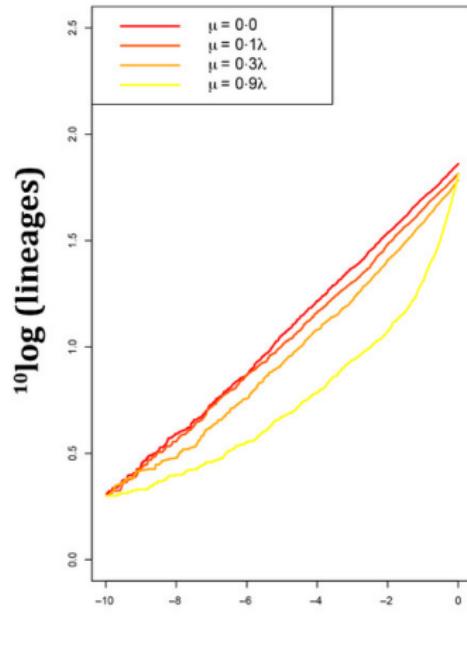
Branch lengths and node heights in substitutions per site or some arbitrary scale.

Useful for studies of micro- and macro-evolutionary processes. For example, identifying regime shifts in macroevolutionary rates using methods such as BAMM, RPANDA, MSBD

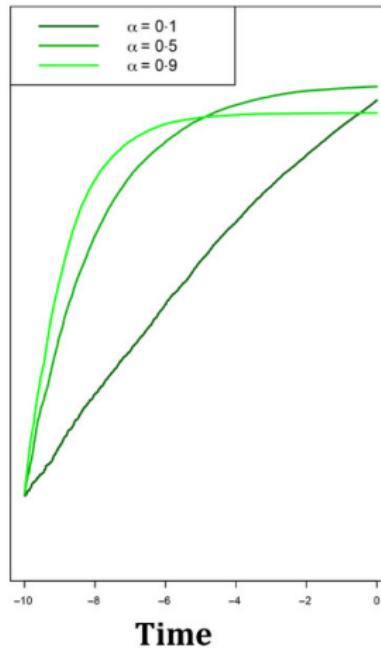


Relative time trees 2

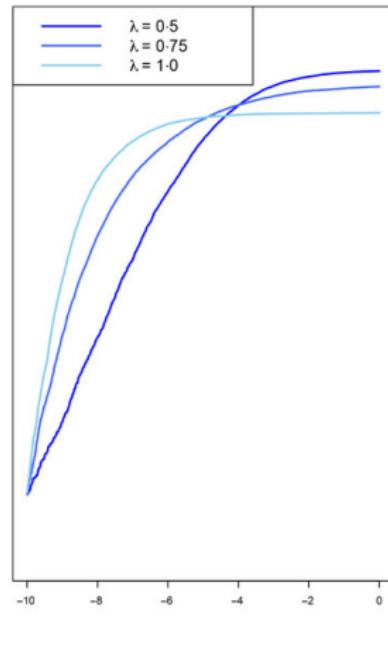
Birth-death



Time-dependent

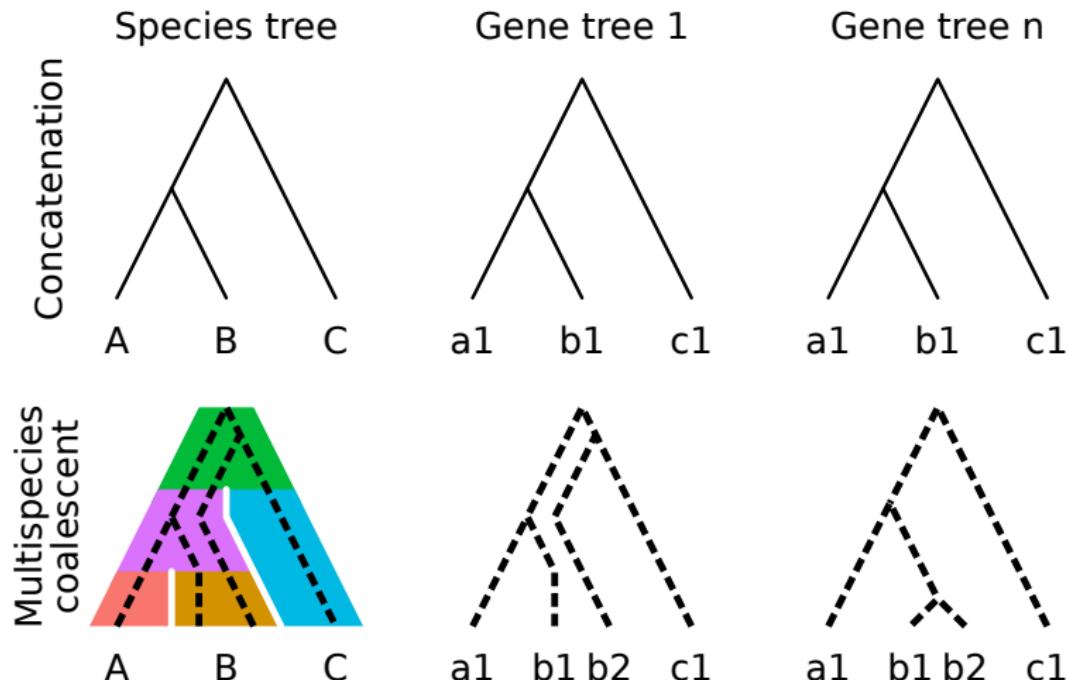


Diversity-dependent

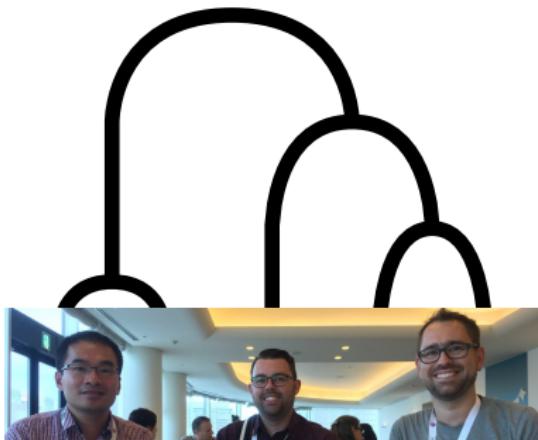


Janzen *et al.* (2015, Methods in Ecology and Evolution)

Inferring time trees from molecular data

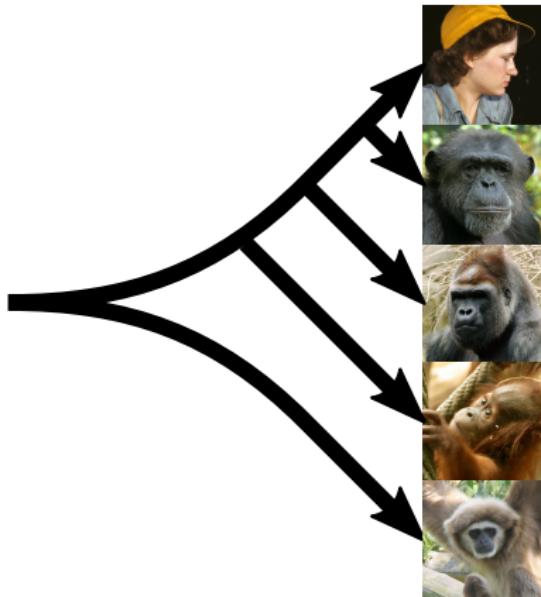


Coalescent processes



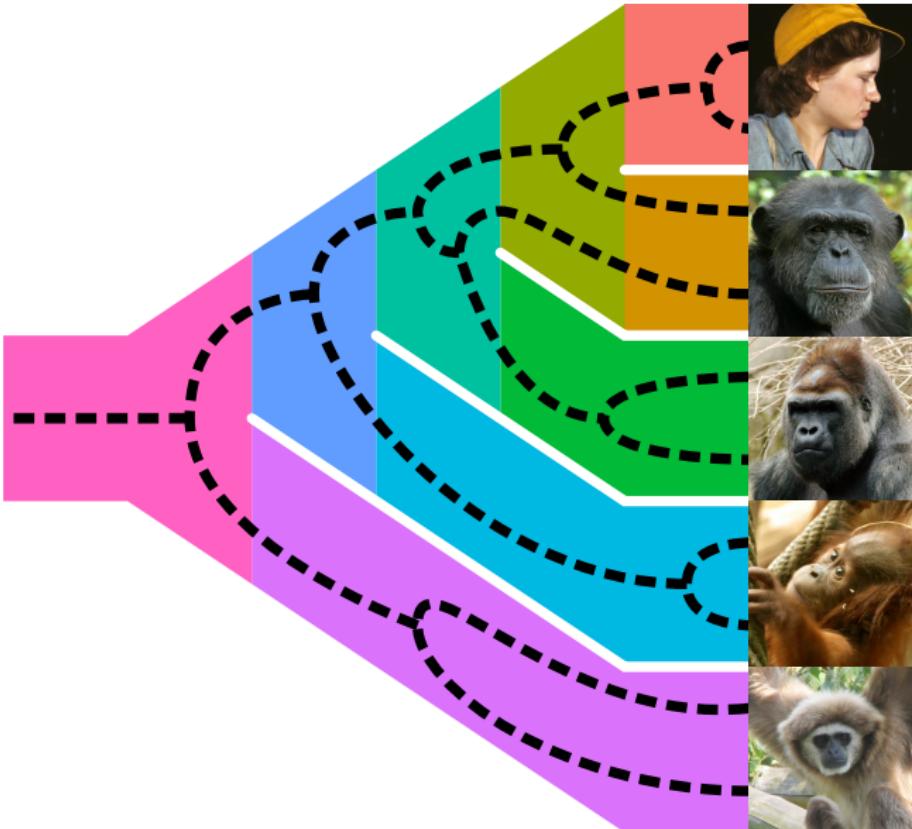
- Evolution of genes
- Dependent on effective population size N_e
- Constant, linear, exponential or stepwise (as in skyline plots)
- Backwards in time

Speciation processes

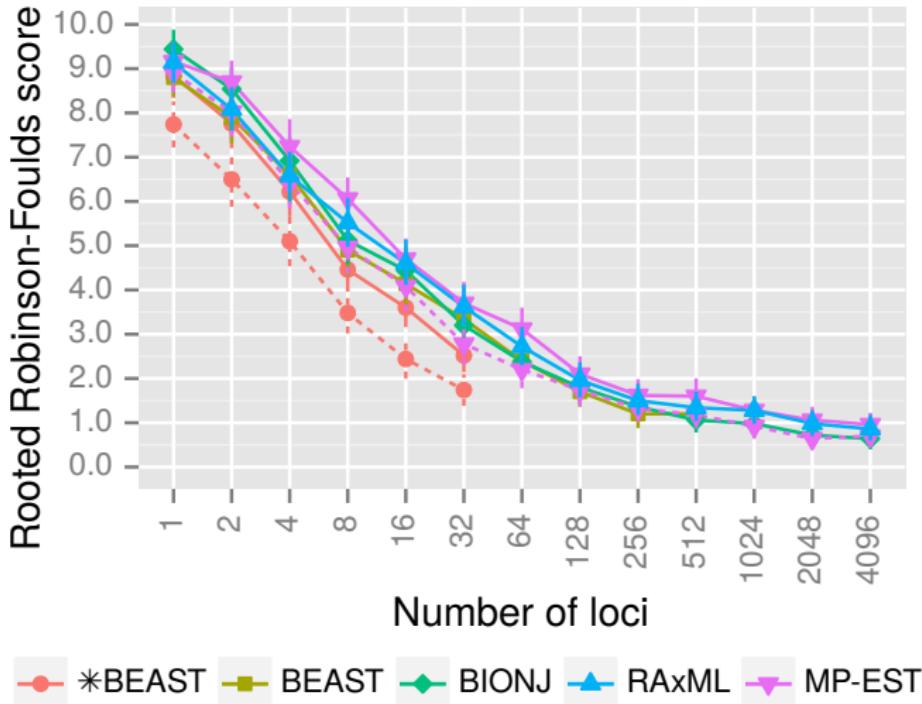


- Evolution of species (or populations)
- Birth-death process (speciation & extinction)
- Diversification rate
 $\text{Birth} - \text{Death} = \lambda$
- Extinction ratio
 $\text{Death} \div \text{Birth} = \nu$
- Forward in time

Putting them together

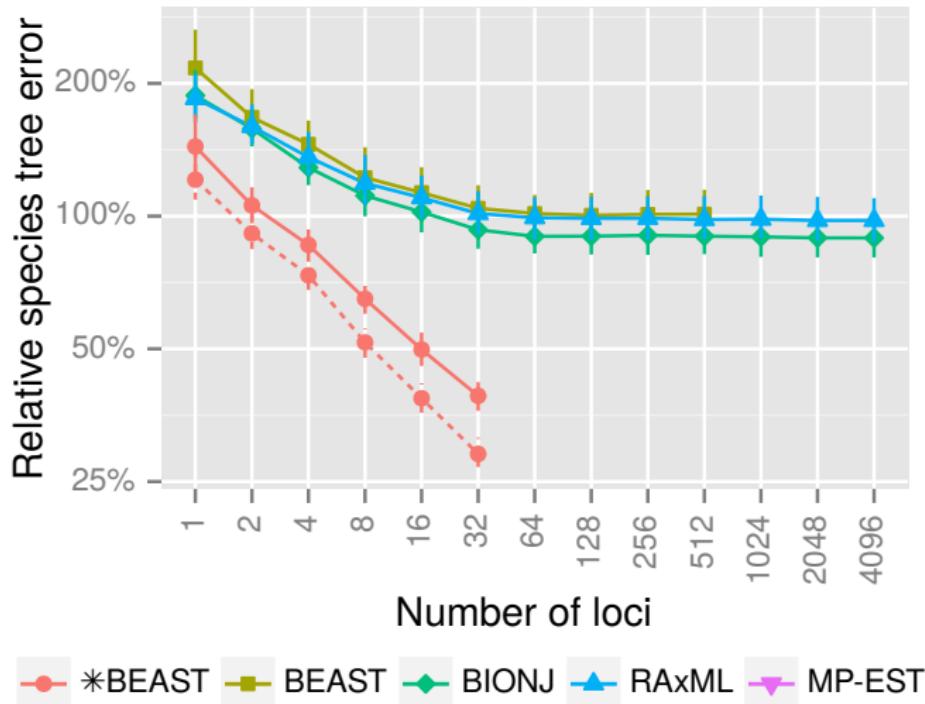


Topological accuracy



Ogilvie *et al.* (2016, Systematic Biology)

Branch length (or divergence time) accuracy

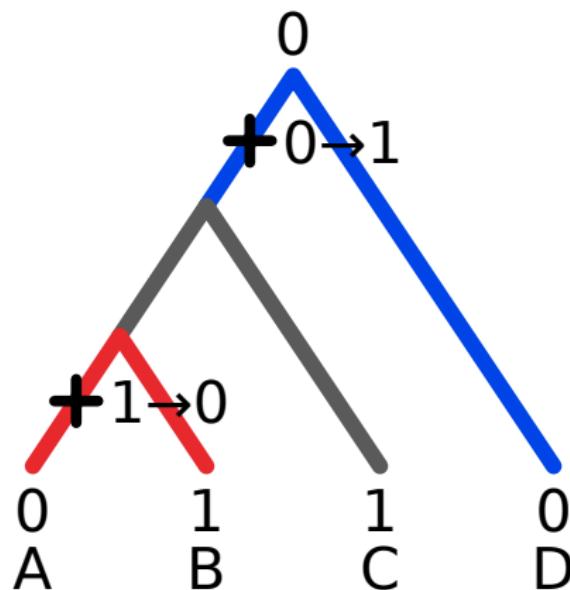
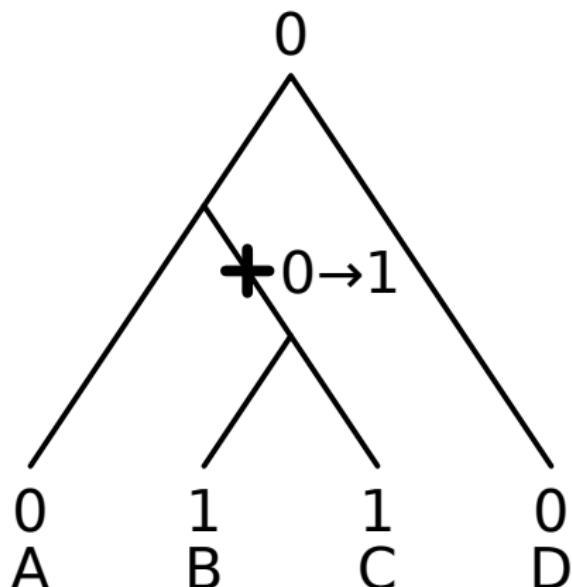


Ogilvie *et al.* (2016, Systematic Biology)

Why? SPILS

“Substitutions Produced by Incomplete Lineage Sorting”

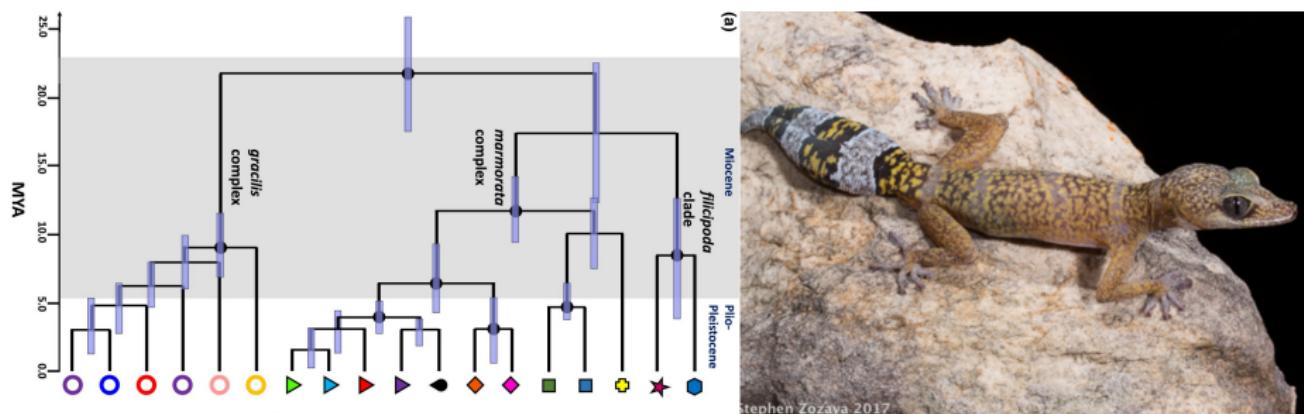
First reported by Mendes & Hahn (2016) in Systematic Biology



Absolute time trees

Branch lengths and node heights in years (often millions or thousands).

Useful for placing the evolution of clades of species in a geological context. For example this clade appears to have diversified largely after widespread cooling and aridification in the late-Miocene.

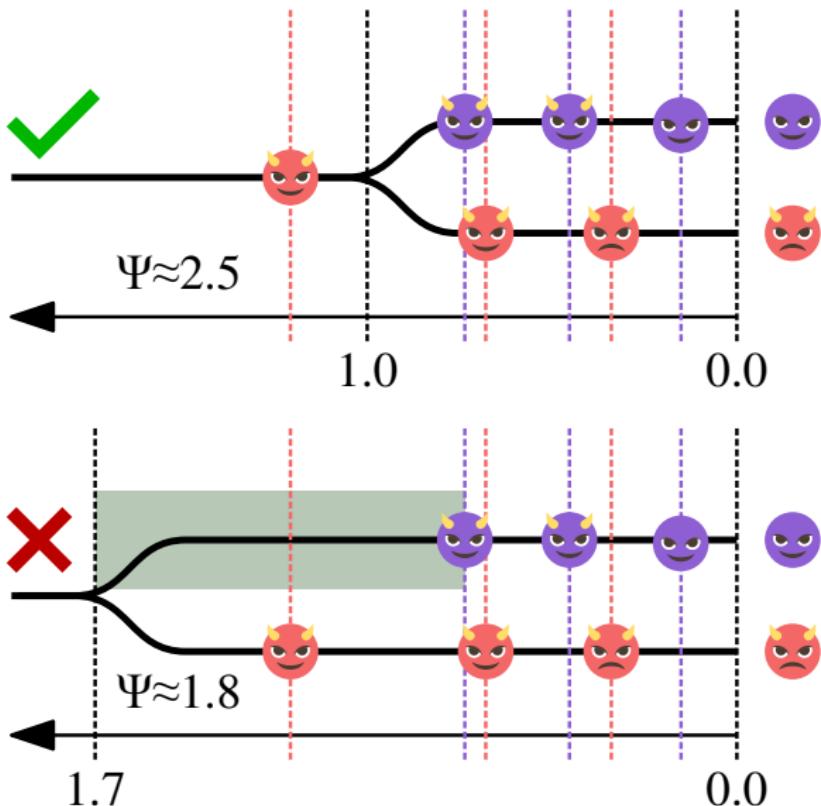


Laver *et al.* (2018, Journal of Biogeography)

Calibrations for absolute time trees

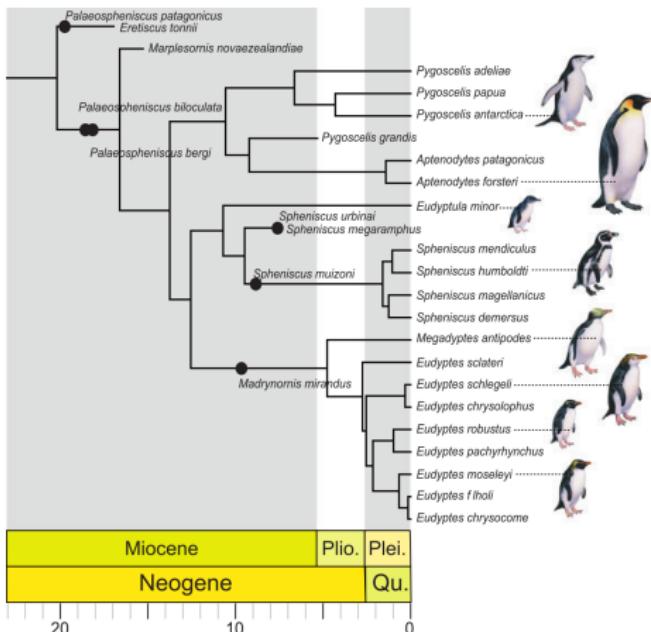
- Fixed clock
 - Set the clock rate of one or more loci to a “known” rate, where available for that clade and locus
- Node dating
 - Specify a prior probability distribution for the age of one or more splits, designed by intuition and based on expert knowledge
- Tip dating
 - Models the process of fossilization to automatically and transparently derive prior probabilities on split times
 - Heath, Huelsenbeck and Stadler (2014) “The fossilized birth–death process for coherent calibration of divergence-time estimates”
 - Gavryushkina, Welch, Stadler and Drummond (2014) “Bayesian Inference of Sampled Ancestor Trees for Epidemiology and Fossil Calibration”

Fossilized birth–death model



FBD–concatenation

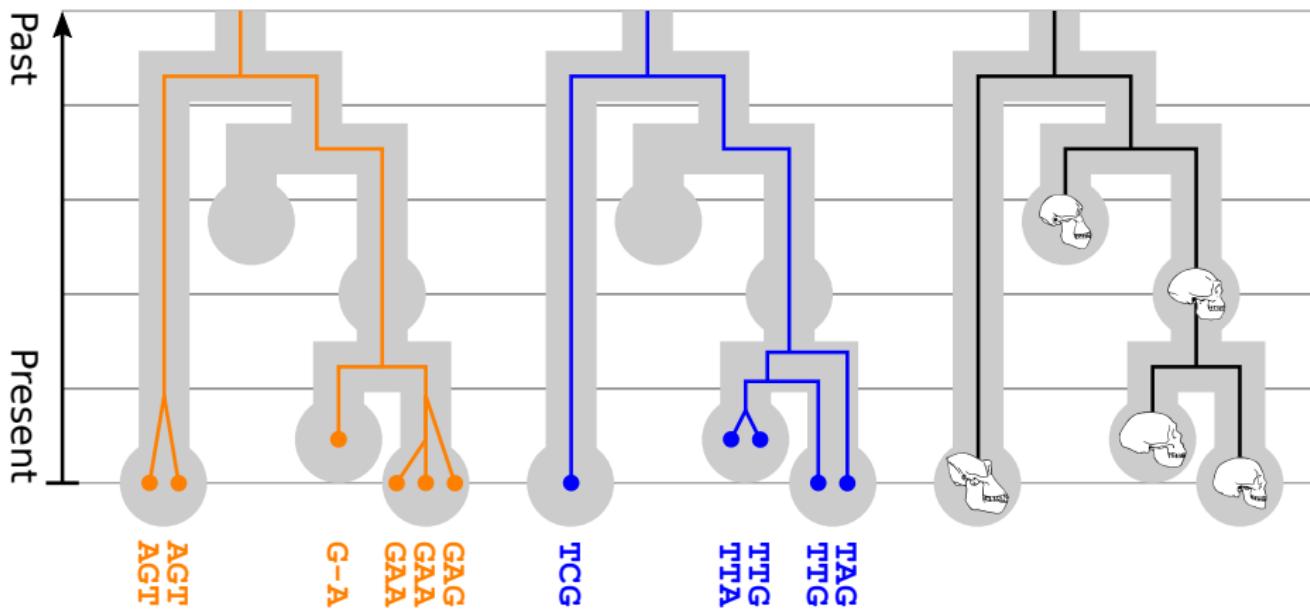
Often called “total evidence,” the same tree is used for all morphological and molecular character matrices.



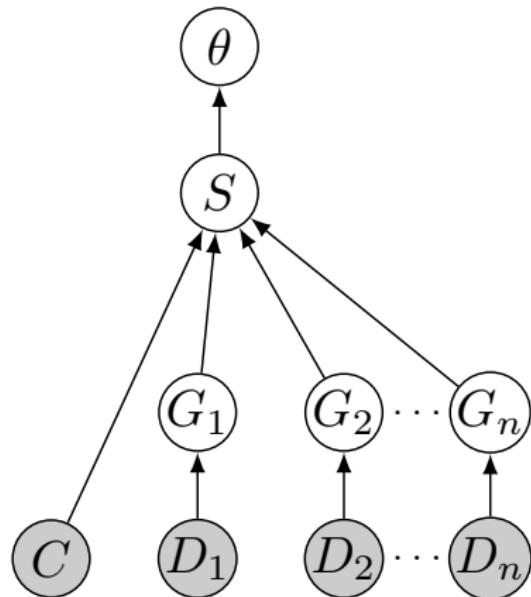
Gavryushkina *et al.* (2017, Systematic Biology)

FBD–MSC model

Separate gene trees for each locus, embedded within the species tree, morphological characters evolve along the species tree.



FBD–MSC as a graphical model



θ : macroevolutionary rates, S : species tree, G_i : gene tree i , D_i : multiple sequence alignment i , C : morphological character matrix

Available in StarBEAST2

BEAST 2 Package Manager			
Name	Install...	Late...	Dependencies
CA		1.3.0	
CodonSubstModels		1.1.0	
DENIM		0.3.1	
EpilnF		7.0.2	SA
GEO_SPHERE		1.2.0	BEASTLabs
Mascot	1.1.1	1.1.1	
MASTER		6.0.0	
MGSMS		0.3.0	
MM	1.1.0	1.1.1	
MODEL_SELECTION		1.4.1	BEASTLabs
MultiTypeTree		7.0.0	
NS		1.0.1	MODEL_SELECTION, BEASTLabs
PhyDyn		1.3.0	
phydynamics		1.3.0	BDSKY
PoMo		1.0.1	
RBS		1.4.0	
SA	2.0.0	2.0.0	BEASTLabs
SCOTTI		2.0.1	
SNAPP		1.4.1	
SpeciesNetwork	0.12.2	0.12.2	
SSM		1.1.0	
STACEY		1.2.4	
StarBEAST2	0.15.1	0.15.1	SA, MM
substBMA		1.2.2	
TMA		1.0.0	phydynamics, TreeStat2, MASTER, BEASTLabs, BDSKY
TreeStat2		0.0.2	
<input checked="" type="checkbox"/> Latest	Install/Upgrade	Uninstall	Package repositories
			Close ?

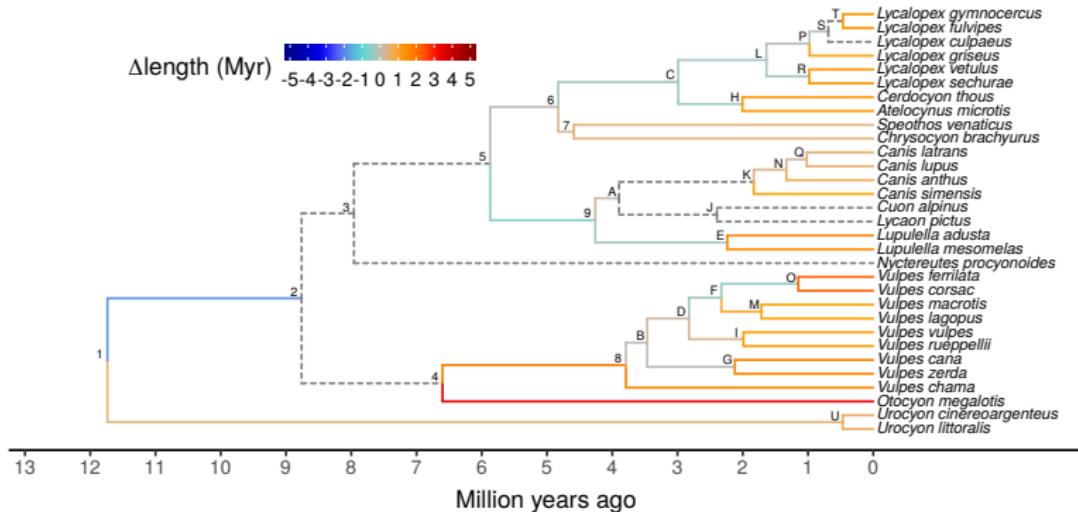
Ogilvie *et al.* (2017, Molecular Biology and Evolution)

Caninae

- Used to test the performance and behaviour of the models
- Family includes extant dogs, wolves and foxes
- Many species across the Americas, Eurasia and Africa
- North American radiation and South American radiation
- Reused morphological matrix from Zrzavý (2018, *Zoological Scripta*)
- Compiled 58 nuclear loci from several sources
- 36 extant and 42 fossil taxa

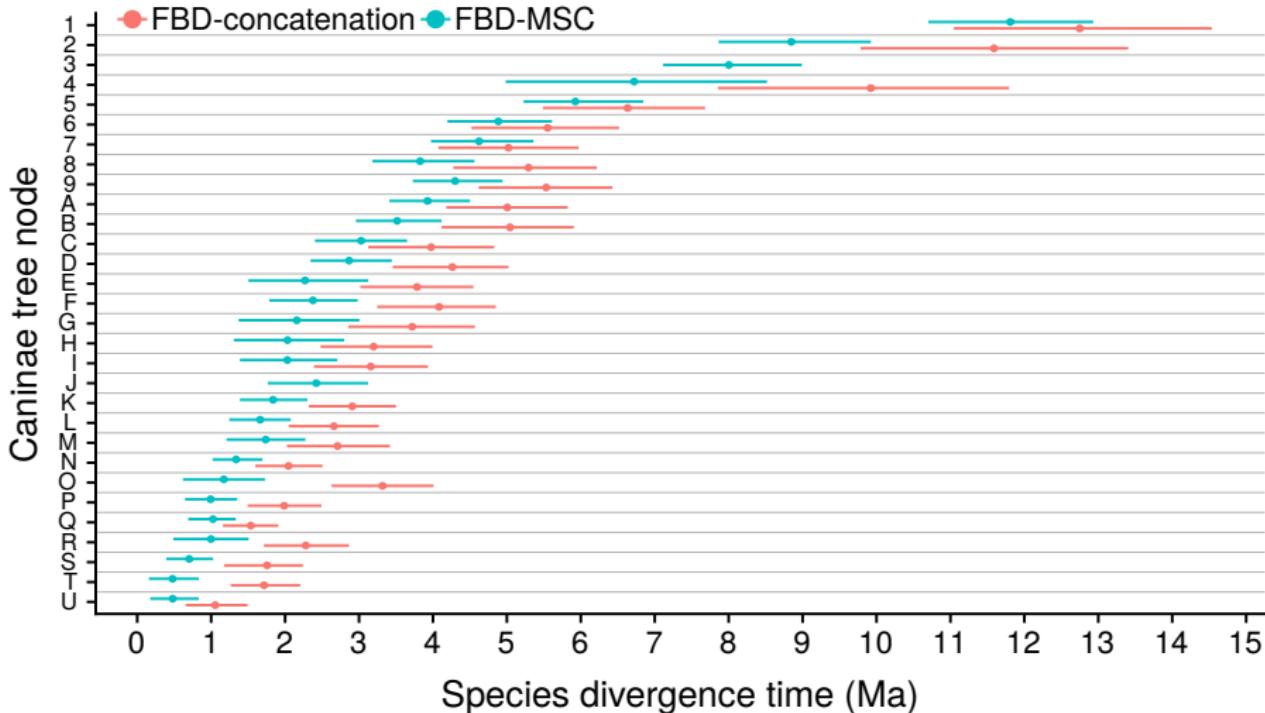
The importance of using FBD–MSC

Branch lengths estimated by FBD–concatenation were more often longer (warm colours) than FBD–MSC, less often shorter (cool colours).

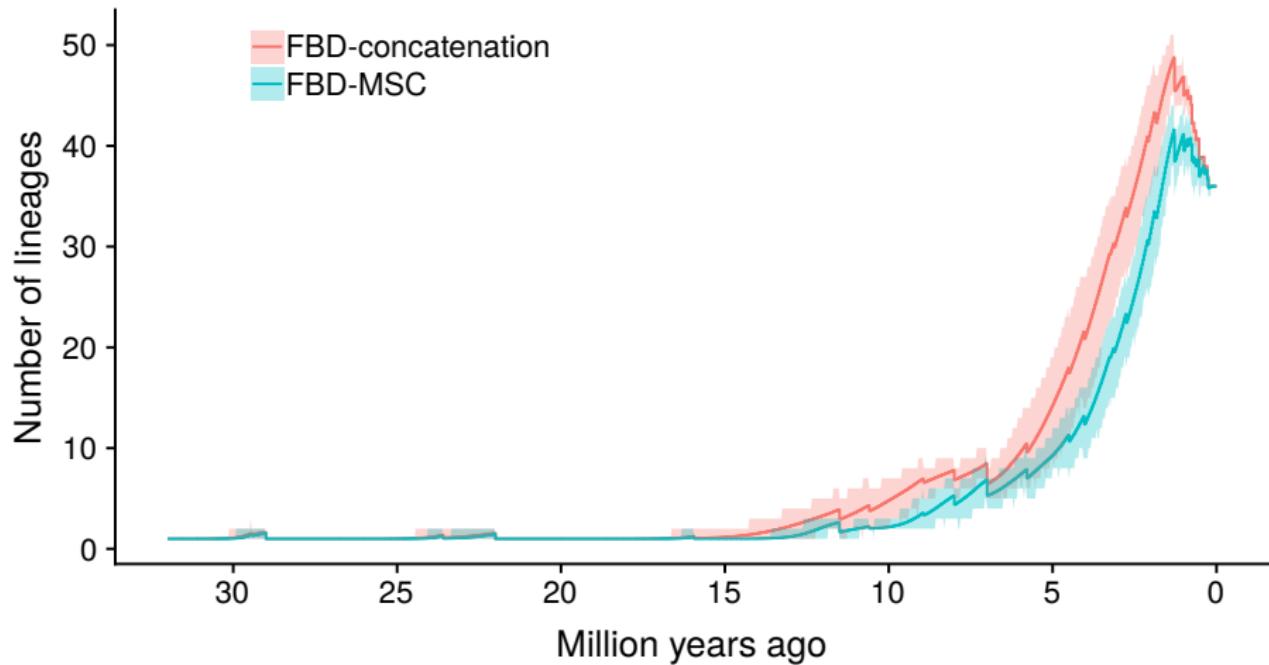


Used posterior predictive simulations to confirm differences are expected as a consequence of model misspecification (concatenation).

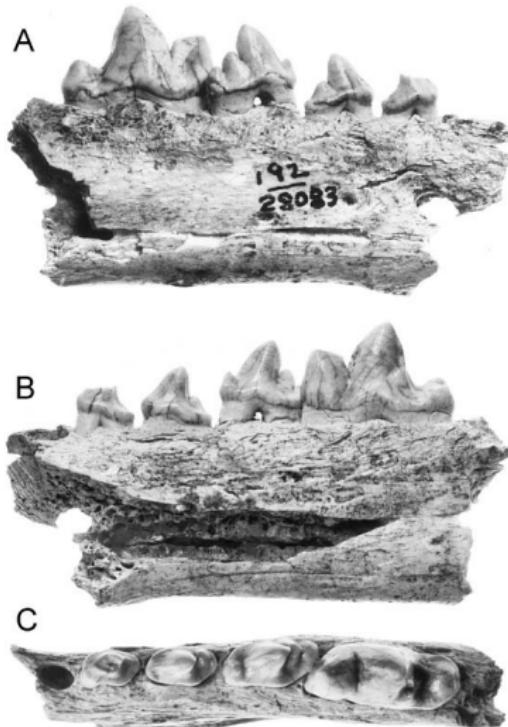
Estimating divergence times with FBD–MSC



Estimating lineages-through-time



A word of caution



- *Cuon alpinus* teeth
- Tedford (2009, Bulletin of the AMNH)
- Morphologically similar to *Lycaon pictus*
- Unlike other Caninae, both are hypercarnivores
- *Cuon*–*Lycaon* clade inferred by FBD–MSC unsupported by molecular data or FBD–concatenation
- Perhaps convergent evolution can mislead the FBD–MSC model?

Hypothesis

When morphological and molecular data conflict:

Using FBD–concatenation, the molecular will “win” because there are usually many more molecular characters

Using FBD–MSC, the morphological data will “win” because the population size parameter can be increased

Ongoing work

1. Publish Ogilvie *et al.* (2018, bioRxiv)
2. Morphological models which account for convergent evolution (with Chi Zhang)

Co-authors, acknowledgements and thanks

University of Auckland:

- **Alexei Drummond**
- **Nick Matzke**
- **David Welch**
- Remco Bouckaert

ETH Zürich:

- **Tanja Stadler**
- **Tim Vaughan**
- Alexandra Gavryushkina
(now at Otago)
- Alex Gavryushkin
(now at Otago)
- Nicola Müller
- Chi Zhang (now at IVPP)

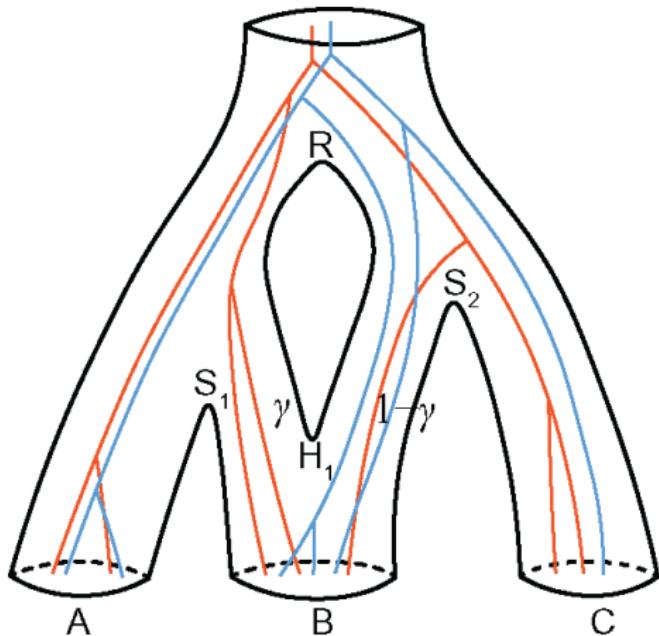
ANU:

- Craig Moritz
- Mozes Blom (now at Museum für Naturkunde Berlin)
- Ian Brennan
- Damien Esquerre
- Jessica Fenker Antunes
- Bec Laver
- Josh Peñalba (now at LMU)
- Ana Afonso Silva
(now at ENS)

University of Chicago:

- **Graham Slater**

Multispecies network coalescent inference



Zhang, Ogilvie, Drummond and Stadler (2018) MBE

Divide-and-Conquer MSNC inference

Bioinformatics

doi.10.1093/bioinformatics/xxxxxx

Advance Access Publication Date: Day Month Year

Manuscript Category

OXFORD

Subject Section

A Divide-and-Conquer Method for Scalable Phylogenetic Network Inference from Multi-locus Data

Jiafan Zhu¹, Xinhao Liu¹, Huw A. Ogilvie¹ and Luay K. Nakhleh^{1,2,*}

¹Department of Computer Science, Rice University, Houston, TX 77005, USA and

²Department of BioSciences, Rice University, Houston, TX 77005, USA