

An integrative model of species evolution accounting for fossilization and coalescence

Huw A. Ogilvie

Department of Computer Science
Rice University, Houston TX

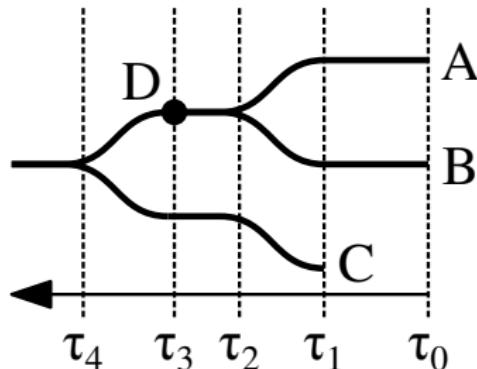
17 August 2023, Taming the BEAST eh!

Introduction

“FBD–MSC”

The **fossilized birth–death multispecies coalescent** is a model for inference of population or species histories with an absolute timescale.

Sampled ancestor trees

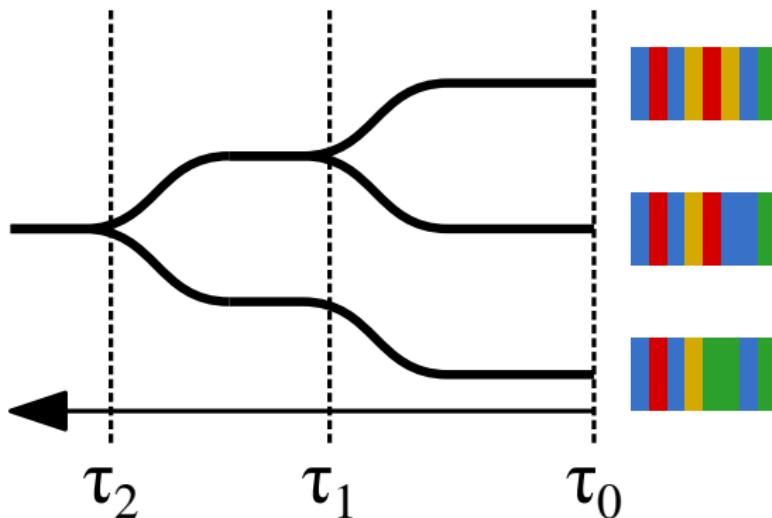


Observations can be:

- 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”)

Gene trees

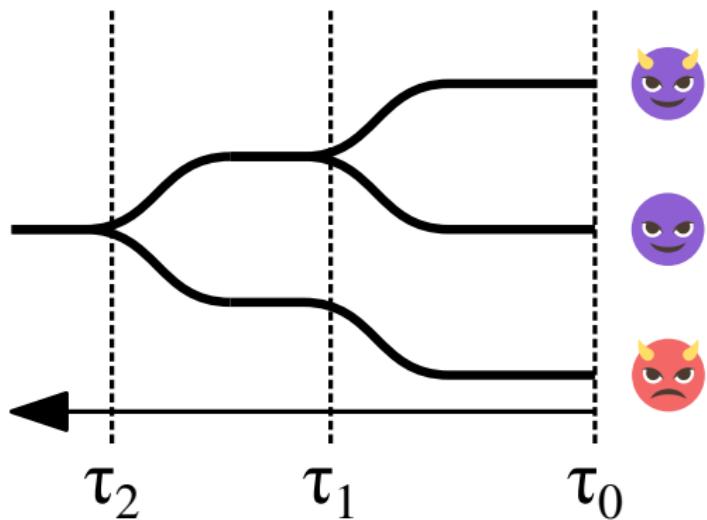
Time trees of genes represent a coalescent history



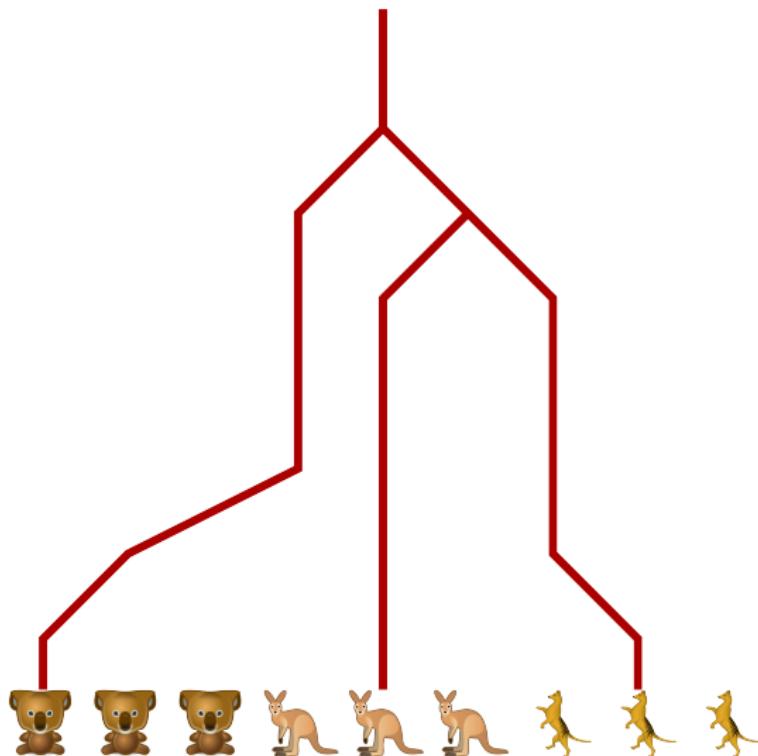
Observed sequences must share a common ancestor (the sequences are "homologous")

Species/population trees

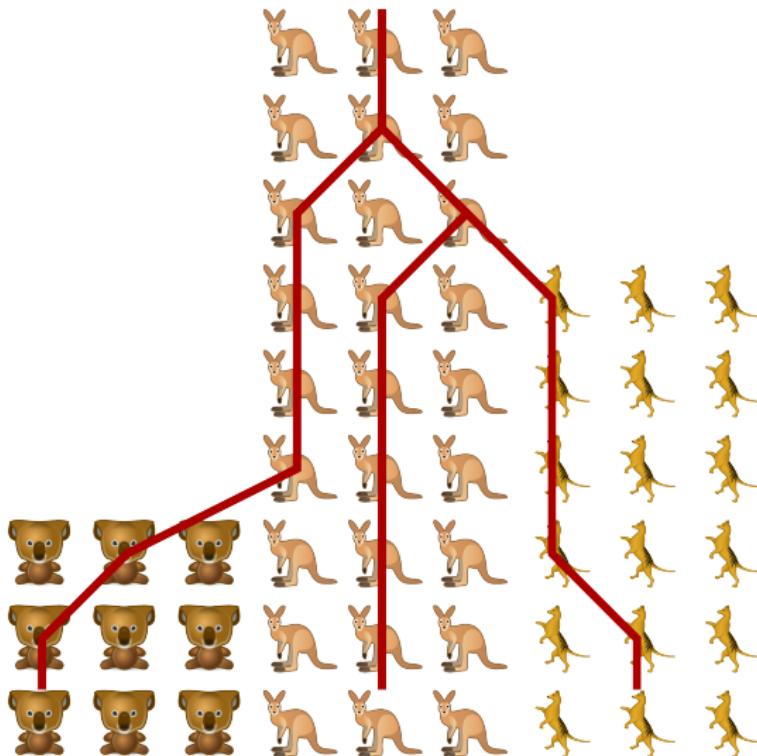
Time trees of species/populations represent the history of speciation/population structure



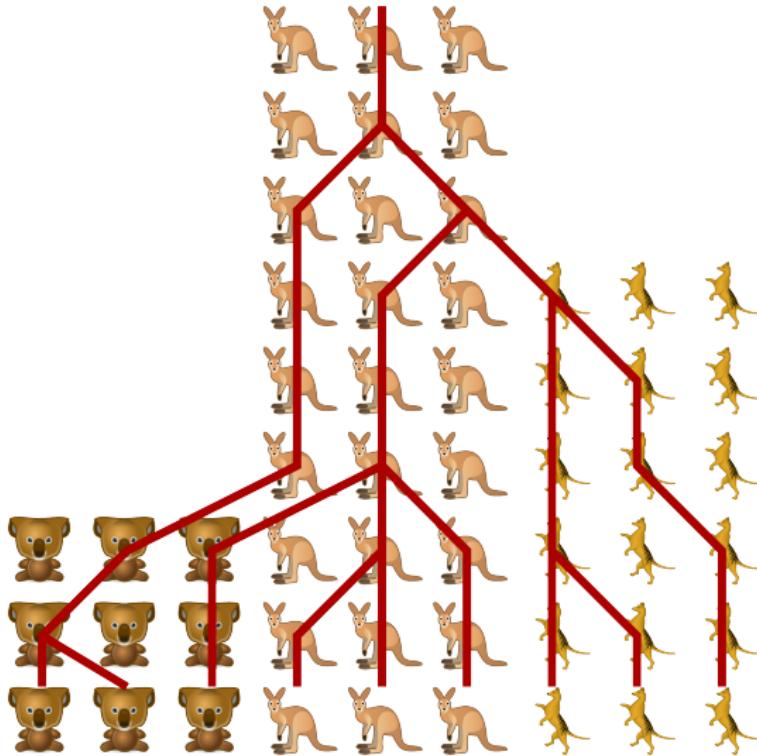
Host jumps example



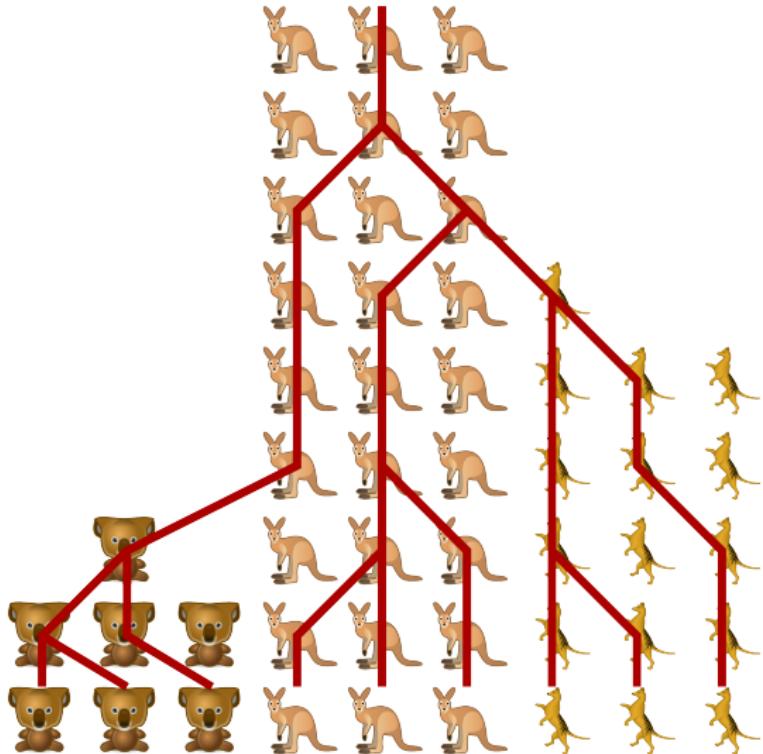
Incomplete lineage sorting



Multiple samples



Bottleneck

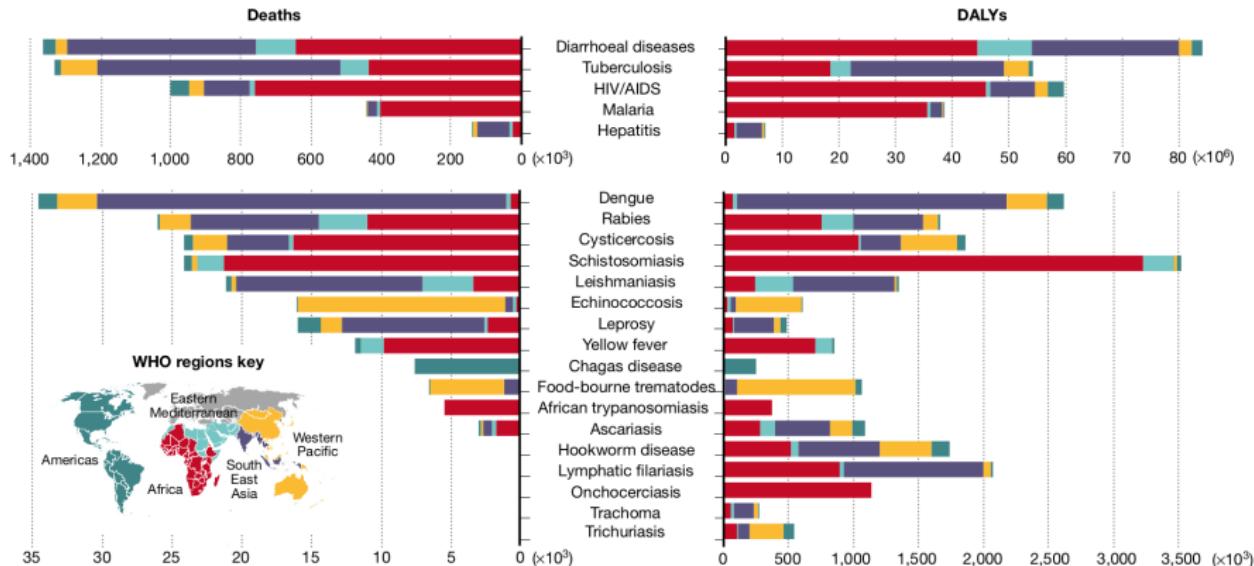


Multilocus methods for MSC inference

Each locus is an independent sample of coalescent times!

- Free recombination between loci
 - Sexual or parasexual reproduction, reassortment
- Random mating within a population
 - Outcrossing
- No intralocus recombination
 - Summary methods require longer loci for precision
 - Longer loci have more recombination; the “recombination ratchet”
 - Use fully Bayesian methods to avoid this

Infectious diseases?

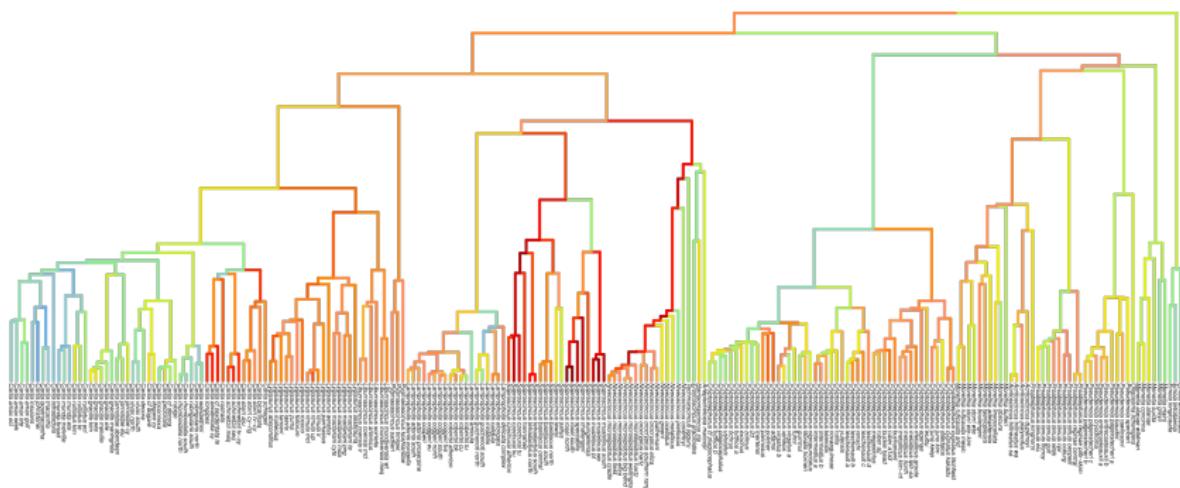


De Rycker et al. (2018) Challenges and recent progress in drug discovery for tropical diseases, Nature 559, pp. 498–506

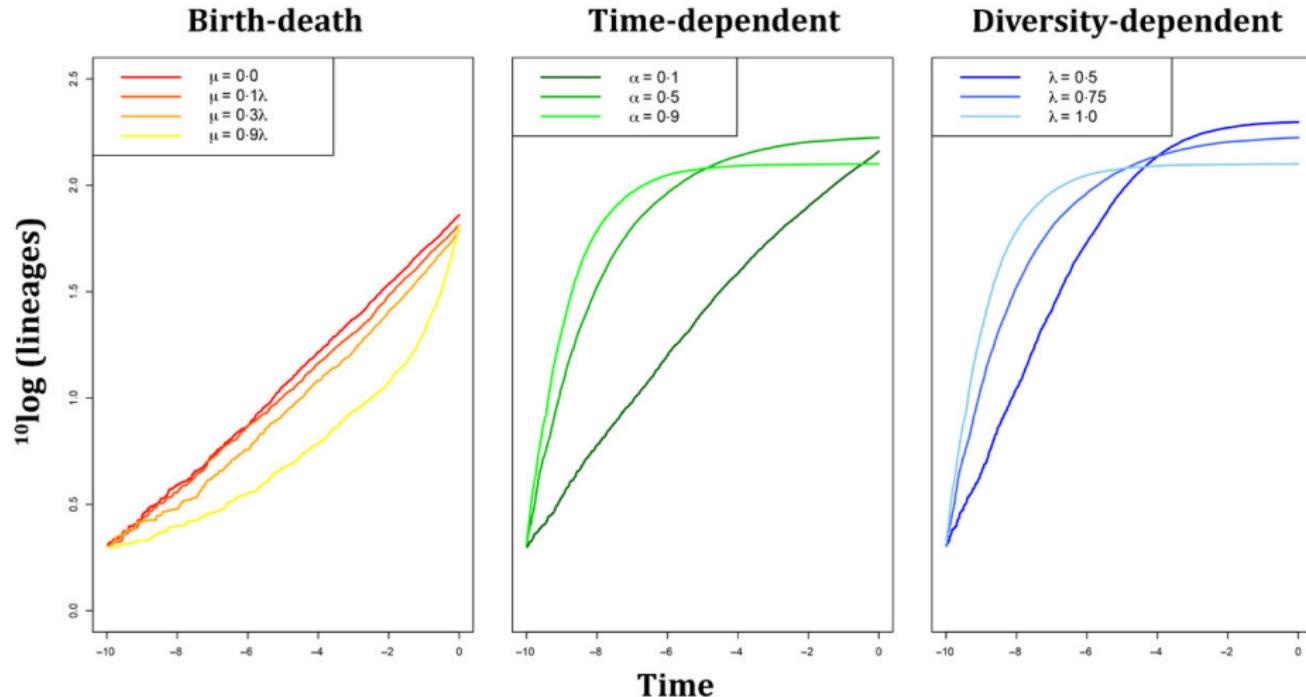
Relative time trees 1

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

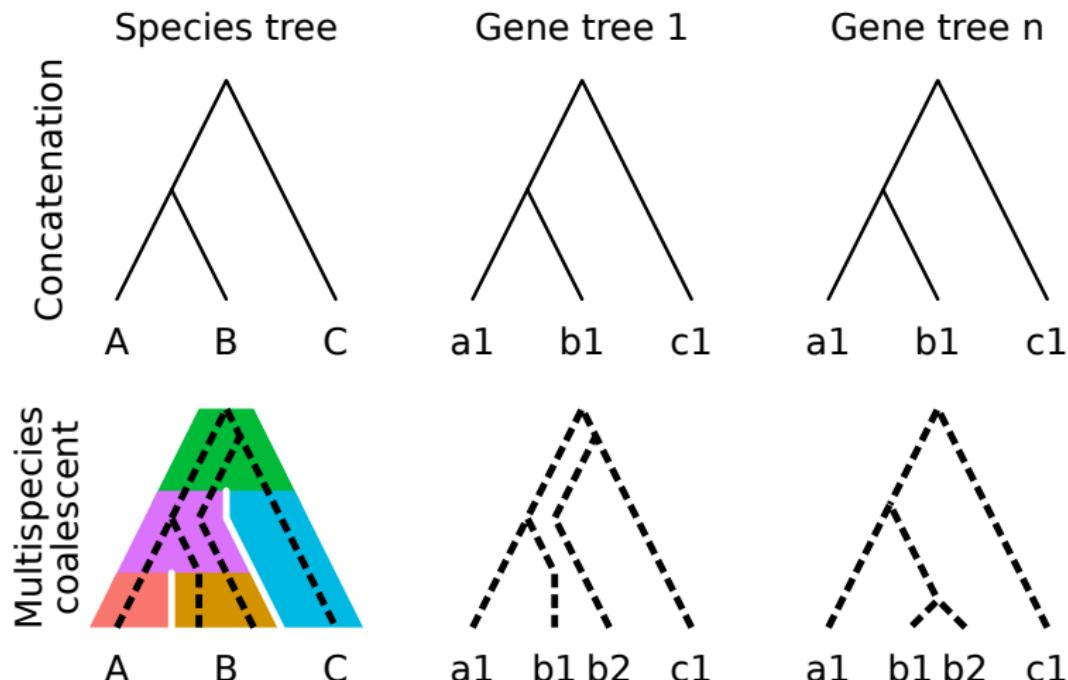


Relative time trees 2

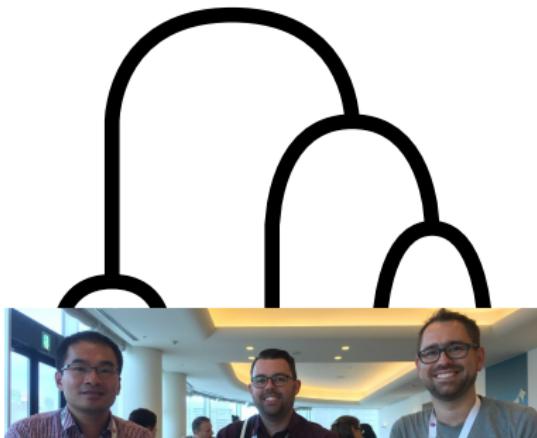


Janzen *et al.* (2015, Methods in Ecology and Evolution)

Inferring time trees from molecular data

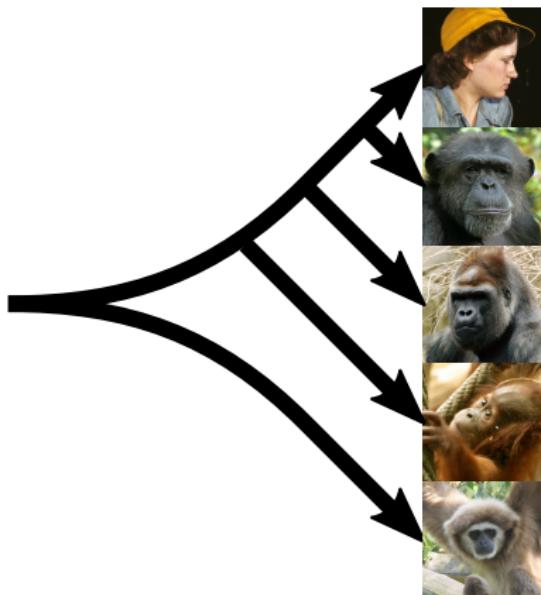


Coalescent processes



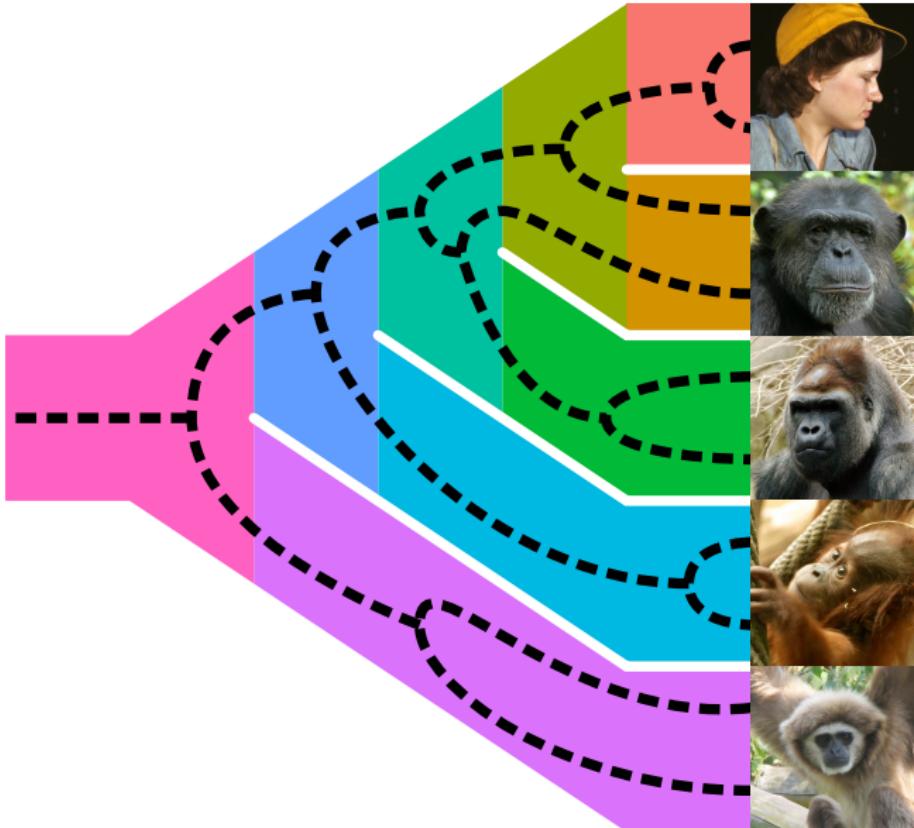
- Evolution of genes
- Dependent on effective population size N_e
- Constant, linear, exponential or stepwise (as in skyline plots)
- Backwards in time

Speciation processes

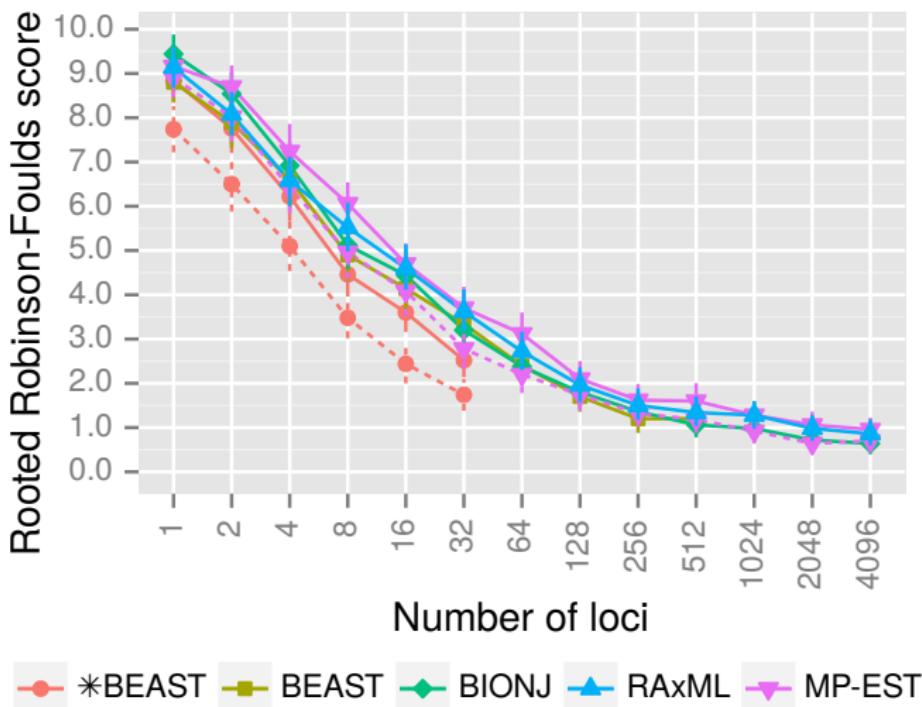


- Evolution of species (or populations)
- Birth-death process (speciation & extinction)
- Diversification rate
 $\text{Birth} - \text{Death} = \lambda$
- Extinction ratio
 $\text{Death} \div \text{Birth} = \nu$
- Forward in time

Putting them together

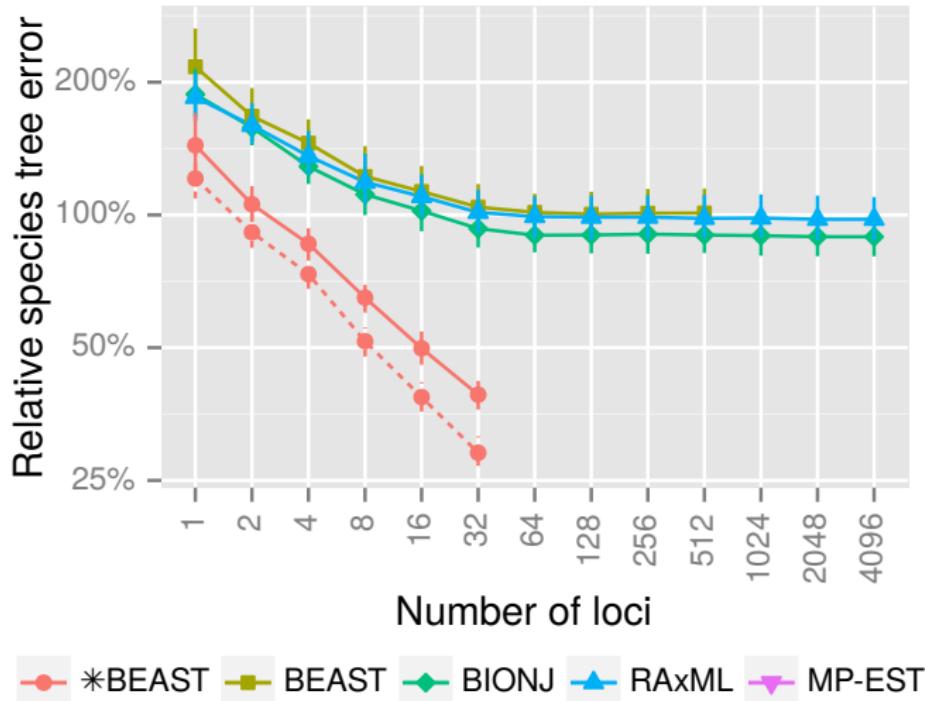


Topological accuracy



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

Branch length (or divergence time) accuracy

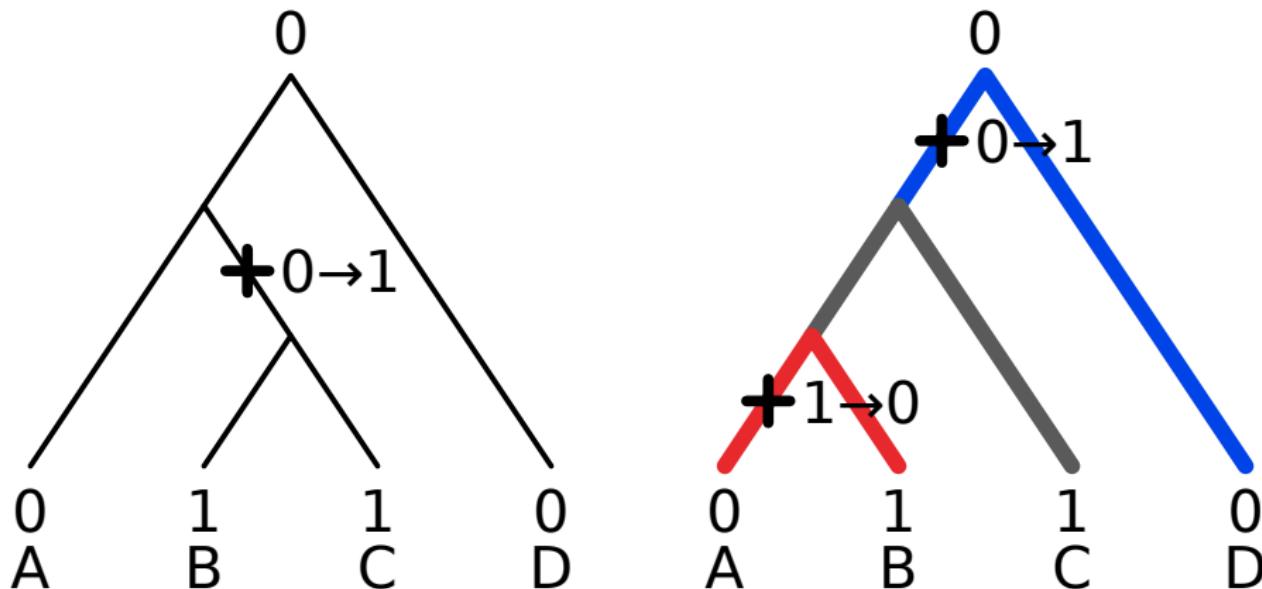


Ogilvie *et al.* (2016, Systematic Biology)

Why? SPILS

"Substitutions Produced by Incomplete Lineage Sorting"

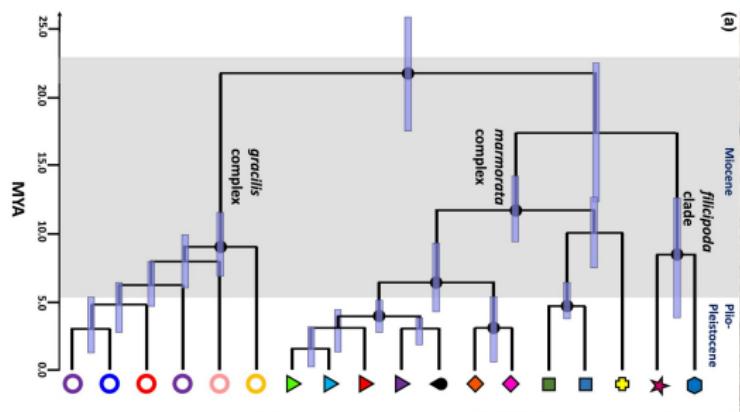
First reported by Mendes & Hahn (2016) in Systematic Biology



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.

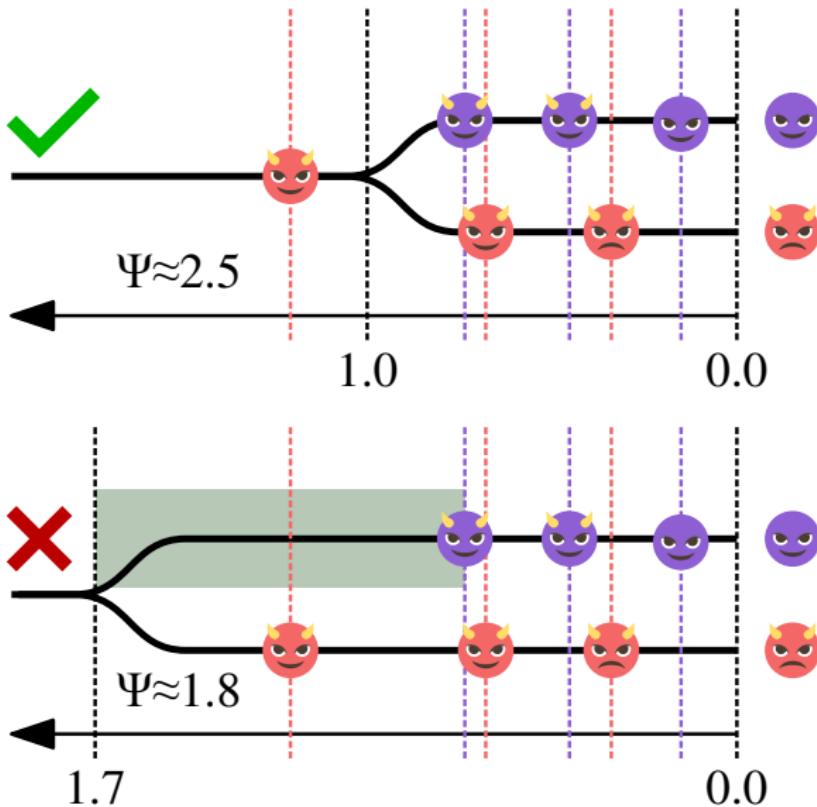


Laver *et al.* (2018, Journal of Biogeography)

Calibrations for absolute time trees

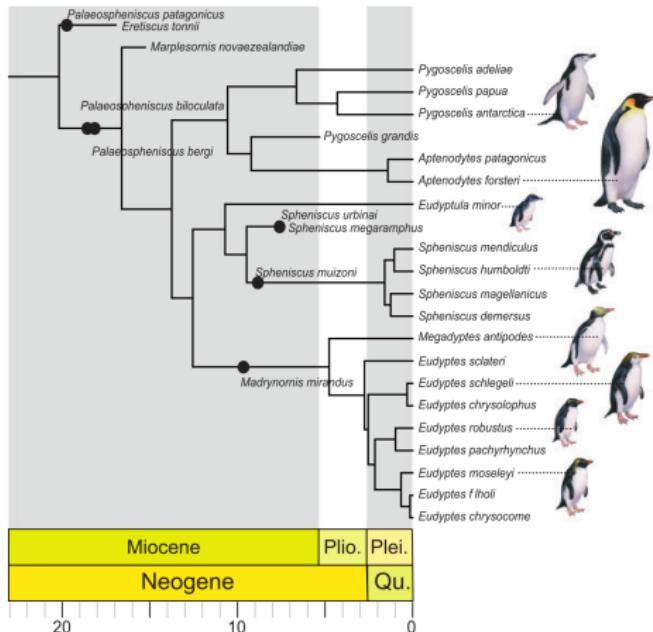
- Fixed clock
 - Set the clock rate of one or more loci to a “known” rate, where available for that clade and locus
- Node dating
 - Specify a prior probability distribution for the age of one or more splits, designed by intuition and based on expert knowledge
- Tip dating
 - Models the process of fossilization 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

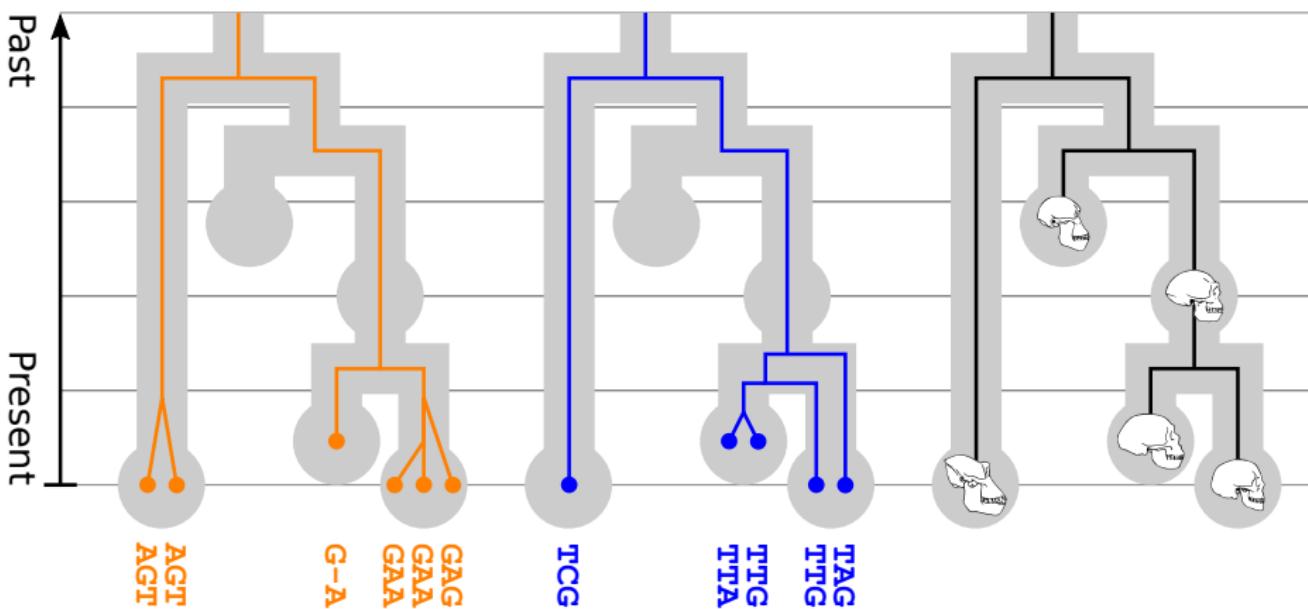
Often called “total evidence,” the same tree is used for all morphological and molecular character matrices.



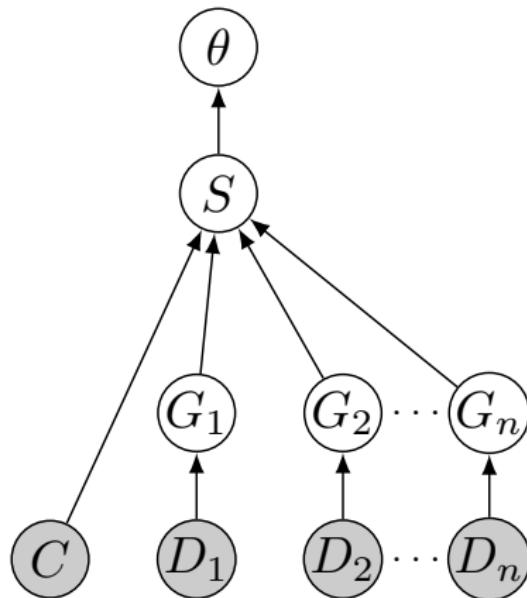
Gavryushkina et al. (2017, Systematic Biology)

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

List of available packages for BEAST v2.7.*

Name	Install...	Latest	Dependencies	Link	
MSBD		2.0.2	BEAST.app, MM, BEAST.base	Implementation of the MSBD multi-type birth-death tree	
MultiTypeTree		8.0.0	BEAST.base, BEAST.app	Structured coalescent inference.	
NS		1.2.1	BEAST.app, BEAST.base, MOD...	Nested sampling for model selection and posterior inference	
OBAMA		1.1.1	BEAST.app, bModelTest, BEA...	OBAMA for Bayesian Amino-acid Model Averaging	
ORC		1.1.2	1.1.2	BEAST.app, BEASTlabs, BEAS...	Optimised Relaxed Clock model
PhyDyn		1.4.0	BEAST.app, BEAST.base	PhyDyn: Epidemiological modelling with BEAST	
phyldynamics		1.4.0	BDSKY, BEAST.base	BDSIR and Stochastic Coalescent	
phylonco		1.0.1	BEASTlabs, BEAST.app, BEAS...	A package for single cell cancer evolution	
PIQME		1.0.0	BDSKY, BEAST.base, BEAST.app	Birth-death skyline-based method efficiently dealing with	
PoMo		1.1.1	BEAST.base, BEAST.app	PoMo, a substitution model that separates mutation and	
Recombination		1.0.1	BEAST.app, BEAST.base	Inference of Recombination networks	
SA		2.1.1	2.1.1	BEASTlabs, BEAST.base, BEA...	Sampled ancestor trees
SCOTTI		3.0.0	BEAST.base, BEAST.app	Structured COalescent Transmission Tree Inference	
SNAPP		1.6.1	BEAST.base, BEASTlabs, BEA...	SNP and AFLP Phylogenies	
snapper		1.1.3	BEAST.base, BEAST.app, SNAPP	Diffusion based SNP and AFLP Phylogenies	
SpeciesNetwork		1.0.0	BEAST.app, BEAST.base	Multispecies network coalescent (MSNC) inference of intr...	
speedemon		1.1.0	BICEPS, BEAST.app, BEASTLa...	Fast species delimitation under the multispecies coalesce...	
SSM		1.2.0	BEAST.app, BEAST.base	Standard Nucleotide Substitution Models	
STACEY		1.3.1	BEAST.app, BEAST.base	Species delimitation and species tree estimation	
StarBEAST2		1.0.0	BEAST.app, MM, BEAST.base, ...	Multispecies coalescent inference using multi-locus and f...	
starbeast3		1.1.7	1.1.7	BEAST.base, SA, BEAST.app, B...	StarBeast3 multispecies coalescent using advanced MCMC
timtam		0.4.0	BEAST.base, BEAST.app	Distribution for birth-death models with occurrence data	
TreeStat2		0.1.1	BEAST.base	Utility for calculating tree statistics from tree log file	

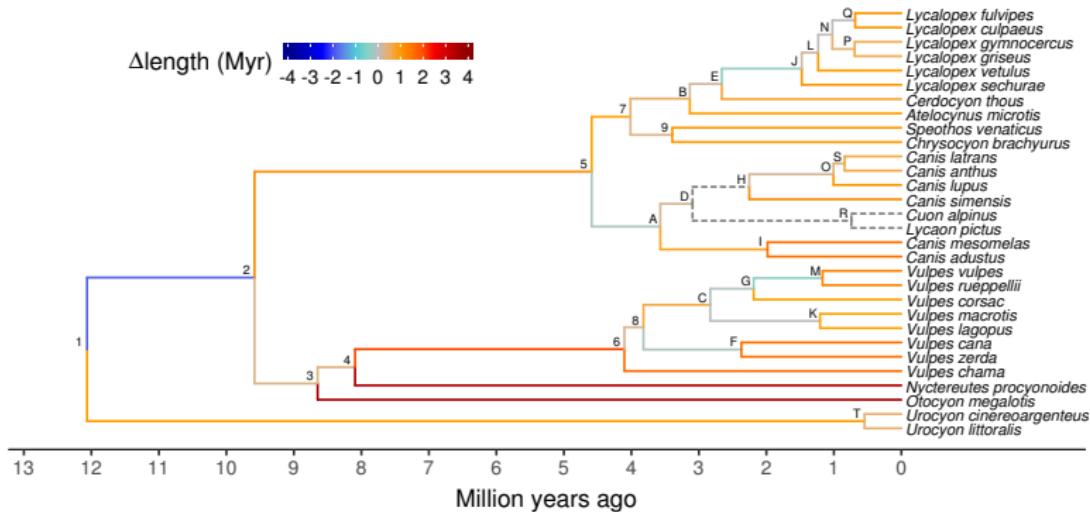
Latest [Install/Upgrade](#) [Uninstall](#) [Package repositories](#) [?](#) [Close](#)

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

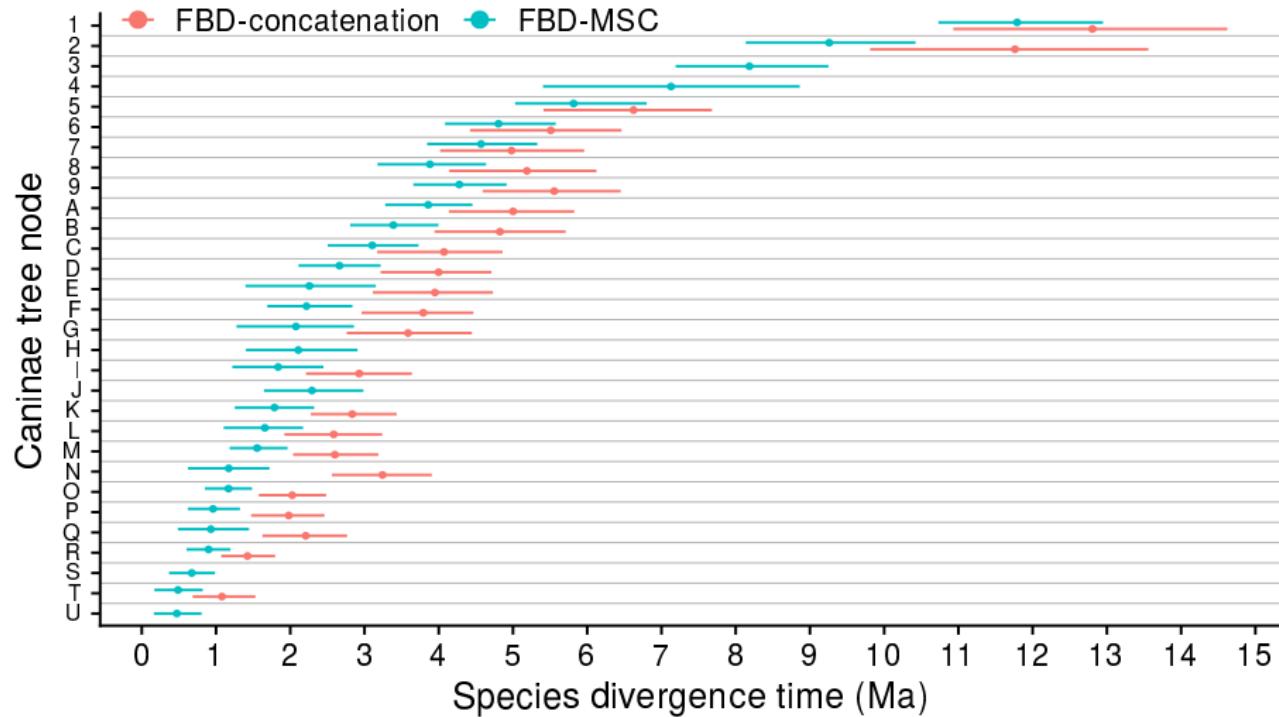
The importance of using FBD–MSC

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

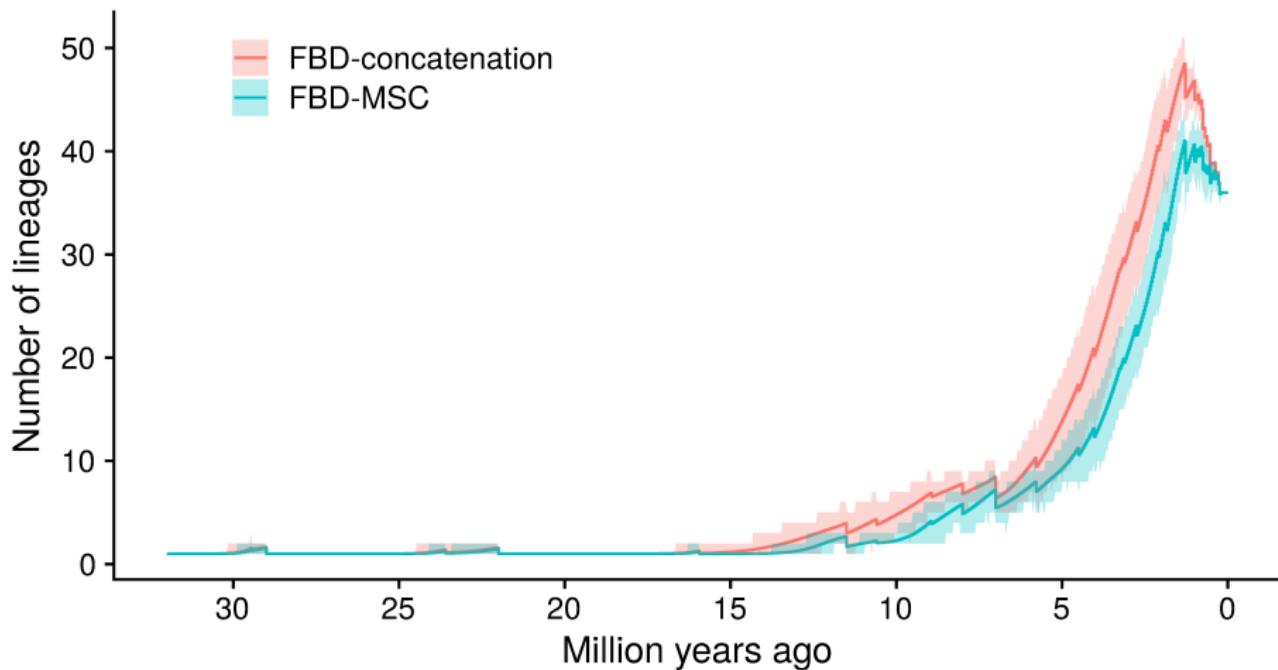


Used posterior predictive simulations to confirm differences are expected as a consequence of model misspecification (concatenation).

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

Read all about it

Novel Integrative Modeling of Molecules and Morphology across Evolutionary Timescales

HUW A. OGILVIE^{1,*}, FÁBIO K. MENDES^{2,4,*}, TIMOTHY G. VAUGHAN^{5,6}, NICHOLAS J. MATZKE^{2,4}, TANJA STADLER^{5,6}, DAVID WELCH^{2,3} AND ALEXEI J. DRUMMOND^{2,3,4}

¹Department of Computer Science - MS-132, Rice University, P.O. Box 1892, Houston, TX 77251-1892; ²Centre for Computational Evolution, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand; ³School of Computer Science, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand; ⁴School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand; ⁵Department of Biosystems Science and Engineering, ETH Zürich, Mattenstrasse 26, 4058 Basel, Switzerland; ⁶SIB Swiss Institute of Bioinformatics, Quartier Sorge - Bâtiment Amphipôle, Lausanne 1015, Switzerland

*Correspondence to be sent to: Huw A. Ogilvie, Department of Computer Science – MS-132, Rice University, P.O. Box 1892, Houston, TX 77251-1892, USA; E-mail: huw.a.ogilvie@rice.edu; and Fábio K. Mendes, School of Biological Sciences, The University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand; E-mail: f.mendes@auckland.ac.nz.
Huw A. Ogilvie and Fábio K. Mendes contributed equally to this article.

Received 19 March 2021; reviews returned 23 June 2021; accepted 29 June 2021
Associate Editor: Simon Ho

Abstract.—Evolutionary models account for either population- or species-level processes but usually not both. We introduce a new model, the FBD-MSC, which makes it possible for the first time to integrate both the genealogical and fossilization phenomena, by means of the multispecies coalescent (MSC) and the fossilized birth–death (FBD) processes. Using this model, we reconstruct the phylogeny representing all extant and many fossil Caninae, recovering both the relative and absolute time of speciation events. We quantify known inaccuracy issues with divergence time estimates using the popular strategy of concatenating molecular alignments and show that the FBD-MSC solves them. Our new integrative method and empirical results advance the paradigm and practice of probabilistic total evidence analyses in evolutionary biology.[Caninae; fossilized birth–death; molecular clock; multispecies coalescent; phylogenetics; species trees.]

Co-authors, acknowledgements and thanks

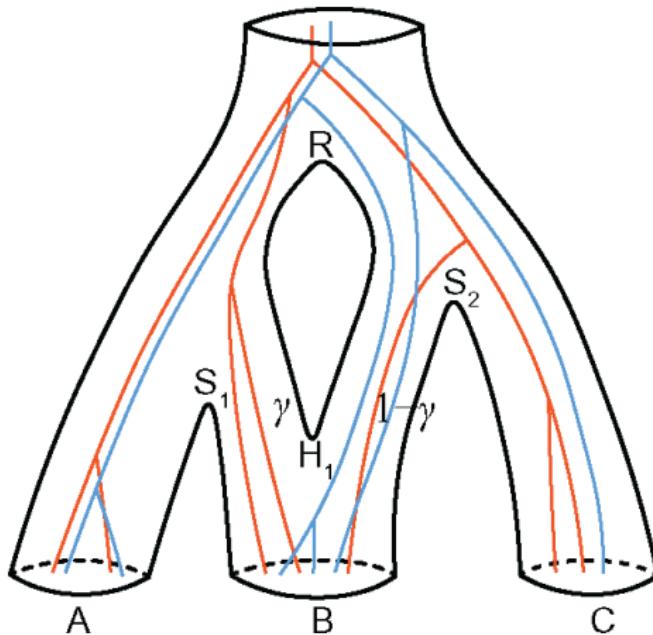
University of Auckland:

- **Alexei Drummond**
- **Nick Matzke**
- **Fábio Mendes**
(now at WUSTL, starting his own lab at LSU next year)
- **David Welch**
- Remco Bouckaert

ETH Zürich:

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

Multispecies network coalescent inference



Zhang, Ogilvie, Drummond and Stadler (2018) MBE

Cancer phylogenetics

CRC04

