

Over 10 cool things to do with BEAST

Remco R. Bouckaert
remco@cs.auckland.ac.nz

Centre for Computational Evolution, University of Auckland
Max Planck Institute for the Science of Human History
Computer Science, Waikato University

Melbourne, 2017

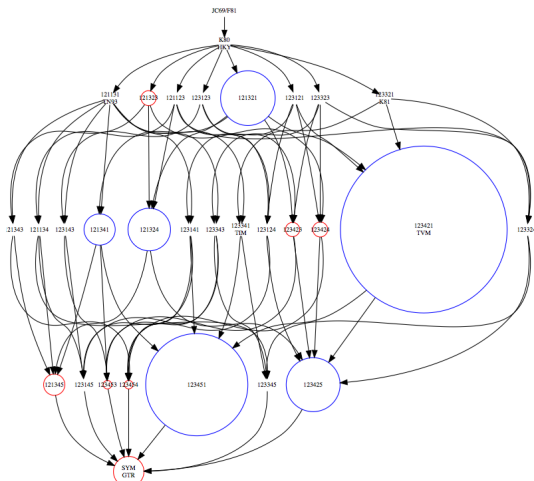


1. Which site model should I use?

bModelTest: Bayesian model averaging over;

- Substitution model (HKY, GTR, etc.)
- Rate parameters
- Estimated/equal frequencies
- With/without gamma rate heterogeneity
- With/without proportion invariant sites

HCV model distribution

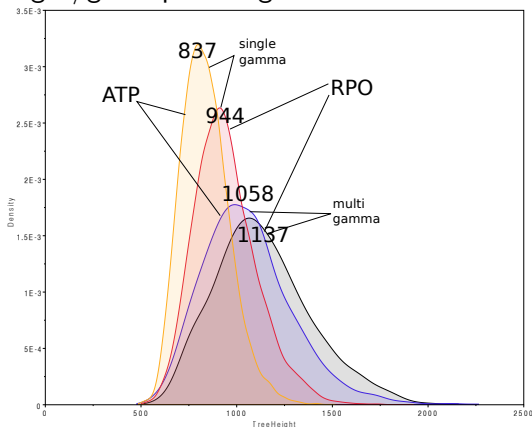


Bouckaert & Drummond, BMC evolutionary biology, 2017, [package: bModelTest](#)

2. What to do with heterotachy?

- Single Gamma site model (Yang)
- Multi Gamma Site Model: one α per branch
- Relaxed Gamma Site model: sample α from (log normal) distribution
- Much better fit for many models
- Significantly different root ages

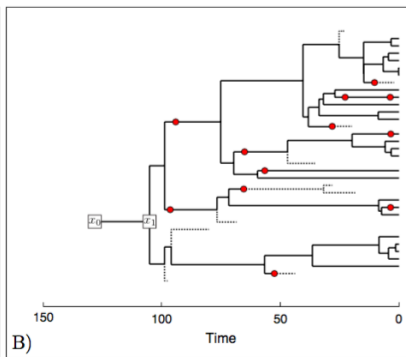
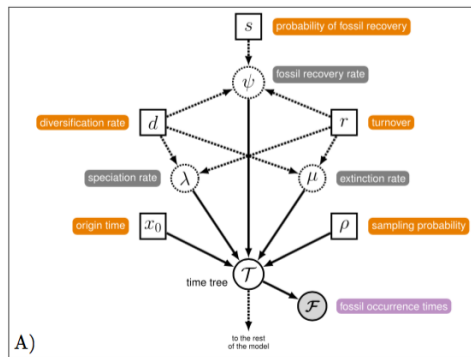
Algae/green plants age



Bouckaert & Lockhart, bioRxiv, 2015, [package: MGSM](#)

3. How to represent fossil calibrations?

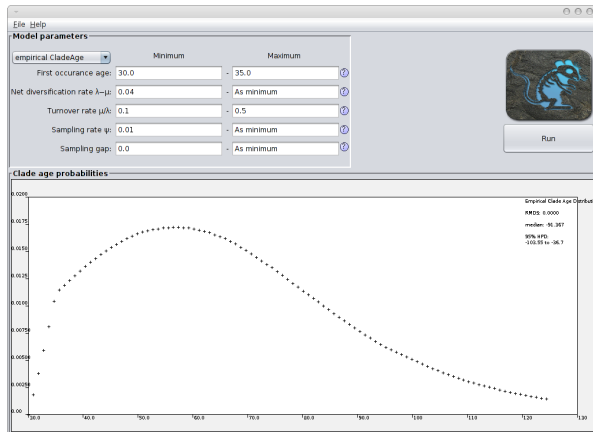
Fossilised Birth Death



Gavryushkina et al, PloS Comp Bio, 2014, [package: SA](#)

4. How to represent fossil calibrations?

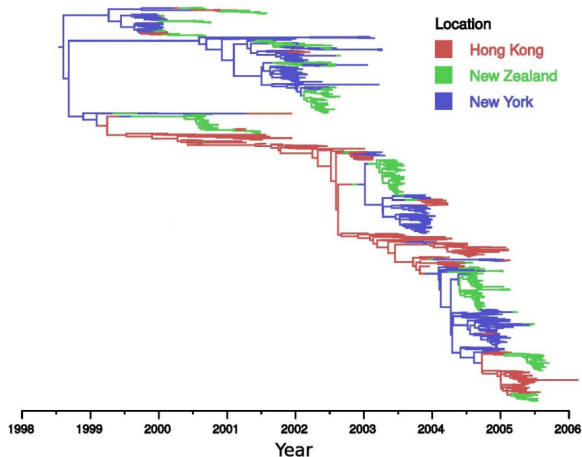
CladeAge



Matschinger et al, Sys Bio, 2017, [package: CA](#)

5. How many states can I have in a structured coalescent ?

- MultiTypeTree, up to 4 demes
- Bayesian Structured Coalescent Approximation: BASTA
- up to 11 demes?
- ongoing work to increase nr of demes



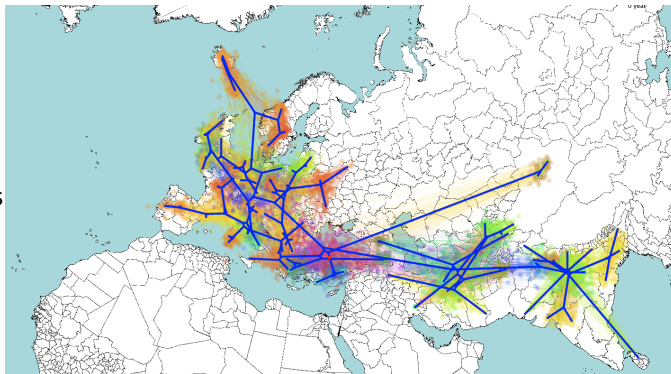
Vaughan et al, Bioinformatics, 2014, [package: MultiTypeTree](#)

De Maio et al, PLoS Genet. 2015, [package: BASTA](#)

6. How to reconstruct geography?

geo-sphere

- diffusion on a sphere
- geographical priors on root, tips, clades
- landscape aware: somewhat inhomogeneous diffusion

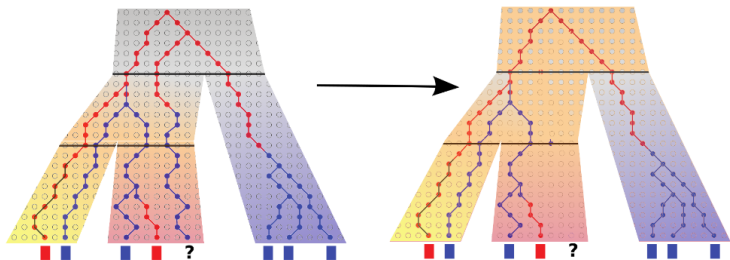


Bouckaert, PeerJ, 2016, [package: GEO_SPHERE](#)

7. But how about SNPs?

SNAPP: multi species coalescent, integrates out gene trees

SNAPP now handles missing data



Coming up:

- better ascertainment correction
- smarter operators

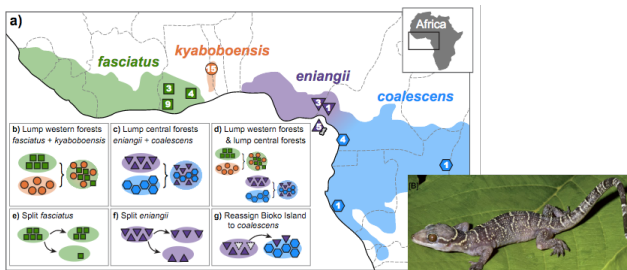
Bryant et al, MBE, 2012, [package: SNAPP](#)

8. How to do species delimitation?

Bayesian Factor Delimitation

- consider different species assignments in a multi species coalescent analysis
- use path sampling to calculate marginal likelihoods (MLs)
- determine Bayes factors from MLs
- BFD for *BEAST, BFD* for SNAPP

Fujita et al, Trends Eco & Evo, 2012, Leach et al, Sys Bio, 2014 , package: [model-selection](#), [SNAPP](#)



8. How to do species delimitation?

Bayesian Factor Delimitation

- consider different species assignments in a multi species coalescent analysis
- use path sampling to calculate marginal likelihoods (MLs)
- determine Bayes factors from MLs
- BFD for *BEAST, BFD* for SNAPP

Fujita et al, Trends Eco & Evo, 2012, Leach et al, Sys Bio, 2014 , package: [model-selection](#), [SNAPP](#)

STACEY: Species Tree And Classification Estimation, Yarely

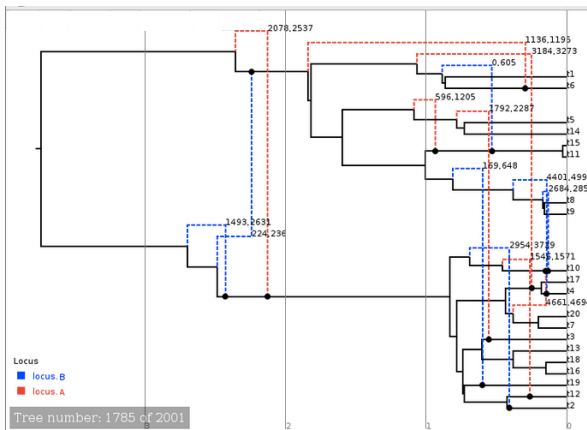
- joint estimate of species tree and species assignment
- considers a species everything joined with branch lengths $< \epsilon$
- faster than *BEAST: integrates out population sizes + better MCMC operators (but keep an eye on StarBeast2 now **>10x faster than *BEAST!**)

Jones, Journal of Mathematical Biology, 2016, package: [STACEY](#)

9. Do I have recombination in my bacterial data?

BACTER

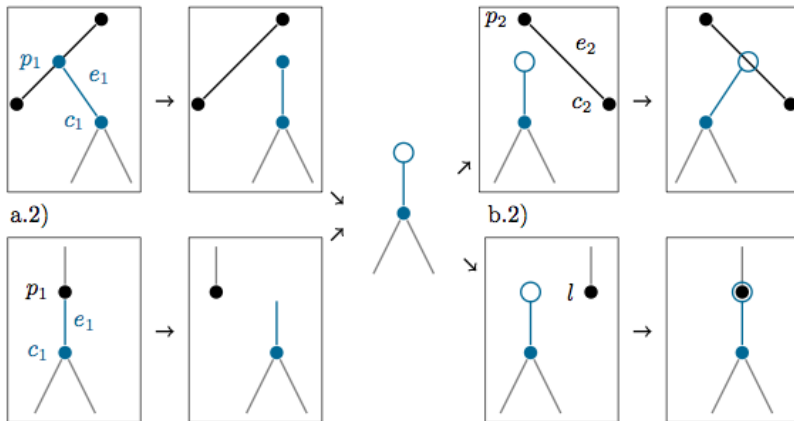
- ancestral recombination graphs
- estimate recombination rate
- estimate expected tract length associated with ARG



Vaughan et al, Genetics, 2017. Didelot et al, Genetics, 2010. [package: BACTER](#)

10. Can I test for ancestrality?

Sampled ancestors model uses reversible jump

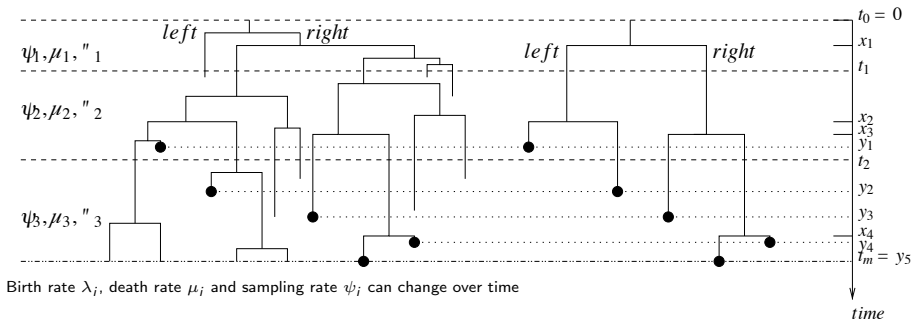


Useful for ancient DNA

Open question: how to specify hyper priors

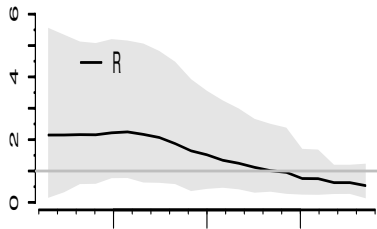
Gavryushkina et al. PLoS Computational Biology 2014 [package: SA](#)

11. Is there a skyline model for birth-death priors?



Reparametrization of birth-death process:

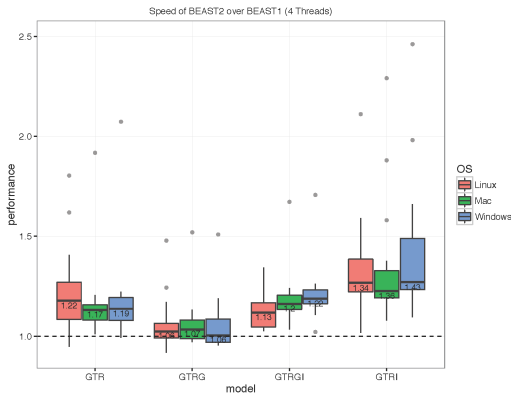
- Reproduction number $R_i = \frac{\lambda_i}{\psi_i + \mu_i}$
- Become-uninfectious rate $\delta_i = \mu_i + \psi_i$
- Sampling proportion $s_i = \frac{\psi_i}{\psi_i + \mu_i}$
- Fixed number of intervals m



12. Are we there yet?

How to speed things up:

- BEAGLE library
- use BEAGLE library with `-beagle_SSE` option
- almost no overhead when using proportion invariant category
- use threads



<http://beast2.org/2016/04/05/beast-1-vs-2-performance-benchmarking/>

13. Are there any others?

Starbeast2 (Canberra)

Protein evolution model (US)

Protracted speciation (Netherlands)

Correlated evolution models (UK)

Density dependent speciation (Switzerland)

Microsatellite models (Netherlands)

SCOTTI (UK)

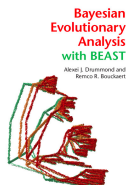
Tree Model Adequacy (Melbourne)

Babel – language evolution

Break-away – model of geography

Nested sampling – convenient marginal likelihood estimation ...

14. Further questions?



- BEAST book
- Tutorials, FAQ, blog: website: <http://beast2.org>
- BEAST users on google groups
<https://groups.google.com/forum/#!forum/beast-users>
- BEAST clinic (later today)



Thanks to all BEAST developers!