

Bioinformatics: Phylogenetics II

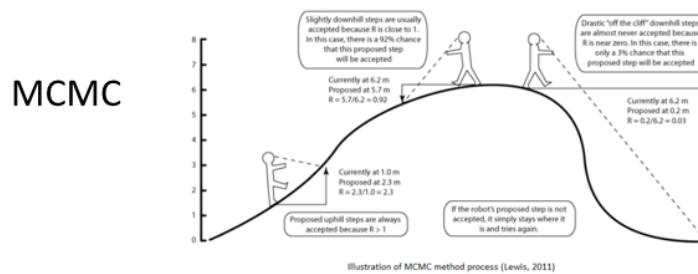
BEAST

Alicia, James and Thomas

a.toon@uq.edu.au

BEAST - Bayesian phylogenetic analysis using molecular sequences

Data, Models of evolution, Priors + Likelihood

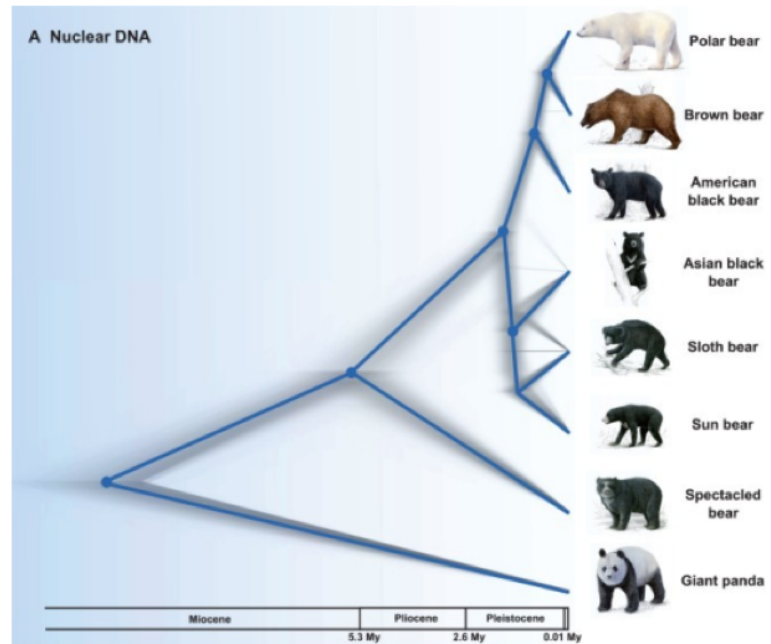


Beast2

Bayesian evolutionary analysis by sampling trees

Search algorithm allows for a collection of posterior trees that can be both used in subsequent analyses and help interpret the phylogenetic result

Search algorithm allows for a collection of posterior trees that can be used in subsequent analyses and help interpret the phylogenetic result.

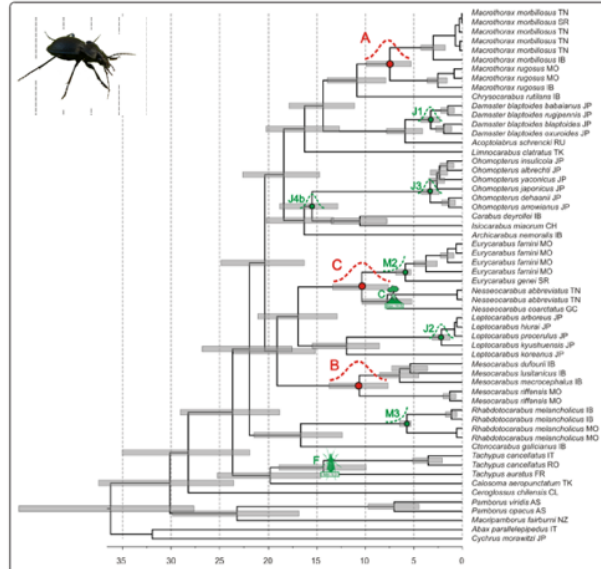


Versatile :
complete toolkit for
phylogeny, timing,
epidemiology, trait
reconstruction



Beast2

Bayesian evolutionary analysis by sampling trees



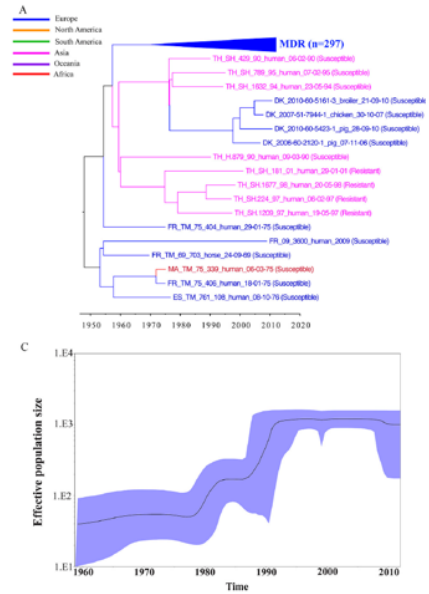
Anduar et al 2012

Versatile :
complete toolkit for
phylogeny, timing,
epidemiology, trait
reconstruction



Beast2

Bayesian evolutionary analysis by sampling trees



Leekitcharoenphon et al 2016

Sampling through time.

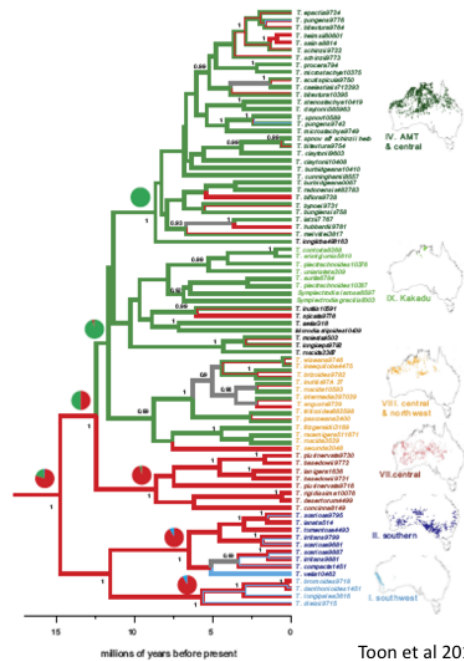
E.g. Tracing the spread of Salmonella resistance

Versatile :
complete toolkit for
phylogeny, timing,
epidemiology, **trait**
reconstruction



Beast2

Bayesian evolutionary analysis by sampling trees



Toon et al 2015

And more

- species delimitation – coalescent model
- model selection
- morphological data for phylogeny estimation
- SNP analysis
- correlations: among traits accounting for phylogeny
- Skyline plots: sampling Ne through time



Beast2

Bayesian evolutionary analysis by sampling trees

Pros and Cons

- Graphical User Interface – Black box
- Reliance on one method – multiple methods to test for robustness of the result given a particular set of assumptions



Beast2

Bayesian evolutionary analysis by sampling trees

Careful not to treat it like a black box.

Installation

BEAST

<http://www.beast2.org>

Beagle library

<https://github.com/beagle-dev/beagle-lib/blob/master/README.md>

Tracer:

<https://github.com/beast-dev/tracer/releases/tag/v1.7.1>

Figtree:

<http://tree.bio.ed.ac.uk/software/figtree/>

Tutorials

<http://www.beast2.org/tutorials/index.html>



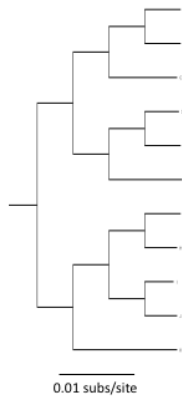
Beast2

Bayesian evolutionary analysis by sampling trees

Beauti

- Load nex file or multiple files
- Load partitioned nex files or partition file within Beauti
- Trait data : need to input manually or use BEAST 1.8
- Package Manager – to add new packages as they are developed
- Follow tutorial: DivergenceDatingTutorialv2.0.3

Clock model – Tree shape



Test for strict clock

- Run relaxed model and check for variation in rates

OR

- Use a model test or model adequacy approach

Relative

Absolute

Model testing can be done within Beast (path sampling).

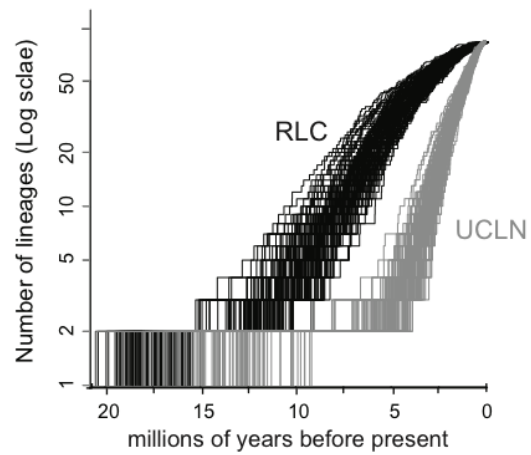
Clock model

Strict: one rate

Uncorrelated Relaxed Clock : allows each branch to have its own clock rate

Random Local Clock: fixed number of rates (uncorrelated or correlated)

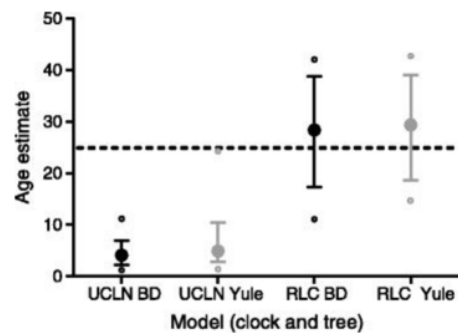
RLC vs UCLN



References

Toon, A., M. D. Crisp, H. Gamage, J. Mant, D. C. Morris, S. Schmidt, and L. G. Cook. "Key Innovation or Adaptive Change? A Test of Leaf Traits Using Triodiinae in Australia." *Sci Rep* 5 (2015): 12398.

RLC vs UCLN



References

Crisp, MD, NB Hardy, and LG Cook. "Clock Model Makes a Large Difference to Age Estimates of Long-Stemmed Clades with No Internal Calibration: A Test Using Australian Grasses." *BMC Evolutionary Biology* 14 (2014): 263.

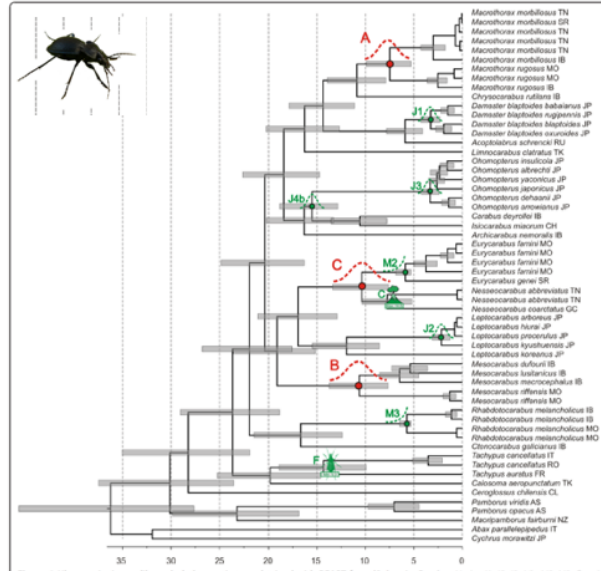
Broom vs bush trees

Fossil calibration



Beast2

Bayesian evolutionary analysis by sampling trees



Anduar et al 2012

Really need some sort of hard bound on a node: many studies will use a maximum age of the group

C volcanic emergence : uniform prior with hard upper limit : equal prob after the event.

F fossil: lognormal: most of distribution around the age of the fossil but a tail that lets the node age be older.

J1-4: breakup (beginning and late) of land: normal broad priors around the age that is known

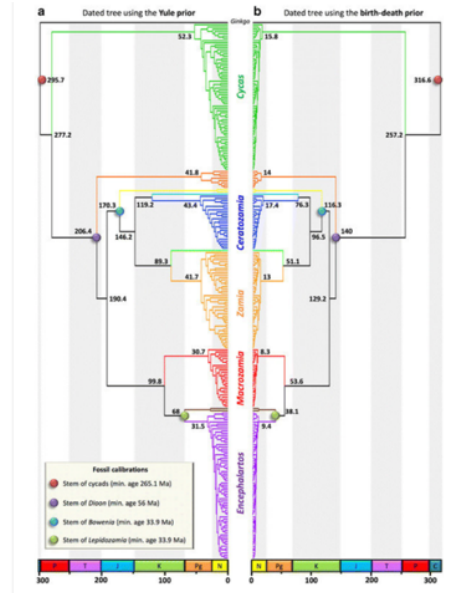
M2-2: opening up of a straight: exponential

We are

going to set $M = 1:78$ and $S = 0:085$ which will specify a distribution centred at about 6 million years with a standard deviation of about 0.5 million years.

This will give a central 95% probability range covering 5-7 Mya.

Yule vs Birth Death



Condamine et al 2015

Yule: birth only (probably only reasonable for closely related organisms)

BD: includes extinction in the model

Running Beast

- MCMC sampling and length : do short trial runs (10 000 000 every 1000)
- Use Beagle library (speeds up analysis)
- Interpret output from run (operator acceptance rates at the end of the run)
- Import to Tracer for analysis (check ESS, estimates, trace)
- RWTY (not in BEAST but very useful for visualizing MCMC convergence)
- Logcombiner (for multiple runs)
- Tree annotator: removes burnin, select MCC tree
- Open tree in Figtree