

# Introduction to Phylogenetics



## Part Three

**ETH** zürich

09.12.21

Bethany Allen

Computational Evolution, D-BSSE

This last section will focus on how we can use phylogenies as a template for other analyses.

# What is phylogenetics?

The science of inferring  
evolutionary relationships,  
often visualised using a phylogeny  
or “tree”

09.12.21

Phylogenetics workshop – Bethany Allen

Slide 2 of 25

Remember this definition from the previous parts. If we have a phylogeny, then we have a map of the evolutionary trajectory of our tips / OTUs.

## Why should I care about phylogenetics?

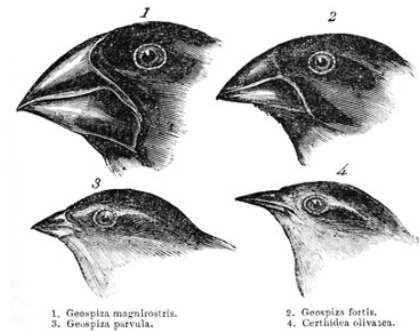
Understanding the evolutionary trajectory of a clade  
enables us to start asking questions about

- Diversity dynamics through time
  - Biogeographic processes
  - Trait evolution
  - Selectivity

...and many other facets of macroevolution

Once we have mapped this evolutionary trajectory, we can start inferring information about the direction and rate of specific changes in a clade's history. This is helpful for lots of questions around understanding evolutionary processes.

## Diversity dynamics



Darwin, 1845



09.12.21

Phylogenetics workshop – Bethany Allen

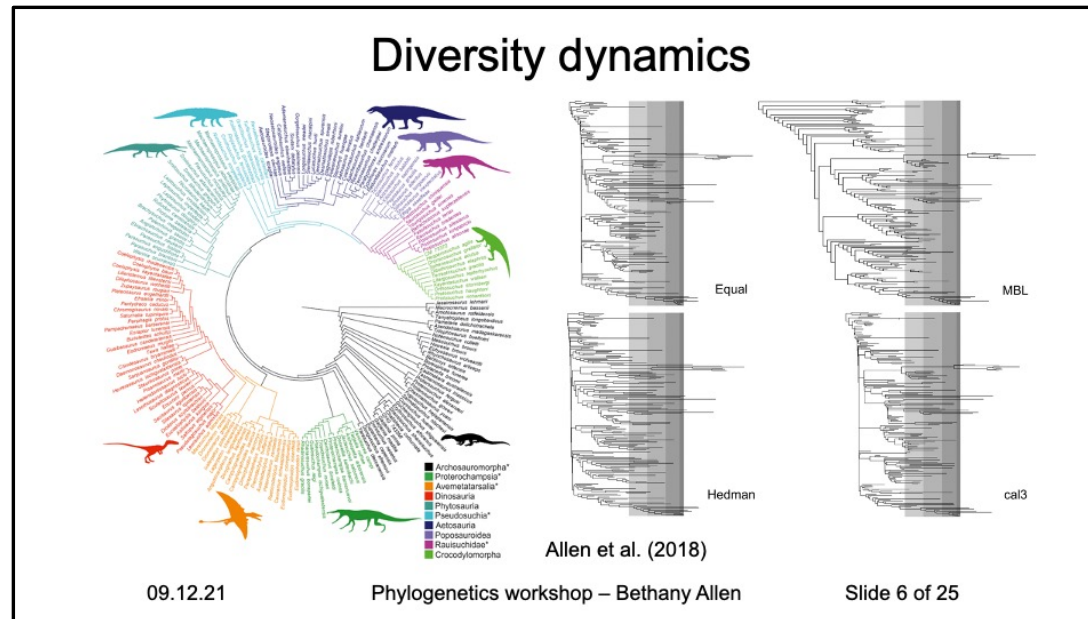
Slide 4 of 25

Understanding diversity patterns is something that I'm very interested in, as are many others. This includes evolutionary radiations, such as how Darwin's finches became so diverse, and mass extinctions – this is a depiction of the Triassic-Jurassic on the right.

## Diversity dynamics

The number of lineages in a time-scaled  
phylogeny at any point in time indicates  
diversity

Phylogenies can provide us with clear information about changes in the diversity of a clade through time, we just have to count the branches.



Here is an example from my own work, my MSc project. I was interested in the impact of the Triassic-Jurassic mass extinction on archosauromorphs (big four-legged animals including early crocodiles, pterosaurs and dinosaurs). I needed to determine which fossil species went extinct during the mass extinction and which did not (either before or after), so I constructed a supertree and used four different methods to time-scale it. The grey bars on the right show four proposed extinction intervals associated with the event. I could then simply count the branches which ended during each interval.

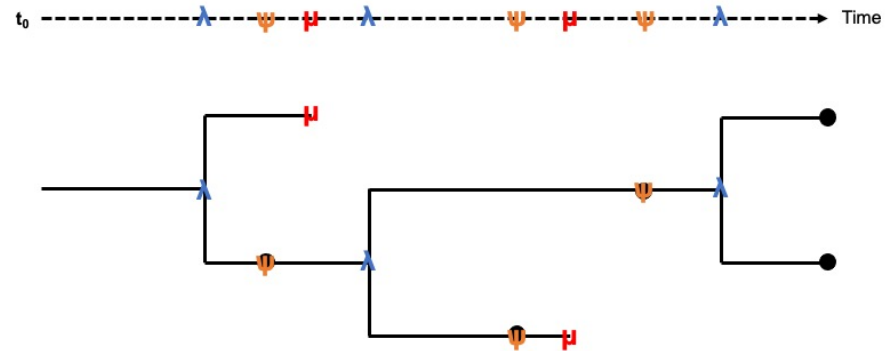
## Diversity dynamics

Phylogenies don't inherently inform us about missing lineages due to incomplete sampling

Integration into a Bayesian framework during phylogeny reconstruction means that "complete" trees can be used

Most phylogenies simply draw relationships between the tips (and sometimes fossils) for which we have samples. However, this ignores sampling bias; we don't know what we don't know, if you like. But sampling bias is affecting our estimates of diversity based on these phylogenies. One way to get around this links back to the last session.

## The fossilised birth-death process



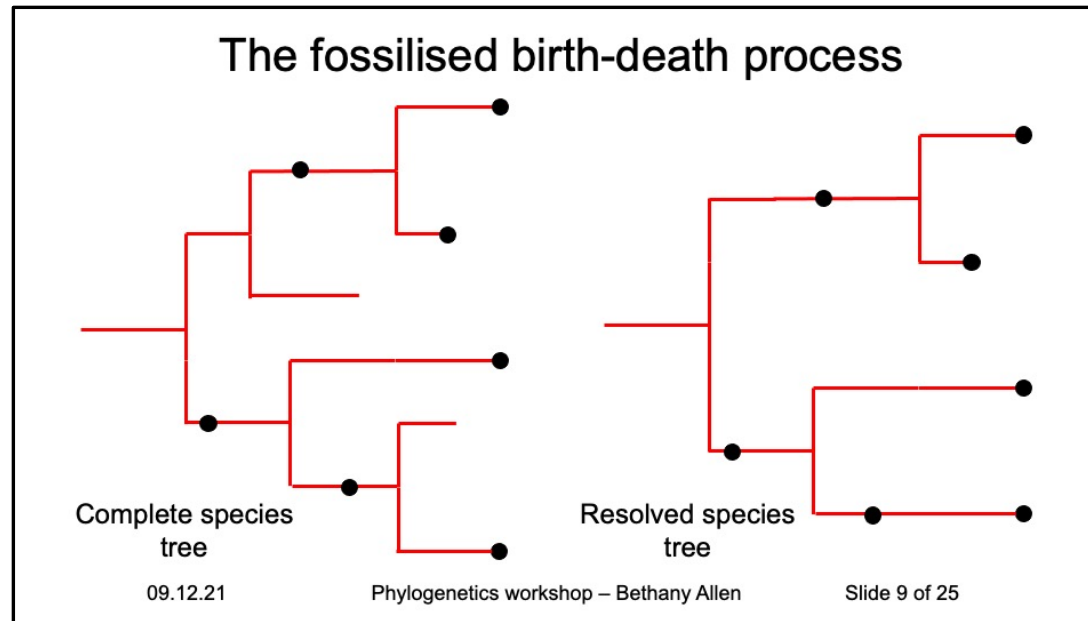
09.12.21

Phylogenetics workshop – Bethany Allen

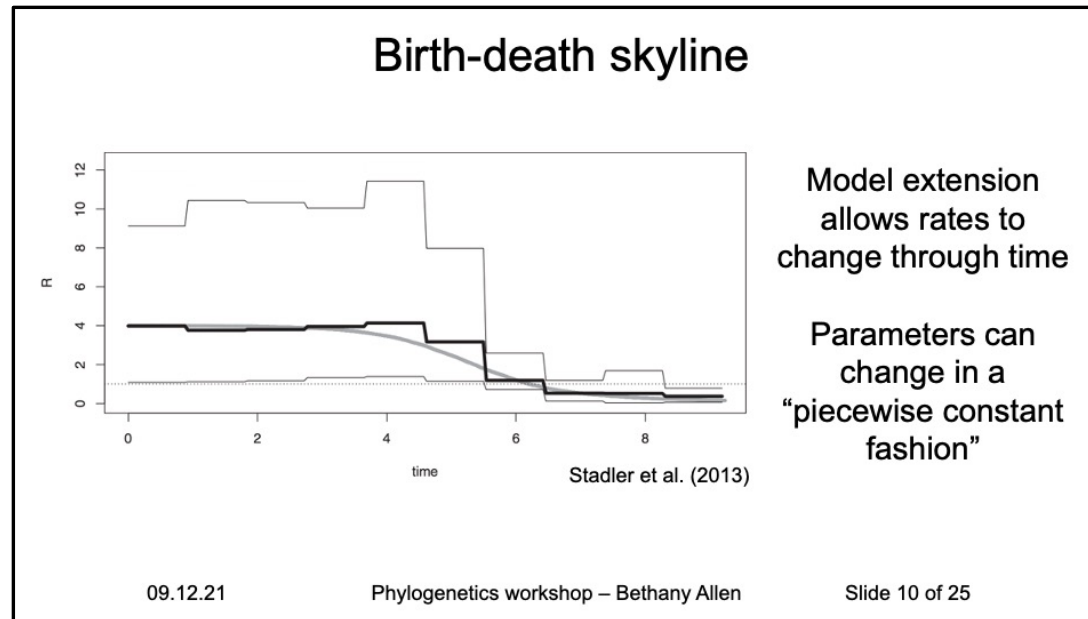
Slide 8 of 25

To quickly recap, the fossilised birth-death process samples values from prior distributions for various parameters in our model, then uses these to simulate a tree.



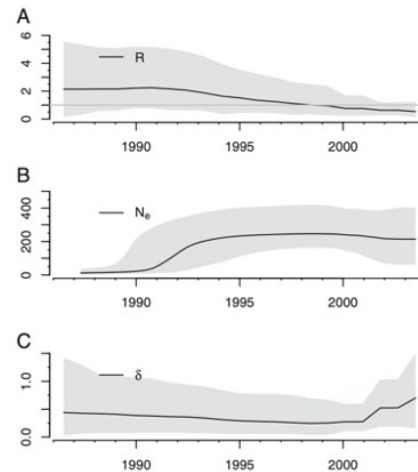


As a result, we have simulated a complete tree, including branches with no sampling. We trim these out to make the resolved tree, for comparison with our data. However, these complete trees give us an insight into what the whole tree might look like, in the absence of sampling bias. Estimating diversity through time might therefore be more accurate using these complete trees.



The original fossilised birth-death process had fixed parameters across the whole tree, but this has subsequently been relaxed. You can now allow your parameters to vary through time, in what we refer to as a “piecewise constant fashion”, where the rate is constant for an interval then reaches a break point and changes. You specify as part of your prior how many break points you expect. The MCMC tests the rates which fit the phylogenies best (although somewhat informed by the priors you set on the rates). This is recorded in a “skyline” plot, named after the tops of buildings which the traces resemble. The black line shows the best values, with the hard grey line showing a smoothed version. The faint grey lines show uncertainty around the parameter.

## Birth-death skyline



Uncertainty is  
captured across  
MCMC iterations

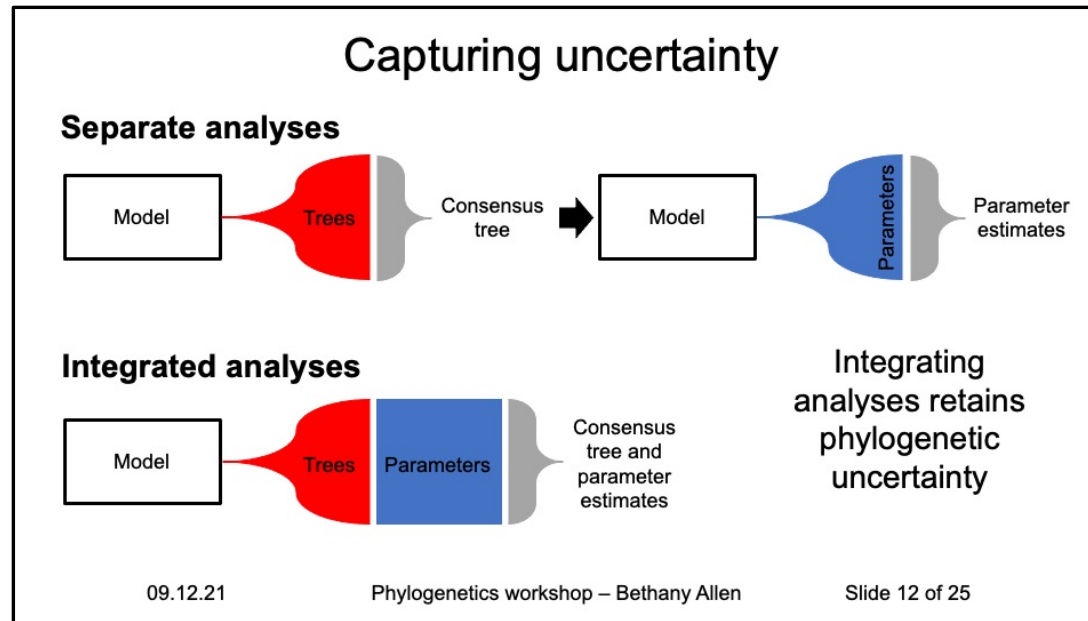
Stadler et al. (2013)

09.12.21

Phylogenetics workshop – Bethany Allen

Slide 11 of 25

Following your analysis you get the tree, the parameter values, and also these plots, which show speciation, extinction and diversity through time (the original examples are using virus epidemics, so the terminology is slightly different in this context). The confidence envelope shows where values are more or less supported. If you know a potentially influential event happened at a particular point in time, these plots can indicate whether a coincident shift in diversity occurred.



As well as basing diversity estimates on complete trees, so taking sampling bias into account, this Bayesian approach is also good because it captures uncertainty in the phylogeny. If you first produce a tree, collapse it into a consensus tree, then use that as the basis for an analysis looking at diversity through time, then the uncertainty you capture only reflects that one phylogeny. But if you take this Bayesian approach, the uncertainty on your diversity estimates also includes the impact of different possible phylogeny shapes.

## Trait evolution

Trait states can be mapped across phylogenies to  
examine their trajectories through time

Implementations exist for both discrete and  
continuous traits

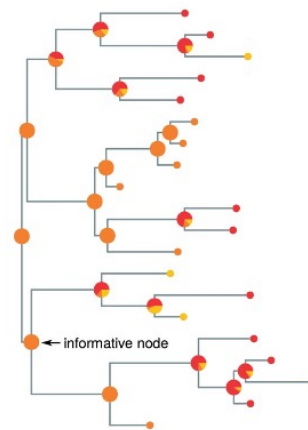
09.12.21

Phylogenetics workshop – Bethany Allen

Slide 13 of 25

Phylogenies can provide an important framework for understanding trait evolution. We can apply traits we know to the tips, and use ancestral reconstruction to trace trait states back through the phylogeny. In combination with branch length information, this can be used to infer the direction and rate of trait evolution through time. This can be useful for lots of different applications, including determining the timing of major evolutionary innovations and examining selectivity (the idea that extinction risk is tied to the possession of certain traits). One big question here is how best to collect data on the trait you are interested in, and whether that means you are thinking about discrete or continuous trait data; the models involved in ancestral trait reconstruction differ between these two types.

## Trait evolution



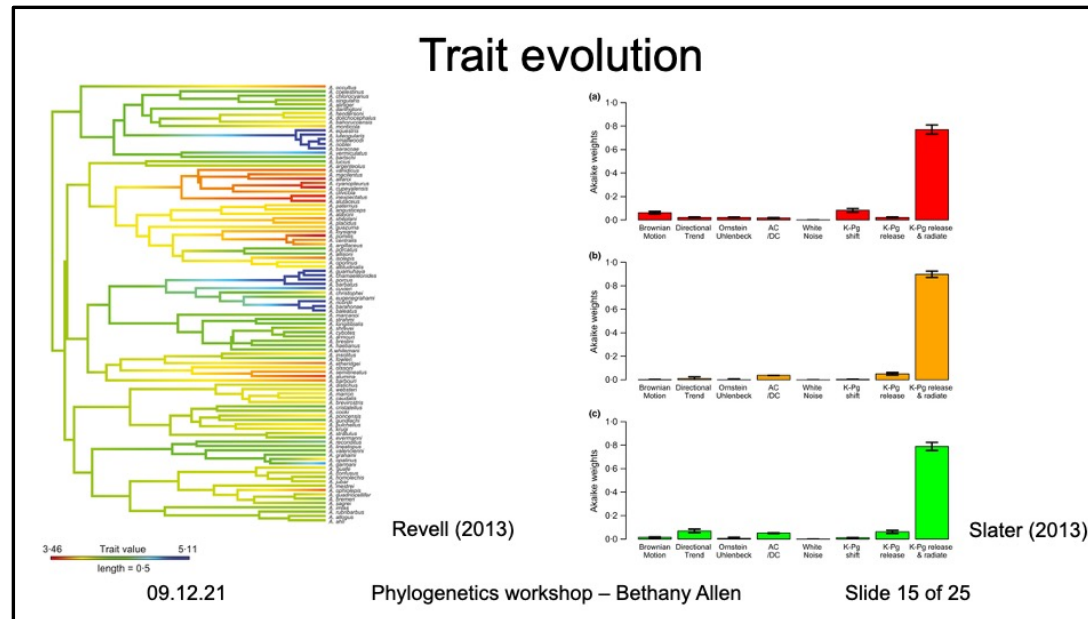
Borges et al. (2019)

09.12.21

Phylogenetics workshop – Bethany Allen

Slide 14 of 25

This is an example of an analysis with discrete trait data. Tips in the tree are labelled with their trait, then at each node, the trait state is determined for that ancestor. This is usually in the form of a pie chart, which provides relative probabilities of each of the different possible trait states.



Body size is a popular continuous trait for analysis. On the left is an example of how you can map body size across a phylogeny, here for *Anolis* lizards. You can see that the phylogeny is coloured based on size, with the range determined by that of tip values. Ancestral body sizes are then reconstructed across the branches. After doing this, you can extract the trait trajectories to investigate their direction and rate. On the right is an example of how you can use model fitting tests to determine which evolutionary model best fits your trait trajectory, looking at whether the K-Pg mass extinction affected the mode of mammalian body size evolution.

# Trait evolution

BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 15 2014, pages 2216–2218  
doi:10.1093/bioinformatics/btu181

Phylogenetics

Advance Access publication April 10, 2014

## geiger v2.0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees

Matthew W. Pennell<sup>1,\*</sup>, Jonathan M. Eastman<sup>1,†</sup>, Graham J. Slater<sup>2</sup>, Joseph W. Brown<sup>1,3</sup>, Josef C. Uyeda<sup>1</sup>, Richard G. FitzJohn<sup>1</sup>, Michael E. Alfaro<sup>2</sup> and Luke J. Harmon<sup>1</sup>

<sup>1</sup>Department of Biological Sciences and Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844, <sup>2</sup>Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, <sup>3</sup>Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA, <sup>4</sup>Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia and <sup>5</sup>Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA 90095, USA

Associate Editor: David Posada

## Methods in Ecology and Evolution

Methods in Ecology and Evolution 2012, 3, 217–223

doi: 10.1111/j.2041-210X.2011.00169.x

APPLICATION

## phytools: an R package for phylogenetic comparative biology (and other things)

Liam J. Revell\*

Department of Biology, University of Massachusetts Boston, Boston, MA 02125, USA

## Methods in Ecology and Evolution

Methods in Ecology and Evolution 2014, 5, 701–707

doi: 10.1111/2041-210X.12199

APPLICATION

## BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees

Daniel L. Rabosky\*, Michael Grundler, Carlos Anderson, Pascal Title, Jeff J. Shi, Joseph W. Brown, Huateng Huang and Joanna G. Larson

Department of Ecology and Evolutionary Biology and Museum of Zoology, University of Michigan, Ann Arbor, MI 48109, USA

## Reading Evolutionary Biology Group

Home People Research Publications Book Software Data Sets

BayesTraits V3.0.5 Aug 2021

Description

BayesTraits is a computer package for performing analyses of trait evolution among groups of species for which a phylogeny or sample of phylogenies is available. This new package incorporates our earlier and separate programs Multistate, Discrete and Continuous. BayesTraits can be applied to the analysis of traits that adopt a finite number of discrete states, or to the analysis of continuously varying traits. Hypotheses can be tested about models of evolution, about ancestral states and about correlations among pairs of traits.

09.12.21

Phylogenetics workshop – Bethany Allen

Slide 16 of 25

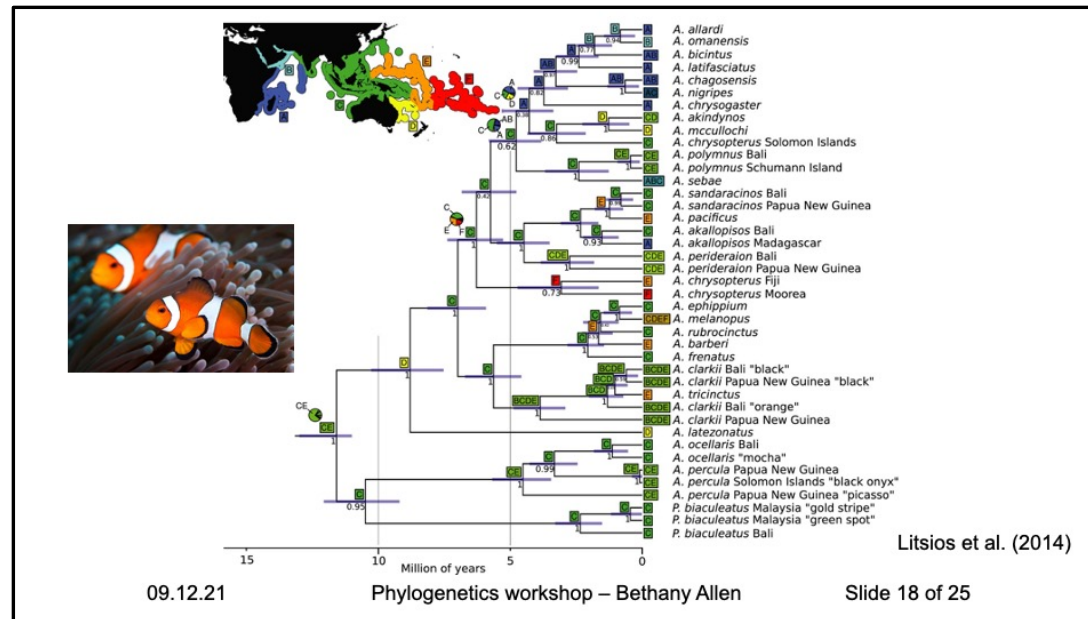
There are many, many different implementations available for running these types of analyses, and each has different approaches for conducting the ancestral reconstruction, different models available for fitting and different metrics for evaluating model fit. It's worth reading into them before deciding which one to use, but all of the examples here have been used widely in previous literature.



## Biogeographic reconstruction

The geographic regions occupied by lineages can be mapped across a phylogeny just like categorical traits, and used to reconstruct past distributions

Another really popular field of questions relates to the evolution of geographic ranges over a clade's history. This information can be mapped across a phylogeny in the same way as a trait, to test many different hypotheses about dispersal within a clade through time.



This example uses a phylogeny of clownfish to examine where the most recent common ancestor of the clade is likely to have originated. The area of interest is divided into regions, and the regions occupied by each extant species are labelled on the tips. This information can then be used to reconstruct ranges back through the tree, with the state of each node estimated based on the distribution of its constituent tips. As with discrete traits, you can see small pie charts to indicate places where there is uncertainty and give a relative likelihood for each possible state, including for the oldest node in the tree.

# Biogeographic reconstruction

Molecular Phylogenetics and Evolution 87 (2015) 46–49

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)

Short Communication

RASP (Reconstruct Ancestral State in Phylogenies): A tool for historical biogeography

Yan Yu<sup>a,b,\*</sup>, A.J. Harris<sup>c</sup>, Christopher Blair<sup>b,d</sup>, Xingjin He<sup>a,\*</sup>

<sup>a</sup>Key Laboratory of Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, Sichuan 610064, PR China

<sup>b</sup>Department of Biology, Duke University, Box 90338, Raleigh, NC 27690-3388, Durham, NC

<sup>c</sup>Department of Biology, Oklahoma State University, 301 Physical Science, Stillwater, OK

<sup>d</sup>Department of Biological Sciences, New York City College of Technology, The City University of New York

Search or jump to...

nmatzke / BioGeoBEARS Public

Code Issues Pull requests Actions Projects Wiki Security Insights

master 1 branch 2 tags

Go to file Add file Code

About

BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts

Readme

Cite this repository

nmatzke rerun example	r387a27 on 29 Aug	184 commits
R	Update BioGeoBEARS_traits_v1.R	4 months ago
inst	rerun example	3 months ago
tests	made functions to convert Biogeographic Stochastic Maps (BSMs) L...	3 years ago

09.12.21

Phylogenetics workshop – Bethany Allen

Slide 19 of 25

There are a few different methods for doing this, and they are implemented in many different packages. One I would suggest looking at is RASP, which is a GUI-based program which runs multiple different methods, you just need to give it your phylogeny and your tip distributions. Including distribution data from fossils as well can be a really good idea, and BioGeoBEARS is a popular R package which lets you do that. Working with spatial distributions in the fossil record can get really tricky, especially when you need to start worrying about continental drift, but I have experience with managing this type of data and can answer questions if you need help.

## Combining analyses

Many of the most interesting analyses can be carried out by combining these methods to ask more complex questions about evolutionary histories and their drivers

Often the questions we have about evolutionary processes are a little more complex than just wanting to understand the trajectory of a single trait or know the distribution of a clade's most recent common ancestor.

## Combining analyses

Was clade diversification associated with expansion into a new geographic region?

Was the evolution of a novel trait, or rapid trait evolution, linked to high levels of speciation?

Was extinction non-randomly distributed across the phylogeny, or associated with certain traits (selectivity)?

Evaluating changes in diversity, traits and geographic distributions in concert can allow us to consider questions such as these. This is a good place to start when thinking about what might be possible for your project given the data you have available.

## Designing analyses

Does my phylogeny capture the temporal interval and resolution I'm interested in?

What trait data do I have available, and is it best to analyse it as discrete or continuous?

How can I divide my species occurrences into regions which I can use to test my biogeographic hypotheses?

The next step is thinking about your research question and designing analyses that address it. It's important to remember that there are many different things that can be considered a “trait”, so getting creative with the data you have could lead to some really interesting results.

Wang et al. (2013). Molecular phylogeny and biogeography of *Percocypris* (Cyprinidae, Teleostei).

Xing & Ree (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot.

Qu et al. (2021). The evolution of ancestral and species-specific adaptations in snowfinches at the Qinghai-Tibet Plateau.

These papers are all relevant to the BECCY project and use phylogenies to test evolutionary questions.

## Find out more

Birth-death skyline plots, Stadler et al. (2013):

[doi.org/10.1073/pnas.1207965110](https://doi.org/10.1073/pnas.1207965110)

Including fossils improves trait evolution modelling, Slater et al. (2012):

[doi.org/10.1111/j.1558-5646.2012.01723.x](https://doi.org/10.1111/j.1558-5646.2012.01723.x)

Phylogenetic comparative methods for understanding trait evolution,  
Harmon (2019):

[lukejharmon.github.io/pcm/](https://lukejharmon.github.io/pcm/)

Harmon's textbook is a really comprehensive introduction to modelling trait evolution using phylogenies.



## Find out more

Reconstructing biogeography using RASP, Yu et al. (2015):  
[doi.org/10.1016/j.ympev.2015.03.008](https://doi.org/10.1016/j.ympev.2015.03.008)

Bayesian phylodynamics in BEAST, Suchard et al. (2018):  
[doi.org/10.1093/ve/vey016](https://doi.org/10.1093/ve/vey016)