**Model-based statistical inference in evolutionary biogeography**

by Nicholas J. Matzke, University of Auckland

Transmitting Science course, March 7-11, 2022

**Summary**

This course will cover the theory and practice of widely used methods in evolutionary biogeography, namely ancestral range estimation on phylogenies, statistical model comparison, and advanced techniques such as Biogeographical Stochastic Mapping, and trait-dependent dispersal modelling. The course will cover both the practical challenges to using these techniques (the basics of R, setting up the files and using the models), and the assumptions that various models and methods make. The course will also cover the statistical fundamentals behind model-based inference (likelihood, AIC, Bayesian techniques). R packages we will use include *ape*, *phytools*, and [*BioGeoBEARS*](http://phylo.wikidot.com/biogeobears).

**Requirements**

Training in the biological sciences, preferably some aspect of ecology or evolutionary biology. Participants must have a computer and be able to install programs (e.g. R, R packages). Prior experience with R is not required – we will cover the basics on day 1 – but will be helpful.

(Alternatively, students can use R Studio Cloud, a free R service that runs through their web browser.)

Zoom:   [https://us06web.zoom.us/j/89300603378](https://us06web.zoom.us/j/89300603378" \t "_blank)

**Setup**

Please:

0. Setup Slack (and get set up with your account)

1. Install RStudio (or get an RStudio Cloud account, if you are going to run R through a web browser).

(if you have prior R experience and prefer e.g. R.app on Macs, plain command-line R, you are welcome to use those; but most people prefer RStudio)

If you already have RStudio, but installed it years ago, please update, as some R packages may not work until RStudio is updated.

2. Once you have installed RStudio, open it, and go to the command prompt. To install the necessary R packages that we will be using, run these commands.

NOTES:

\* If R asks you to choose a CRAN mirror, pick one near you, or you can just use "0-Cloud" which is in the UK and which is always the most up-to-date.

\* If R asks you "Which packages would you like to update?", choose "CRAN packages only."

\* If R asks to install dependencies, say yes. Install them from CRAN (which has precompiled binaries, much easier to install).

**Installation code to run at the R prompt:**

install.packages("devtools")

library(devtools)

install.packages("Rcpp")

install.packages("ape")

install.packages("phytools")

install.packages("FD")

install.packages("GenSA")

install.packages("snow")

install.packages("rexpokit")

install.packages("cladoRcpp")

library(devtools)

# BioGeoBEARS now lives on GitHub instead of CRAN

# To install

# BioGeoBEARS version 1.1 from GitHub, install with:

devtools::install\_github(repo="nmatzke/BioGeoBEARS")

# NOTE: If R asks you "Which packages would you like to update?", choose "CRAN packages only."

# Check that your installed packages will load:

library(ape)

library(phytools)

library(GenSA)

library(FD)

library(snow)

library(parallel)

library(rexpokit)

library(cladoRcpp)

library(BioGeoBEARS)

**Outline**

*Day 1: Background on Historical Biogeography, and Intro to R*

Introductions and participant research projects

Short history of historical biogeography methods and assumptions

Introduction to the R statistical programming environment

R Studio

Interacting with the console

Working with variables

Control flow

Writing functions

Writing and reading scripts

Basic graphics

Introductory phylogenetics in R

How to read and use phylogenies

Running a basic biogeography model

*Day 2: Statistical Model Comparison, generally and in biogeography*

Models in biogeography, and their assumptions

Likelihood-based statistical model choice

Likelihood and Maximum Likelihood

AIC, AICc, BIC, vs. Bayesian analysis

Basic biogeography tutorial: comparing biogeographical models

*Day 3: Adding realism to historical biogeography*

Distance-dependent dispersal

Connectivity-dependent dispersal

Biogeographic Stochastic Mapping

*Day 4: Phylogenetic biogeography*

Trait-dependent dispersal models

Other current methods

*Day 5: Discussion, students’ research*

Critical thinking about flaws in our models

Upcoming methods and their strengths and weaknesses

1. Showing where BioGeoBEARS-type models are equivalent to ClaSSE models (in R)
2. Ree & Sanmartin (2018) critique; Matzke (2022) response

3. A biogeography SSE model in Julia – in development

Help sessions