

PHYLOGEOGRAPHY WORKSHOP

INTRODUCTION

Alex Beams

Department of Mathematics
Simon Fraser University

PhylogeogRaphy Workshop

September 29, 2025

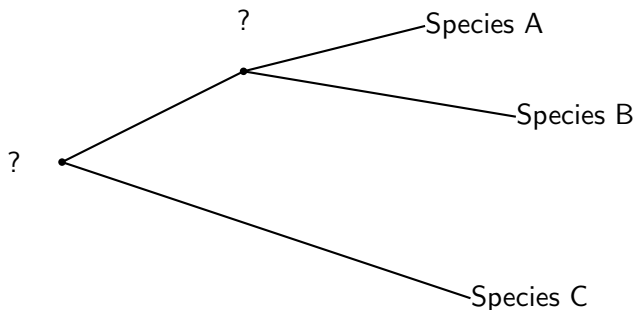
- 1 WHAT IS PHYLOGEOGRAPHY?
- 2 EXAMPLE DATASETS AVAILABLE TO USE
- 3 OVERVIEW OF THIS WORKSHOP

TABLE OF CONTENTS

- 1 WHAT IS PHYLOGEOGRAPHY?
- 2 EXAMPLE DATASETS AVAILABLE TO USE
- 3 OVERVIEW OF THIS WORKSHOP

FIRST, PHYLOGENIES:

Relationships among organisms are frequently represented with trees:



MARKOV CHAINS UNDERLIE PHYLOGENETIC RECONSTRUCTION

Nucleotide changes over time are modeled with continuous-time Markov chains

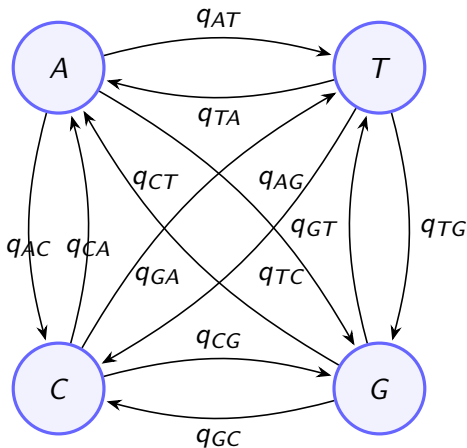


TABLE OF CONTENTS

- 1 WHAT IS PHYLOGEOGRAPHY?
- 2 EXAMPLE DATASETS AVAILABLE TO USE
- 3 OVERVIEW OF THIS WORKSHOP

AVAILABLE DATASETS

We will use some publicly available datasets in this workshop. Many of them are available at taming-the-beast.org, and others have been taken from published articles carrying out phylogeographic/ancestral character estimation analysis.

The main datasets are

- Primate mitochondrial DNA
- Influenza H3N2
- Influenza H5N1
- Bat rabies
- MERS coronavirus

PRIMATE MITOCHONDRIAL DNA

This dataset is taken from the introductory BEAST2 tutorial
(located at
<https://taming-the-beast.org/tutorials/Introduction-to-BEAST2/>)

12 sequences, 898 sites, no location data – for intro BEAST2 runs only

This might be useful for you to use on Thursday afternoon
(BEAST2 intro)

INFLUENZA H3N2

This dataset is also taken from the “Taming the Beast” tutorials (located at <https://taming-the-beast.org/tutorials/Structured-coalescent/>)

60 sequences, 1762 sites, 2 locations – New Zealand and Hong Kong

INFLUENZA H5N1

This dataset is taken from an older “Taming the Beast” tutorial (located at <https://beast2-dev.github.io/beast-docs/beast2/PhylogeographyDiscrete/AR.html>)

43 sequences, 1698 sites, 5 locations (Guangdong, Guangxi, Fujian, Hunan, Hong Kong)

MERS CORONAVIRUS

MERS coronavirus genomes are available from a published article¹

174 genomes from human infections, 100 from camel infections

Using phylogeographic methods to learn about the types of hosts
the virus infects

¹Gytis Dudas et al. "MERS-CoV spillover at the camel-human interface".

In: *elife* 7 (2018), e31257.

TABLE OF CONTENTS

- 1 WHAT IS PHYLOGEOGRAPHY?
- 2 EXAMPLE DATASETS AVAILABLE TO USE
- 3 OVERVIEW OF THIS WORKSHOP

PLAN FOR THE WEEK

We will cover a lot of topics, but the materials on these web pages are designed to help you make the most of this week. The first two days will focus on discrete trait analyses, the second two days will focus more on phylodynamics.

- Monday: Markov models, ancestral character estimation, and stochastic character mapping
- Wednesday: Birth-death models, State-dependent speciation and extinction (SSE) models
- Thursday: Intro to Phylodynamics models and BEAST2
- Friday: More advanced topics in BEAST2
 - Bayesian stochastic search variable selection
 - comparisons with discrete trait analyses

MONDAY: ACE AND SIMMAP

Monday's topics include the simplest models underlying “discrete trait analysis”

These methods assume a phylogenetic tree is given, and aim to reconstruct trait/location evolution on the tree

Phylogenies will be provided; in practical R sessions, the codes also work with simulated trees to help you get a feel for the methods/models

We will will work with the R functions `ace` and `make.simmap`, which are both based on similar types of Markov models describing discrete state changes on a phylogeny

- `ace`: Ancestral character estimation, R package `ape`
- `simmap`: stochastic character mapping, R package `phytools`

WEDNESDAY: BIRTH-DEATH MODELS AND BiSSE

Wednesday's topics include more advanced models underlying “discrete trait analysis” State-dependent speciation and extinction (SSE) models include the following functions:

- BiSSE: binary SSE
- MuSSE: multi SSE

These are in the diversitree package in R

These methods were originally developed for macroevolution (i.e., ultrametric trees) but the models they are based on will help us bridge to more advanced phylodynamics models

Analyses with BiSSE and MuSSE typically treat phylogenies as given, but the models can be extended to help estimate phylogenies (phylodynamics)

THURSDAY: PHYLODYNAMIC MODELS AND BEAST2

Thursday's topics include an introduction to phylodynamic models, and in the afternoon we will begin working with BEAST2

By **phylodynamics**, I am broadly referring to models falling into two classes:

- multi-type birth death (MTBD) models
- structured coalescent models

We will see that multi-type birth death models are a natural extension/generalization of the SSE models. Coalesecents construct genealogy in backward time; birth-death models build trees in forward time

THURSDAY: PHYLODYNAMIC MODELS AND BEAST2

In BEAST2, MTBD and structured coalescent models are referred to as “tree priors”

A full phylodynamics analysis in BEAST2 jointly estimates a nucleotide substitution model and a phylodynamic model using Markov Chain Monte Carlo

In the afternoon, we will try to get BEAST2 to work on the Fir server overnight and see how it does by Friday morning

There will also be BEAST2 output files available for you to analyze if all else fails

FRIDAY: MORE ADVANCED BEAST2 TOPICS

On **Friday**, we will begin the day by examining our overnight BEAST2 runs We will then try to cover some more advanced topics:

- Bayesian stochastic search variable selection (BSSVS)
 - Used to infer migration patterns when there are lots of locations, but movements are thought to be relatively rare

We will then spend the afternoon carrying out comparative analyses using the various discrete trait analyses (ace, simmap) with more computationally intense phylodynamics analyses in BEAST2

NEXT:

Now, we will begin the first practical session!