# 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

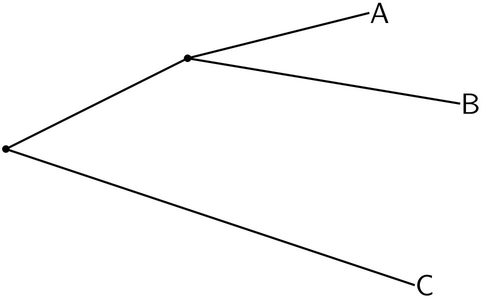
# First, phylogenies:

Phylogenies describe ancestral relationships among organisms

Our samples could be representatives of species (macroevolution), or they could represent sequenced viruses from infections (microevolution)

# How do we interpret trees?

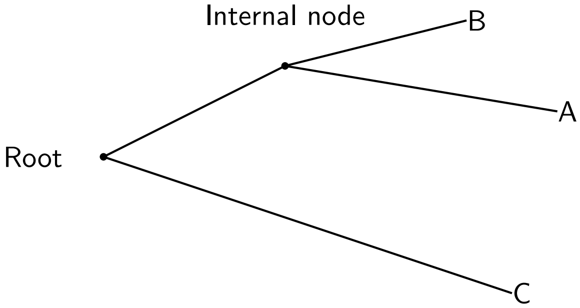
Trees have a **topology** tells us about the ancestry relationships of the taxa (species, infections, etc.)



This tree's topology can be written: ((A,B),C)

# How do we interpret trees?

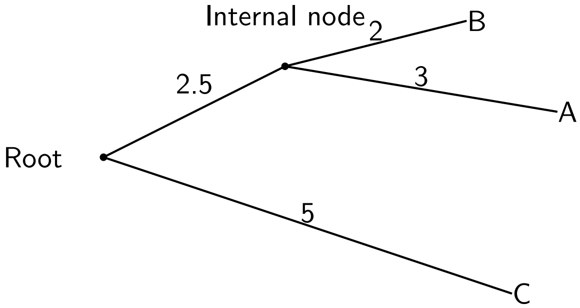
Trees have a **topology** tells us about the ancestry relationships of the taxa indicated at the leaves (species, infections, etc.)



This tree's topology also can be written: ((A,B),C)

# HOW DO WE INTERPRET TREES?

The branches/edges of trees have lengths. Units of branch length are often substitutions/site/year, but molecular clock models can rescale these to units of time



These branch lengths can be used to define **patristic (or phylogenetic) distances** between the taxa. For example:

$d(A, B) = 2 + 3,$

$d(\text{Internal node}, C) = 2.5 + 5$

# Lots of things phylogenies can be used for

In addition to describing evolutionary relationships among organisms, phylogenies are useful for understanding evolution of traits

- phylogenetic comparative methods
- discrete traits

Phylogeography can be treated in discrete or continuous space, but we will focus on discrete-state phylogeography in this workshop

# HUMAN CULTURAL EVOLUTION

Phylogeographic analysis of humans supports movement of populations out of Africa to the rest of the world[1]

Useful for exploring hypotheses related to cultural evolution

- associations of language with phylogeny
- hunter-gatherer societies: more matriarchal, similar mitochondrial DNA $\sim$ similar language
- agricultural societies: more patriarchial, Y chromosome similarity $\sim$ similar language

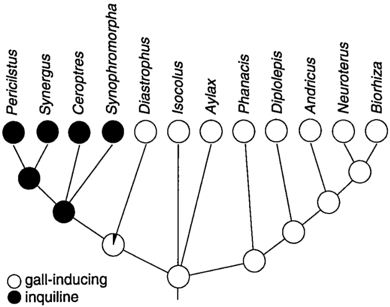
---

[1] Alexander H Harcourt. "Human phylogeography and diversity". In: *Proceedings of the National Academy of Sciences* 113.29 (2016), pp. 8072–8078.

# EVOLUTION OF DIET IN BIRDS

How did iniquilinism in wasps evolve?[2]

Some wasps induce galls in plants to lay their eggs; **iniquilinism** is when some species of wasps steal galls from other wasps

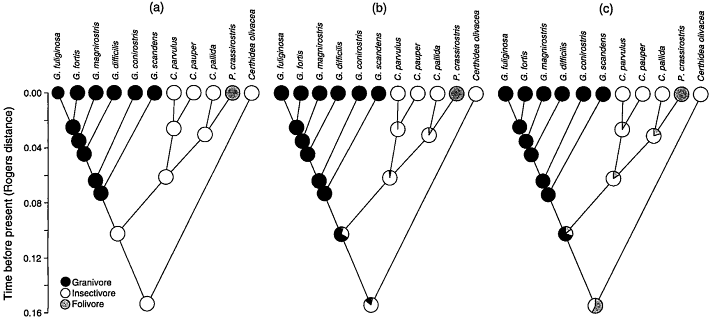


---

[2] Dolph Schluter et al. "Likelihood of ancestor states in adaptive radiation". In: *Evolution* 51.6 (1997), pp. 1699–1711.

# Evolution of diet in birds

How did the modern diets of finches evolve?[3]

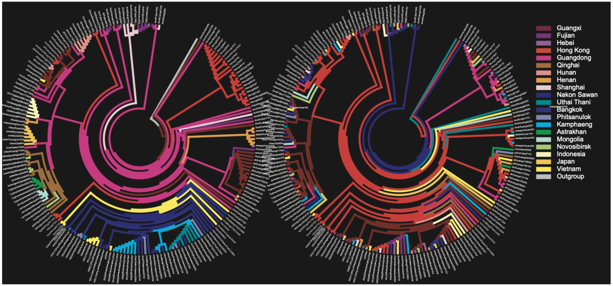


---

[3] Dolph Schluter et al. "Likelihood of ancestor states in adaptive radiation". In: *Evolution* 51.6 (1997), pp. 1699–1711.

# INFLUENZA H5N1

How does Influenza H5N1 move around the world?[4]

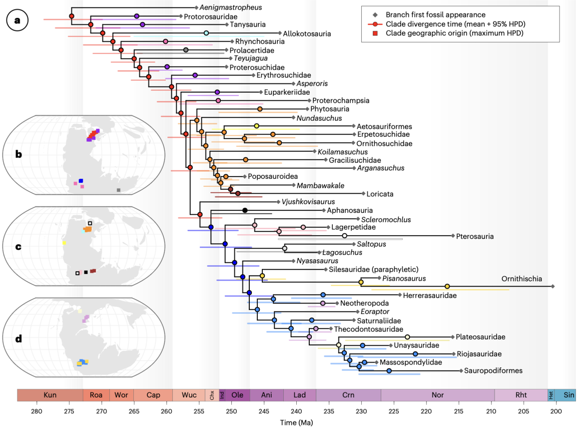


---

[4] Robert G Wallace et al. "A statistical phylogeography of influenza A H5N1". In: *Proceedings of the National Academy of Sciences* 104.11 (2007), pp. 4473–4478.

# REPTILES

Where do particular species originate?[5]

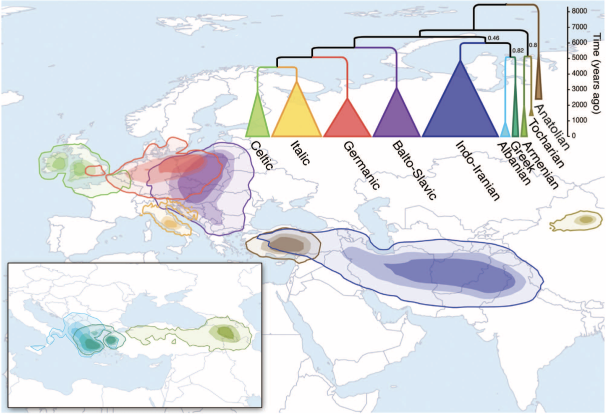


---

[5] Joseph T Flannery-Sutherland et al. "Landscape-explicit phylogeography illuminates the ecographic radiation of early archosauromorph reptiles". In: *Nature Ecology & Evolution* (2025), pp. 1–15.

# Language

How do languages evolve?[6]



---

[6] Remco Bouckaert et al. “Mapping the origins and expansion of the Indo-European language family”. In: *Science* 337.6097 (2012), pp. 957–960.

# PHYLOGEOGRAPHY BY ANOTHER NAME

Synonyms for phylogeography (as far as this workshop is concerned):

- ancestral character estimation
- ancestral state reconstruction
- ancestral trait reconstruction

Phylogeography is a bit of a misnomer: really, the method we will be learning about can be used to study evolution of any discrete trait

There are methods for analyzing continuous traits, but we will not explore those in this workshop

# TWO CATEGORIES OF PHYLOGEOGRAPHIC ANALYSIS

- Discrete trait analysis
- phylodynamics

# DISCRETE TRAIT ANALYSIS

The analyses we learn about in the first half of this week fall under the category of **discrete trait analyses**

These treat phylogenetic trees as fixed, known without error.

These analyses can all be carried out in R (and that will be our focus)

# PHYLODYNAMICS

The second half of the workshop will focus on **phylodynamic** methods

These jointly reconstruct phylogenetic trees and phylogeographic histories

We will take R as far as it can go, but will need to learn how to use BEAST2

# SSE MODELS: BRIDGING THE GAP

On Wednesday, we will learn about state-dependent speciation and extinction (SSE) models, which bridge this gap – they treat phylogenies as fixed, but the underlying models are the basis of the phylodynamics methods

# TABLE OF CONTENTS

# 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 article[7]

174 genomes from human infections, 100 from camel infections
Using phylogeographic methods to learn about the types of hosts the virus infects

---

[7] Gytis Dudas et al. “MERS-CoV spillover at the camel-human interface”. In: *elife* 7 (2018), e31257.

# NEXTSTRAIN

Let's navigate to Nextstrain and download some data from there

# VertLife

Check out this site: https://vertlife.org/data/

# Table of Contents

# 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 extinctinon (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. Coalescents 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

# SOME COMMENTS ABOUT HOW THIS WORKSHOP WILL FUNCTION

As you may notice from the website, modules for this workshop contain either slides or code demos.

Feel free to come and go as necessary to meet your obligations for classes, etc.

The practical sessions in R are designed to help you gain familiarity with the different methods we are learning about in this course

The codes are designed to work with simulated trees, but you may wish to edit things to load example phylogenies from real datasets

As we progress through the week, there will be opportunities to compare results from the different types of analyses available to us

# NEXT:

Today, we will be working in R. You will need the following packages installed:

- ape
- expm
- phytools
- coda

# SCHEDULE FOR TODAY:

- 1030-1130 – practical session in R
  - Working with phylogenetic trees in R
  - Simulating nucleotide substituion models
- 1130-1230 – Ancestral character estimation (slides)
- 1230-1330 – lunch break
- 1330-1430 – practical session in R
  - Simulating and fitting Markov models with ace
- 1430-1530 – Stochastic character mapping (slides)
- 1530-1630 – practical session in R
  - Simulating stochastic character mappings
  - Fitting models with stochastic character mappings

Feel free to work through the optional MCMC primer (practical session in R) if there is downtime