R Assignment 5: Combining Phylogenetics with Biogeography

#### Motivation

Add section…

#### Prerequisites

Before starting this week’s tutorial you will have:

1. completed Assignment 2 (Mapping) and Assignment 4 (Phylogenetics),
2. completed some pre-assignment reading on the use of phylogeography in plant taxonomy,
3. more…

#### Outcomes

By the end of this tutorial you will have:

1. the ability to combine distribution data from GBIF with sequence data from Genbank to create phylogeographic maps of species distributions,
2. more…

#### Activity 1: Creating a Phylogeographic Map from GBIF and Genbank Data

In previous Assignments we mapped species distributions and constructed phylogenetic trees. In this Assignment, we’ll combine both lines of research to simultaneously visualize the phylogenetic and geographic relationships of five *Opuntia* species.

In this lab, we’ll explore further the genus *Opuntia*:

*Opuntia basilaris* (Mojave, Colorado, Utah) *Opuntia fragilis* (Northwestern) *Opuntia humifusa* (Eastern) *Opuntia oricola* (Coastal sage and chaparral of S. California) *Opuntia polyacantha* (Great Plains, foothills of Rocky Mtns)

**MB** If we’re using this example, some background will have hopefully been imparted in the last Assignment.

Open this week’s .R file (‘324\_Phylogeography.R’) and navigate to the source pane. Install and attach the necessary packages using install.packages() and library(). Start by importing your species distribution data (from GBIF, Assignment 2) and your phylogenetic tree (Assignment 4). By highlighting and running the entire script, the data are combined and plotted.

#### Activity 2: Dechyphering some code

Let’s step through the script in order to understand what is happening here.

**Lines 14-18**: The library() function loads various packages designed specifically for organizing data (tidyverse), creating custom colour palettes (viridis), and mapping (mapdata, phytools).

**Lines 24-27**: Import your GBIF and phylogenetic data from previous assignments.

**Lines 30-38**: To reduce clutter in our plot, we’ll rename the rows in our dataframe with shortened versions of the species’ names.

**Lines 41-44**: We’ll extract the latitude and longitude information from our larger dataframe, and store it as a matrix style object. (This is because phytools::phylo.to.map() can only handle lat-long info of the matrix class.)

**Line 50**: Here, we’re removing any rows that have missing lat-long information (NAs).

**Line 53**: In order to combine the phylogenetic and distribution data, both data objects need to have the species names in the same format.

**Lines 71-81**: The phytools::phylo.to.map() function combines our distribution data latlong with our phylogenetic tree con\_tree\_rooted. We use the regions parameter to specify what parts of the world to plot, and we’ll use the ‘Hi-resolution World’ setting under database.

**Lines 84-92**: We can plot our combined dataset using the phytools::plot() function.

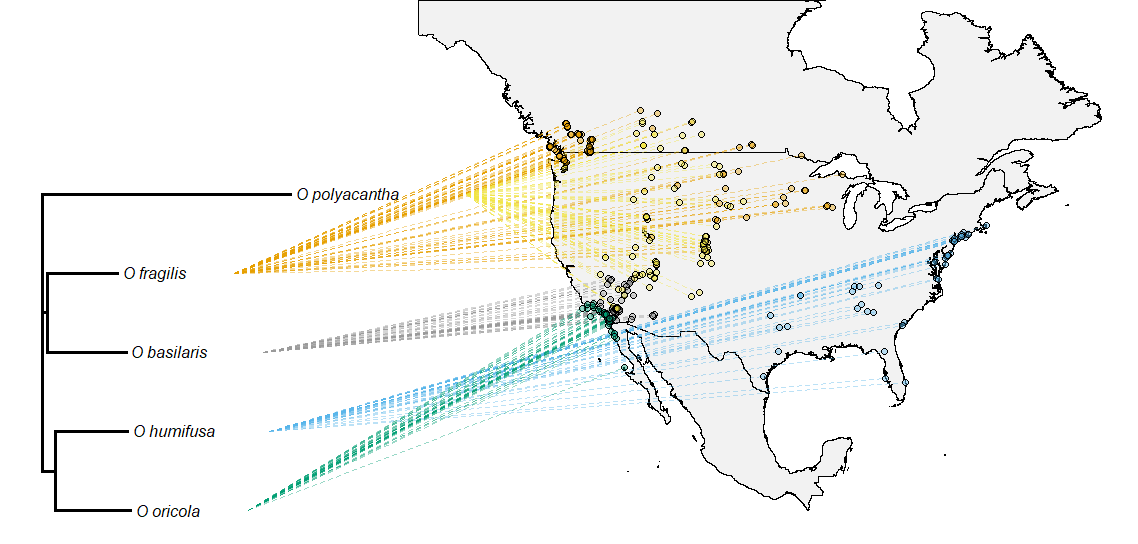


Figure 1: Phylogeographic distribution (with points) of some North American *Opuntia*.

**Lines 84-92**: In this example, we’ve set cex.points to zero, which tells the plotting function not to plots circles at each lat-long location. This can make our map a bit cleaner when plotting hulls.

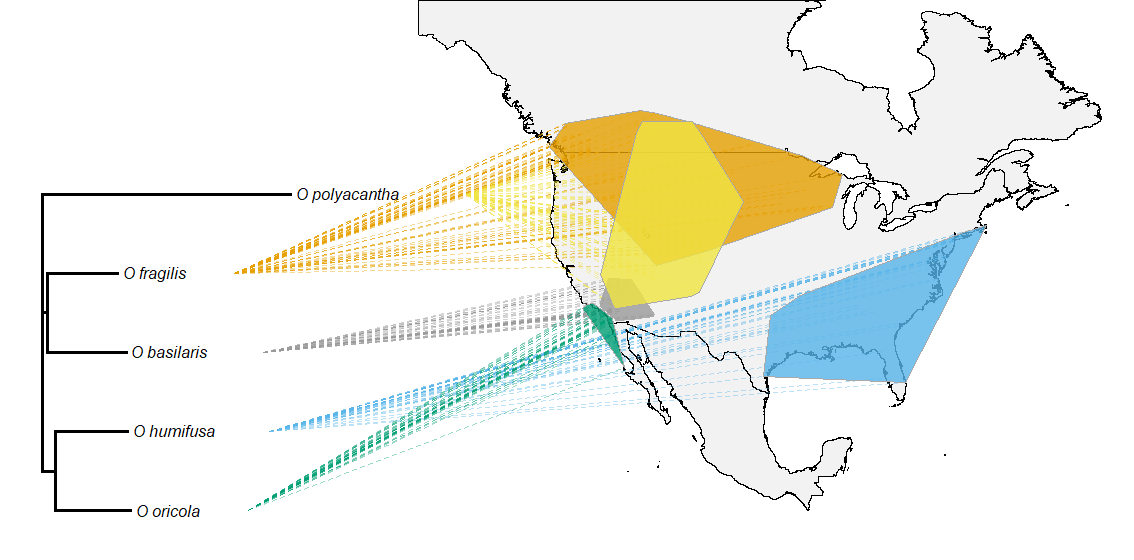


Figure 2: Phylogeographic distribution (with hulls) of some North American *Opuntia*.