

Primary working group members: Ty Tuff, Bruno Vilela, and Carlos Botero

VHS (vertical, horizontal, or selection) model

Overview:

We are part of a larger working group studying the ecology of human cultural evolution. This group has identified a critical knowledge gap, which is to identify the mechanisms by which ideas are transmitted through the environment. There are four possible modes of transport for ideas, vertical transmission between generations, horizontal transmission between cultures, and selection from the environment or society. These four mechanisms are all playing a role in the transmission of language and we would like to disentangle their relative contributions to the process. These questions are currently beyond the scope of field experimentation, so we will use a modeling approach to establish the boundaries of our current understanding and inform future experimental designs. A key function of this model must be that it outputs phylogenetic trees that can be compared against the most trustworthy data on the spread of cultural ideas, language phylogenies. This contribution should provide valuable contributions to both ongoing studies (e.g. the spread of agriculture) and future questions that generically involve the spatial spread of cultural ideas. We must maintain independence between the modes in the model so that we can isolate the roll of each in a nested experimental design. To test these various modes of idea transmission, we focus on the evolution and spread of human agriculture because it is a transmittable trait (cultural idea) with many ecological underpinnings that can be tested against data. We compete these models with both ecological data (e.g. abiotic conditions, suitability of environment for agriculture, and species richness of domesticated plants and animals) and evolutionary data (e.g. phylogeny of human language and culture) in a contest of direct inference (i.e. competing all available hypotheses against eachother to decide which most effectively describe the data).

Project summary:

Historians, archaeologists and geneticists have struggled to establish how and why agriculture originated abruptly and then quickly spread across the globe as it was adopted by a wide diversity of different cultures across a variety of environmental conditions. Currently, there are two primary mechanisms proposed for the transmission of culture through space that correspond to the vertical and horizontal components of evolutionary transmission. Cultural diffusion describes the *horizontal* diffusion of ideas and technology between societies, while Demic Diffusion describes the *vertical* transmission of ideas from one generation to the next through population expansion of the first farmers that subsequently displaced foraging societies. These two mechanisms have been difficult to disentangle because the scarcity of data and the limited application of quantitative analysis, so they have generally been treated as mutually exclusive and tested independently. We propose that the two process (Cultural and Demic diffusion) will produce different patterns of language evolution and that such differences could be captured by modern phylogenetic methods applied to phylogenies of human language.

After, we plan to compare the modelling results with the real world linguistic phylogeny to indicate how the farming technology was spread among human societies. The results of this work has

the potential to answer an important question regarding human history, that after all could have been always there, hidden in the relationship of our languages altogether.

Scientific question

- What relative contributions do the four modes of cultural transmission make to the spatial spread of cultural ideas? Can changes in these mechanisms be detected in the phylogenies of cultures that have undergone spatial change?

Hypotheses:

- 1) When provided a phylogenetic tree, it is possible to reconstruct the spatial process that created that tree from the information in the tree.
- 2) Each of the four spatial mechanisms alone will produce a different phylogenetic outcome.
- 3) Each combination of spatial mechanisms will produce a different phylogenetic outcome.

Explicit assumptions:

- 1) Vertical transmission is described by two different processes, speciation (where a society splits into un-colonized space by creating a new branch on the tree) or neighbor takeover (where a society drives a neighbor extinct so the invader speciates into the neighbors spot and the extinct society is removed from the tree).
- 2) Horizontal transmission is described by a simple diffusion process where one culture gives their neighbor a trait, but the neighbor maintains their own tip on the tree.
- 3) Selection is caused by the underlying habitat. Currently this habitat, and therefore the fitness surface, is defined as a binary choice between matching or mismatching habits for each transmission mode.

Predictions:

A priori, what shape of tree do we expect each transmission mode to produce?

Vertical

Horizontal

Env Selection

Social selection

We expect that the spread of farming through Demic Diffusion (vertical transmission) would lead to more speciation of farming societies (new societies driven by expansion and conquering), more societies extinctions of foragers (caused by conflicts) and an aggregation of farming technology among societies that descend from the first farmers. On the other side, the cultural diffusion (horizontal transmission) of farming would represent an adaptation of already established societies, only decreasing the general extinction rates. Assuming that the selection of farming should be at the society level, we aim to generate a society based modelling simulation to first confirm if there could be differences among phylogenetic trees exposed to different cultural spread modes (Cultural and Demic diffusion), as we suspect, and what would be such differences.

Discussion points:

- 1) Do we have any way of identifying the geographic location of different nodes in the tree?
- 2) How good are our time estimates for the time between nodes?
- 3) Agriculture: will societies that are successful with ag be more likely to take over another ag society because they have ag?

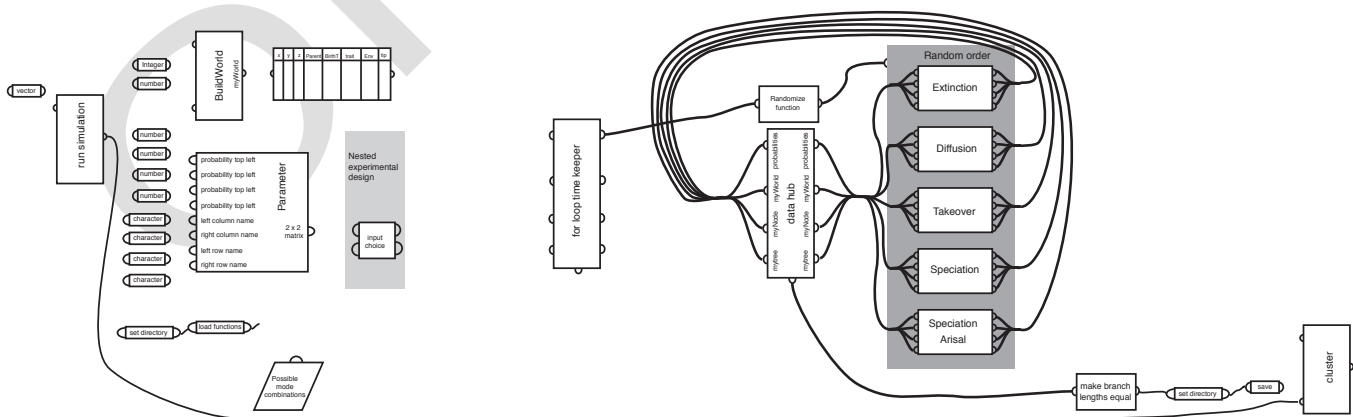
Experimental design:

- 1) We compete 4 different scenarios in a direct inference comparison. The for competing models are: 1) Speciation, Extinction, background takeover, and Arisal (combo #25); 2) Speciation, Extinction, background takeover, Arisal and Diffusion (combo #29); 3) Speciation, Extinction, background takeover, Arisal, and Takeover (combo #28); 4) Speciation, Extinction, background takeover, Arisal, Diffusion, and Takeover (combo #31).
- 2) For each scenario, a series of phylogenetic and spatial metrics will be calculated.
- 3) We will choose the parameters of each of the scenarios using an Approximate Computational Bayesian (ABC) approach that maximize the likelihood of the simulation phylogenetic and spatial metrics to the metrics calculated for the real dataset.
- 4) Each of the four calibrated scenarios will be repeat 1000 times. This will allow us to identify the key metrics that change among the scenarios and what scenario best fits the real dataset.

Model overview:

The model for this project is a sequence of three separate modules. The first module produces simulated data by running the simulation. These data are written to the hard drive at the end of the simulation. The second module loads those files and analyzes the data they produced. The third module plots the analysis outputs from the second module.

We are designing the model using the R(CRAN) programming language. This is a graphical schematic of our algorithm design.



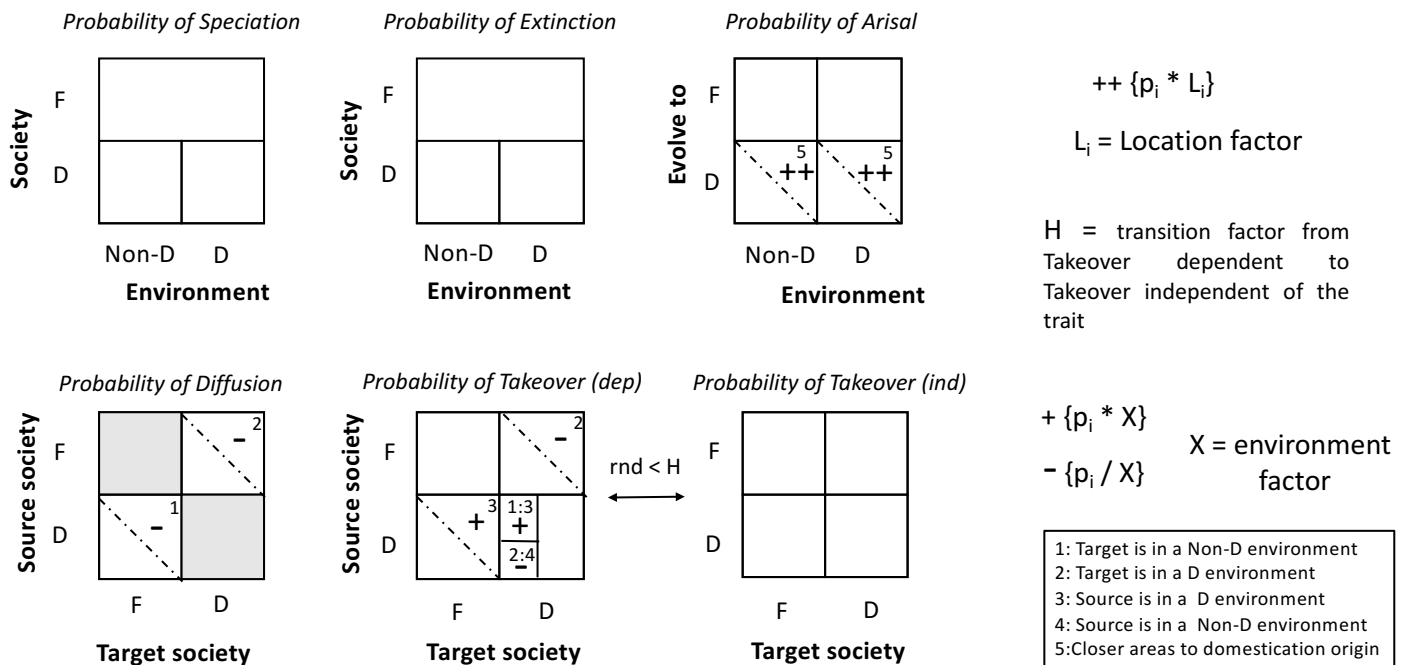
Methods:

The model is divided into three modules that are run independently so that there are natural breaks in running the model. Module one accepts input parameters and produces a simulation from those inputs. Module one produces a realized landscape and the associated phylogenetic tree for that realization. Module two accepts simulation replicates (landscapes and trees) and performs a large series of analyses on those data. The output from module two is a list of summary statistics from those analyses. Module three accepts the list of summary statistics produce by module two and plots those analyses to a visualization dashboard for visual quality control and science communication.

Simulation module – (this is where replicates happen)

- a) Define input parameters: This is a phenomenological model where the probabilities of different events are defined by the user.
 - a. The total number of time steps (T) in which a simulation will run is defined *a priori* by the user.
 - b. All the parameters involving the probabilities related to each action module are decided at the beginning of the simulation by the user.
 - c. Any module can be turned off at the simulation by setting its probabilities to zero.
 - d. In locations that are not suitable for farming, it is X times harder to diffuse farming. It is also X times harder to diffuse foraging to farming societies that are in suitable locations for farming. Farming societies have X times higher probability of takeover and resistance from takeover when in suitable locations for farming. Being X defined *a priori* by the user.
- e.

Parameters space



- b) Randomize the order that processes (speciation, extinction, arisal, takeover,

diffusion) will happen. Different experimental models use different combinations or numbers of these processes to test different hypotheses, but the order they are applied each time step is always randomized.

- a. All the modules will happen at every step. The modules order of action is always shuffled at the beginning of every time step.
 - b. Inside the modules, the actions will involve all existing societies at a specific time step (N_i). However, for the Speciation, Takeover and Diffusion modules, the actions cannot happen for all the societies at the same time, because what happen to one society may affect the behavior of other societies in the same module. So, the simulations inside these three modules happen sequentially among societies within sub-time steps ($T_{sub} = 1 / N_i$), such sequence is decided by shuffling the order of societies at the begging of each module. Societies that get extinct during the Takeover module are excluded from the sequence.
- c) Run the modules, which operate according to the following rule sets:
- a. Landscape –
 - i. A single society can only occupy one spatial location. The locations are determined by the occurrence points in the real dataset (N) [Note that any set of occurrence points is accepted to make the model flexible]. Hence, the maximum number of societies is restricted by the number of locations.
 - ii. The locations are connected by the 7 closest neighbors that keep all the connections symmetric. Previous research shows that his number represent better the overall connection among societies in this specific dataset.
 - iii. All the societies have one trait associated, either foraging or farming.
 - iv. Each location has one binary value indicating whether a society can farm or not (environmental conditions), determined by previous studies using real data.
 - v. The model starts with one foraging society. The location of the starting society can be pre-determined by the user or randomly assigned by the model.
 - vi. The spatial table, will include for each location their ID, latitude, longitude and environmental conditions. If a society is occupying the location at the time T , the table will also include the society trait and tip label (corresponding to the phylogenetic tree).
 - b. Phylogenetic tree
 - i. During the simulation, all the evolutionary relationships are tracked using a phylogenetic tree.
 - ii. The branch lengths (BL) correspond to the sum of time steps from the origin of a node (k) or tip to the current time of the simulation ($BL_k = T_i - T_k$).
 - iii. During a speciation process (through the Speciation or Takeover module) one society will bifurcate into two new societies in the phylogenetic tree (so that the tree is always binary).

- iv. All societies extinctions (through the Extinction or Takeover module) will be pruned from the phylogenetic tree.
- v. At the end of each time step, all the tip lengths are extended by adding the current time step so that the tree is always ultrametric.
- vi. The output phylogenetic tree will include all existing societies at the time T (no extinct society will be included). The final phylogenetic tree shall be binary and ultrametric.

c. Speciation –

- i. Every society has a chance to expand into un-colonized neighbor locations at each time step. Because societies can only occupy one location, a speciation occurs when a society expand into a new location, and a new society is created (Speciation module). The decedent society will keep the same trait of the parent society.

d. Extinction –

- i. Every society has a chance to go extinct at each time step.

e. Arisal –

- i. Every society has a chance to change their trait from foraging to farming at every time step.

f. Diffusion –

- i. Every society has a chance to diffuse their trait to other neighbor society at every time step

g. Takeover –

- i. Every society has a chance to takeover a neighbor society (with speciation of the source society and extinction of the target society) at every time step (Takeover module). Takeover only happens when all neighbor cells are occupied [Note that because the Speciation module requires empty neighbors, while Takeover only occurs in the absence of them, only one of these two modules will happen at each time step per society. This avoids an artificial inflation of the speciation rate with Takeover]. Farming societies will always prefer to invade locations suitable for farming if they are available.

- d) Track the development of both spatial pattern and phylogenetic pattern through time.
- e) Write a file with the spatial and phylogenetic data that the simulation produced. The file name for each file embeds the parameters used for that simulation including, the probabilities used for each module, the replicate number of each simulation, and the number of time steps each simulation ran for.

Data analysis module – (input multiple replicates)

- a) Load files from simulation output and extract data
- b) Subset data from files to analyze each model results separately.
- c) There are 5 subcategories of analysis.
 - a. Spatial analyses

- b. Richness tree metrics (sum of tree units)
- c. Divergence tree metrics (mean of tree units)
- d. Regularity tree metrics (variance of tree units)
- e. Macro evolutionary tree metrics (rate and rate change in tree units)
- d) Write a file with all of the output values.

Data plotting module – (input 1 data analysis file from each model type)

- a) Load analysis files
- b) Subset analysis files for plotting
- c) Use layout() function to make a display dashboard
- d) Currently taking the form of 4 pages
 - a. Page 1-Overview
 - i. a panel with the input parameters used
 - ii. a panel with the important information about the run (e.g. date, authors, number of time steps run, and number of tree tips replicated)
 - iii. a map for each model
 - iv. a tree for each model
 - b. Page 2 – tree metrics
 - i. Four rows showing the four units available for comparing trees to each other (e.g. branch length, pairwise distance between tips, evolutionary distinctiveness, and tree topology)
 - ii. Four columns show the four levels of summary statistics for each of those units (e.g. raw units, sum of units, mean of units, and variance in units).
 - c. Page 3 – Evolutionary rates
 - i. Five rows with the type of rate to be viewed (e.g. lineage through time with variance in time, lineage through time with variance in lineage number, speciation vs. extinction growth rates, distribution of speciation vs. extinction rates, and net diversification)

Important dates:

Idea proposed to PI: 6 June 2016

Start date: 13 June 2016

Establish Github: 13 June 2016

Start documentation: 20 June 2016

Finish fully running draft of simulation module: 29 July 2016

Apply for funding to continue project: 15 (chemistry?) and 18 (Templeton foundation) August 2016