**Methods. 3000 words max (Nature format)**

To disentangle the relative contributions of cultural diffusion and takeover to the modern

distribution of cultural traits, we simulated the evolution and global spread of agriculture. Our model simulates an initial colonization of the world by hunter-gathers and subsequently allows agriculture to arise and spread from a limited number of arisal locations. Different model variants allow for different mechanisms to drive these processes. The most basic variant simulates the spread of agriculture solely via vertical transmission. More complex variants allow for the addition of diffusion, takeover, or both. The outputs of these simulations are then summarized through 11 highly-targeted phylogenetic and spatial statistics (each model variant is replicated 100K times). These resulting distributions are then used to classify the realized historical trajectory according to its probability of belonging to one of those distributions.

The model is divided into three modules that are run independently but in sequence so that there are natural breaks in running the model. Module one accepts input parameters and produces a simulation from those inputs. Module one returns a simulated landscape and the associated phylogenetic tree for that simulation. Module two accepts landscapes and trees objects (simulated or real) and calculates a targeted list of summary statistics to describe those simulations. Module two returns a list of summary statistics from those provided worlds. Module three accepts all of the simulated and real world summary statistics from module 2 and returns estimates for which of the four simulation types the realized human history more closely resembles.

**Module 1: Simulate the spread of a cultural trait**

Simulations are initiated as two data storage objects, a spatial layout object and a phylogenetic tree object, that are updated each time step as the simulation progresses. The spatial layout object is a set of ~6000 points marking the centroids of all extant languages that are linked together using a symmetrical full connected network defining available neighbors for each point. This object is stored as a matrix where each row contains the latitude, longitude, a list of neighbors for that point, tip labels, and trait identification for matching locations to the phylogenetic tree. Each of these points is then assigned an agricultural habitat suitability score based on the number of domesticated species found at a particular centroid location compared to the global pool of domesticated species. This integer value is converted to a binary indicator (1, 0) of agricultural suitability (good for farming or not) using the threshold value that maximized the relationship between suitability and the observed subsistence method. Each location point was connected to the 7 nearest points (this number of neighbors better controlled for the spatial autocorrelation in our observed farming data; Vilela et al. *in prep*). However, using 7 nearest points created an asymmetrical connectivity between neighbors (i.e. a point A is connected to B but B is not connected to A). We then forced a symmetrical connection between all points, which occasionally increased the number of neighbors for some points (max of 15). Societies can only be assigned a single trait state (agriculturalist or foraging society) and a single state of environmental suitability (1, 0). Each simulation starts by choosing a single random location to assign as the first foraging society and all other societies radiate from that single origin as the simulation progresses. The probability of agriculture arising from a foraging society is greatest within a 1000m buffer of the known origins of agriculture.

The phylogenetic tree object is a matrix object where each row contains the name of a single tip or node, and its correspondent parent node. During the simulation, all the evolutionary relationships are continuously tracked and modified using this phylogenetic tree object. The branch lengths created during the simulations correspond to the sum of time steps from the origin of a node or tip to the current time of the simulation. During a speciation process, one society will bifurcate into two new societies so the tree is always binary and non-reticulated. Extinct societies are pruned from the phylogenetic tree each time step, so the output tree will include all extant societies but no extinct societies. At the end of each time step, all the tip lengths are extended to the current time step so that the tree is always ultrametric. The final phylogenetic tree will always be binary and ultrametric. When the simulation is finished the matrix used to track the phylogenetic relationships is then transformed into a *phylo* object(see package ape).

Replicate simulations each start with identical initial conditions and then they are subjected to a series of simulation rule sets to modify both the spatial and phylogenetic objects each time the simulation progresses. Each time step, these objects are modified a maximum of 4 times by up to 5 separate and randomly ordered simulation rule sets named: arisal, speciation, extinction, diffusion, and takeover. Each of these modules simulates a single process contributing to the spatial spread and persistence of a society and/or its trait. The number of modules used is defined by the model type run. The basic model uses only the arisal, speciation, and extinction modules. The diffusion model adds the diffusion module to the set of basic modules, while the takeover model adds the takeover module to the set of basic modules instead. The full model includes all 5 modules. These modules are randomly reordered each step as an experimental control because there is no intrinsic hierarchical relationship between them and we didn’t want to introduce a bias artificially assuming an ordered relationship between these causes.

Rule sets within each module cannot be applied to all centroid locations at the same time because the modification of one point can interfere with the modification of the next, so points are sampled without replacement and modified one by one. This sampling scheme randomly reorders the sequence of spatial points each time a new rule set is applied and prevents spatial bias when applying rule sets. Since the speciation rule set requires empty neighbors, while the takeover rule set only occurs in the absence of them, only one of these two sets will happen at each time step per society. This is necessary to avoid artificially inflating the total speciation rate.

Each rule set function accepts a 2x2 matrix of probabilities defining how those rules are applied throughout a particular replicate simulation. These matrices are created before the simulation begins by drawing random numbers from a uniform distribution to populate each of the boxes in 5 separate 2x2 matrices. The rows of these matrices define the trait identity for the source population (forager of domesticator) that is trying to expand into a neighboring centroid. In the function defining the basic model (speciation, extinction, and arisal rule sets), column values of these matrices define the environmental suitability for areas targeted by expanding societies. In the functions expanding on the basic model (takeover, diffusion, and takeover with diffusion rule sets) columns describe the target societies’ susceptibility to receiving a trait through diffusion or takeover from the source society. The cells of these matrices define the individual probabilities for each possible outcome (e.g. domesticator in good environment for domestication, domesticator in a bad environment for domestication, a forager in a good environment for foraging, and a forager in a bad environment of domestication). These numbers are redrawn for each replicate simulation but they stay the same across time within a given replicate. The values in each box defines a probability threshold for triggering a rule set or not. Within a rule set, a random number is drawn from a uniform distribution. If that drawn value is less than the value defined in the probability matrix, then the rule set takes its defined action. If the drawn value is greater than the value provided in the matrix, then the rule set action is skipped for that time step.

**Rule sets:**

1. Arisal – Every society has a chance to change their trait from foraging to farming at every time step.
2. Speciation – Every society has a chance to expand into un-colonized neighbor locations at each time step. Because one societies can only occupy one location, a speciation event occurs when a society expands into a new location and a new diaspora is created. The decedent society will keep the same trait of the parent society.
3. Extinction – Every society has a chance to go extinct at each time step.
4. Diffusion – Every society has a chance to diffuse their trait to occupied neighbors at every time step
5. Takeover – Every society has a chance to takeover a neighbor society to cause speciation of the source society and extinction of the target society at every time step. Takeover only happens when all neighbor cells are occupied. Farming societies will always prefer to invade locations suitable for farming if they are available.

**Module 2: Calculate summary statistics**

The spatial layout object and phylogenetic tree are both passed directly to the second module without modification. These input objects are immediately trimmed to match the 1200 societies in the realized history dataset. Those trimmed objects are then passed to 11 individual functions which each produce a single summary statistic. Those statistics are then aggregated into a single list and returned as the output. The statistics used in this module are organized according to their basic unit of measurement and the type of data summary they represent. There are 6 subcategories of analysis:

1. Branch Lengths (basic unit, sum of unit, mean of unit)

The basic unit of measurement for this section is the branch length, which is the unit length between each node or between a node and a tip. Phylogenetic diversity (*PD*) is the summation of all branch lengths connecting species together from root to tip (Faith 1992). PD is a richness measure; it describes how much evolutionary history is associated with a set of tips. Average phylogenetic diversity (*avPD*) (Clarke and Warwick 2001) is the mean branch length calculated as *PD* divided by the total number of tips in the tree.

1. Pairwise distance between tips (basic unit, sum of unit, mean of unit, variance of unit)

The basic unit of measurement for this section is the pairwise distance. This is the patristic distance, meaning the sum of the branch lengths following the shortest distance between two tips in a tree, implemented as a distance matrix where every tip is compared to every other tip. The sum of all pairwise distances, *F*, is more formally called ‘Extensive quadratic entropy’ (Izsak and Papp 2000). Just as it was with branch lengths, this is a richness measure and, accordingly, describes the amount of distance between compared tips. The divergence metric for pairwise distance is mean pairwise distance, *MPD* (a.k.a. *AvTD*, and *Δ+*), which is the mean inter-species distances between species. (R. Clarke and Warwick 1998; Webb et al. 2002; Webb, Ackerly, and Kembel 2008; Kembel et al. 2010). Variance in pairwise distances, *VPD* (a.k.a. *VarTD and Λ+*), is a regularity indices. (Clarke and Warwick 2001) Variance is relative to tips, not to total branches.

1. Phylogenetic isolation of each tip (basic unit and sum of units)

A phylogenetic isolation index represents the relative isolation of a given species within a phylogenetic tree. Several indices have been proposed so far but we focus here on the evolutionary distinctiveness index called ‘Fair Proportion’ as proposed by Redding (2003) and Isaac et al. (2007). The sum of phylogenetic isolation units is evolutionary distinctiveness (*ED*), which is a richness indices.

1. Tree topology (root-to-tip symmetry of the full phylogeny)

The most trusted index for the root-to-tip symmetry of a phylogenetic tree is the gamma index, γ, which characterizes the distribution of branching events within the tree. Trees with γ < 0 have relatively longer branches towards the tips of the phylogeny, whereas trees with γ > 0 have relatively longer inter-nodal distances towards the root of the phylogeny. (Pybus and Harvey 2000)

1. Phylogenetic signal

The extend of phylogenetic closer species to have similar traits is measured as the Phylogenetic signal. The most robust measure of phylogenetic signal for binary traits is the *D* proposed by Fritz and Purvis (2010). D is based on the sum of differences between pairs of sister-clade. I varies from 0 (Brownian motion expectation) to 1 (random expectation).

1. Macro evolutionary tree metrics (rates and rate changes).

Using a maximum likelihood function to fit a birth-death model to a phylogenetic tree branching times it is possible to estimate the rates of extinction (*μ*) and speciation(*λ*) for a phylogenetic tree (Nee et al. 1994). Diversification rate is then obtained by subtracting extinction from the speciation rate (*μ* - *λ*). Clade turnover is obtained dividing the extinction by the speciation rate (*μ*/*λ*). The trait specific diversification rate and clade turnover was obtained using the multistate speciation and extinction model (FitzJohn 2012).

1. Spatial analyses

The join count test was used to measure the spatial autocorrelation in each trait. The statistic is calculated subtracting the number of observed connections between locations with the same traits (O) by the expected value of connections (E), and finally dividing this value by the expected standard deviation (Cliff and Ord 1981).

**Module 3: Classify outcomes**

Module three accepts the list of summary statistics produced by module two and uses a random forest machine learning algorithm to first classify all of the replicate simulations according to model type and once the algorithm is trained use the decision trees created during that classification to categorize the realized linguistic phylogeny estimated from linguistic, and archeological data. Module three returns a list ranking the value each summary statistic brings to classifying each simulation, a matrix describing how good each decision tree in the random forest was at classifying each simulation, and a list recording the classifications that all of the individual decision trees assigned to the realized historical phylogeny. This final count represents the probability attributed by the random forest that the realized historical phylogeny came from each of the proposed models.

**Running the simulation using High-performance computing clusters**

We ran the simulations from 17 January 2016 to 15 May 2016 across 4 separate high-performance computer cluster at three different institutions. –ran 100k times….