

Supporting Information

Premo and Hublin 10.1073/pnas.0809194105

SI Text

Ancient Structured Populations (ASP) is programmed in Objective-C and makes use of the Swarm libraries (1). The commented source code is freely available upon request. This model description follows the ODD protocol (2).

Purpose. The purpose of this model is to understand how culturally mediated migration, a social behavior practiced by individuals living in groups, affects the level of genetic diversity displayed by the metapopulation. We use this spatially explicit, agent-based model to investigate whether culturally mediated migration has a significant effect on average gene diversity across neutral loci in a (nearly) constant sized population. We run the model with and without differential reproductive fitness to demonstrate that this effect is due to group-level selective sweeps.

State Variables and Scales. The model comprises 4 hierarchical levels, from lowest to highest: the gene/cultural trait, the individual, the group, and the metapopulation. An individual contains many genetic loci and cultural traits, a group contains many individuals, and the metapopulation is composed of many groups.

Genes and cultural traits occupy the lowest level in the model. Each gene is characterized by the following state variables: fitness, type (neutral or selective), and allele size [this value refers to the length of the short tandem repeat (STR)]. Each cultural trait is characterized by the following state variables: fitness, type (neutral or selective), and expression (i.e., variant). Each individual contains a haplotype (a list of 1,000 gene agents) and a cultural repertoire (a list of 100 cultural trait agents). In addition, each individual is characterized by the following state variables: death probability, age at first reproduction, migration rate, cultural similarity threshold, and fitness. Each group is characterized by the following state variables: maximum group size (φ) and search radius (τ). In addition, each group has a cultural identity, which is a dynamic list of all of the expressions that its members display for each trait in the cultural repertoire. Although individuals' cultural repertoires cannot change after they have passed through social learning and innovation, both of which occur only once (at birth), a group's cultural identity changes each time an agent joins it (via birth or immigration) or leaves it (via death or emigration). The metapopulation refers to the collection of all groups and, therefore, all living individuals. The metapopulation is characterized by the following variables: average gene diversity (H), average gene diversity within subpopulations (H_S), average gene diversity between subpopulations ($H-H_S$), population size, number of groups with members, average group size, average generation length, and cumulative number of migrations.

The temporal scale of the model is defined as arbitrary time steps. The average length of a generation (i.e., the mean age at first successful reproductive event of all extant individuals) is ~ 3 time steps. Each simulation run lasts 10,000 time steps, or roughly 3,000 generations of individual agents. The model is spatially explicit, and this is one of the central reasons why we chose to use the agent-based method to operationalize it. The spatial environment is a torus of regularly spaced and shaped square cells. Each cell contains a single group, which can contain between 0 and φ individuals. Although cell size does not equate to an empirical value, one can think of it as the area needed to support the subsistence activities of a single social group within

an idealized, homogeneous landscape. It is important to note that, at any given time during a simulation run, many of the cells contain no individuals. Empty cells represent open areas that can be colonized by the descendants of reproductively successful individuals following group fission.

Process Overview and Scheduling. During each time step, each individual goes through 3 processes in the following order: migration, reproduction, and death (Fig. S1). The order in which individuals go through each routine is randomized (i.e., their order is shuffled) at the beginning of each time step. As a result, even though each individual completes each of the processes before any individual begins the next, by randomizing the queue of individuals we emulate what would occur concurrently in nature while ultimately updating each individual's state variables asynchronously. Along the way, the model calculates values that the individuals will need in their submodels, like their relative individual fitness and the population size coefficient (see *Submodels*). Only after individual and group agents have passed through all of the submodels do we calculate measures of genetic diversity at multiple levels. Data collection occurs at an interval of 50 time steps.

Design Concepts. Adaptation. In selective simulations, evolution occurs at the level of the metapopulation due to selection acting on differential fitness associated with heritable variation in individuals. Those individuals that possess the highest relative fitness temporarily occupy positions atop adaptive peaks on a dynamic fitness landscape. When CST is high, CMM effectively raises a set of individuals atop the same adaptive peak. This set of individuals is defined by membership in the same group. With CMM, the cultural identity of this group serves as a buffer to gene flow when CST is high and/or when groups are very different from one another culturally, such that its members enjoy for many successive generations the fitness benefits of the suite of favorable mutations they accumulated serendipitously.

Emergence. Population-level characteristics emerge from the uncoordinated behaviors of individuals, which serve as containers for cultural and genetic variation. The average gene diversity of the entire population is affected by the way in which relative fitness is partitioned within and between groups, which in turn is affected by the movement of individuals between groups. With time, low gene flow leads to internally homogeneous but externally heterogeneous groups, which yields low variance in relative fitness among members of the same group but high variance in relative fitness among members of different groups. The movement of individuals between groups is affected by geographic distance, the severity of the cultural similarity threshold, and the cultural identities of groups (which, of course, are dependent upon their unique, dynamic, and mutually exclusive sets of members).

Sensing. Each group is assigned a search radius (τ), which defines the size of the surrounding Moore neighborhood of cells in which its members can search when trying to locate a teacher or move to a new group. If $\tau = 1$, then this neighborhood includes 8 cells (9 when including the cell in which the individual is located). If $\tau = 2$, then this neighborhood includes 24 cells (25 when including the cell in which the individual is located). Group agents also use τ to find a cell in which they can place half of their members following group fission.

Stochasticity. Stochasticity is included throughout the model. Examples include (but are not restricted to): chances at migra-

tion, reproduction, and death are all interpreted as probabilities; the sequential order in which individuals go through their submodels is randomized each time step; the sequential order in which groups go through their submodel is randomized each time step; teachers are chosen at random from all individuals living within τ ; and the group to which an individual migrates is chosen randomly from the set of groups that satisfies CST. Of course, this list is not exhaustive, but it illustrates that stochasticity is incorporated into many aspects of this nondeterministic model.

Observation. For model analysis, we collect data at the individual level and the metapopulation level at an interval of 50 time steps. We record each individual's current group, cultural repertoire (i.e., a list of variants for all cultural traits), and haplotype (i.e., a list of STR lengths for all loci). At the metapopulation level, we collect mean H , mean H_S , number of individuals, number of groups with members, average group size, and cumulative number of migrations. Finally, we collect data pertaining to each genetic mutation that increases a locus's fitness to a value that is greater than or equal to 150% of its previous fitness value. We collect the locus number, new allele size (i.e., STR length), old fitness value, new fitness value, and group ID for each of these so-called "big" mutations.

Initialization. Each simulation run is initialized with a metapopulation that is highly genetically and culturally diverse, both at the within-subpopulations and metapopulation levels. In other words, total diversity is high and between-group diversity is low in populations at the start of each simulation run. At the start, each group is randomly assigned to an unoccupied cell. Then, individuals are evenly distributed among the starting groups. Cultural variants and allele sizes are randomly assigned to each individual by drawing randomly from a uniform distribution of integers ranging from 0 to 20 for each trait or locus. The state variables of groups and individuals are then set according to user-defined values (Table S1).

Submodels. Calculating relative individual fitness. Our model deals only with relative individual fitness—fitness values that are scaled relative to the highest combined fitness value in the population during each time step. Each individual's absolute cultural fitness (simply the sum of the fitness values at all selective cultural traits) is first scaled relative to the highest absolute cultural fitness in the population. Next, each individual's absolute genetic fitness (simply the sum of the fitness values at all selective genetic loci) is scaled relative to the highest absolute genetic fitness in the population. This is done separately for cultural traits and genetic loci to allow for cultural and genetic fitness to contribute equally to one's relative fitness. Relative individual fitness is then calculated by scaling each individual's combined fitness (relative cultural fitness + relative genetic fitness) to the highest combined fitness value. If an individual were to possess the highest (absolute) genetic fitness and the highest (absolute) cultural fitness, then, after scaling, its combined fitness would be $1 + 1 = 2$, and its relative individual fitness would be 1, because 2 is the highest possible combined fitness value ($2/2 = 1$). To maintain a nearly constant population size, all relative individual fitness values are weighted by a coefficient that is sensitive to the difference between the current population size and the target population size. If the current number of individuals is less than the target population size, then the coefficient is calculated as follows: [(target population size – current population size) / expected number of offspring], where the expected number of offspring is simply the sum of all living individuals' relative fitness values. If the current population size is greater than or equal to the target population size then the coefficient is set equal to 0, resulting in no reproductive events

during that time step. Relative individual fitness does not affect survival; it only affects reproduction.

Reproduction. Reproduction is sexual in the sense that it requires 2 individuals, but for the sake of simplicity we do not model 2 different sexes (male and female). If a number drawn randomly from a uniform distribution between 0 and 1 is less than or equal to an agent's relative fitness, it will randomly choose a partner from within its current group. Then, if another number drawn randomly from a uniform distribution between 0 and 1 is less than or equal to the partner's relative fitness, the couple produces 1 offspring, else they do not have an offspring.

Genes and Mutation. Genetic material is modeled as if contained on a single chromosome composed of 1,000 partially linked STR loci. At birth, genetic loci undergo mutation ($\mu = 0.001$ per locus) according to the single step mutation model with the additional assumption that all mutations result in a value of 1 being added to the allele size (i.e., STR length). A proportion (η) of the genetic loci are neutral. Their fitness value is always 0, regardless of allele size. At the start of each simulation run, neutral loci are spread randomly throughout the artificial genome. Each locus that has a selective effect has its own unique nonlinear fitness function, which is the same for all individuals. When the STR length of a selective locus increases by 1, it is assigned a new real number, which is drawn randomly from a uniform distribution between 0 and 1, as its fitness value. Note that a change in STR length could result in a positive or negative change in the locus's fitness value. The algorithm for the fitness function is designed so that the values drawn randomly from a uniform distribution are consistent across individuals for any combination of locus number and allele size. This means that an allele size of 5 at locus 2 has the same fitness value for any individual that displays it, just as an allele size of 6 at locus 2 has the same fitness value for any individual that displays it (even though both of these unique values are chosen randomly from a uniform distribution between 0 and 1). This also means that an allele size of 5 at locus 2 is highly unlikely to have the same fitness value as an allele size of 5 at locus 3. Note also that there is no interaction between genetic loci (i.e., epistasis); the fitness function at each locus is independent of all others.

Genetic Recombination and Transmission. A hypothetical pairing between 2 agents ($P1$ and $P2$) illustrates how genetic recombination is executed in this model. We start with the haplotypes of 2 parents. For the sake of illustration, these are only 10 loci long:

Haplotype $_{P1}$ [1 3 2 4 2 1 1 4 1 1]

Haplotype $_{P2}$ [2 1 1 1 2 5 2 5 2 1]

Next, we randomly choose the position at which each haplotype will be split and recombined. Imagine that our randomly chosen locus for this recombination is the sixth locus. This position is marked by an asterisk below.

Haplotype $_{P1}$ [1 3 2 4 2 1* 1 4 1 1]

Haplotype $_{P2}$ [2 1 1 1 2 5* 2 5 2 1]

After recombining the segment of Haplotype $_{P2}$ that follows the asterisk with the segment of Haplotype $_{P1}$ that precedes the asterisk (and vice versa), we have created two new haplotypes ($O1$ and $O2$) from $P1$ and $P2$ via sexual recombination.

Haplotype $_{O1}$ [1 3 2 4 2 1 2 5 2 1]

Haplotype $_{O2}$ [2 1 1 1 2 5 1 4 1 1]

The offspring of $P1$ and $P2$ is assigned one of these new haplotypes with equal probability. In this case, imagine it is assigned Haplotype $_{O1}$. Finally, each locus mutates with probability μ . The asterisk denotes mutation.

Haplotype $_O$ [1 3 2 4 3* 1 2 5 2 1]

The offspring will carry haplotype $_O$ for the rest of its life.

Cultural Traits, Cultural Transmission, and Innovation. Each individual's cultural repertoire is composed of 100 traits, a proportion (ϵ) of which are neutral. Each offspring "learns" all of its neutral variants from one of its parents. The remaining cultural traits have fitness effects. The fitness function of each selective cultural

trait is described by a simple linear relationship with a slope of 0.001. After adopting the more fit of the variants displayed by its parents (i.e., the higher integer) for each selective cultural trait, each offspring has an opportunity to learn a better technique from a “teacher.” An offspring randomly chooses 1 individual from those that live within the search radius of its own group (including those that live within its own group with the exception of itself) to serve as its teacher. After the offspring has the chance to learn from both of its parents and a teacher, each of its cultural traits (selective and neutral) undergoes innovation with probability ι . Like genetic mutation, cultural innovation follows a single stepwise model. However, for the case of neutral traits, it does not follow the additional assumption that the change must be a positive value. In other words, innovation has an equal probability of increasing or decreasing (by 1) the integer that represents the variant of a neutral cultural trait, but innovation can only be positive for a selective cultural trait. Thus, innovations to selective cultural traits always increase the absolute fitness of that trait by 0.001. Here is an illustration of how cultural transmission is operationalized in the model. Note that the first 3 traits are neutral and the last 7 are selective.

Step 1: Transmission from parents to offspring.

Cultural Repertoire_{P1} [4 3 2 4 5 7 3 2 1 9]

Cultural Repertoire_{P2} [5 7 7 3 5 8 3 1 0 5]

Cultural Repertoire_O [4 3 2 4 5 8 3 2 1 9] (Offspring adopts neutral traits from Parent₁ in this case)

Step 2: Transmission of selective traits from randomly chosen teacher to offspring.

Cultural Repertoire_T [9 6 0 5 4 4 5 1 1 4]

Cultural Repertoire_O [4 3 2 5 5 8 5 2 1 9] (after adopting “better” variants at selective loci from the teacher)

Step 3: all traits undergo innovation at probability ι .

Cultural Repertoire_O [3* 3 2 6* 5 8 5 2 1 9] (* denote innovations)

The offspring will carry this cultural repertoire for the rest of its life.

Cultural Identities of Groups. Each group maintains a list of all of the cultural variants its members display at each cultural trait. Each group’s cultural identity is the sum of its members’ cultural variants. Cultural identities are dynamic; they are updated whenever an individual joins or leaves a group. For example, imagine group *A* has 2 members, whose cultural repertoires are only 2 traits long:

Cultural Repertoire₁ [1, 2]

Cultural Repertoire₂ [3, 4],

where the integers represent cultural variants. Given these 2 members, the cultural identity of group *A* would be:

Cultural Identity_A [[1, 3], [2, 4]].

Imagine that individual 1 were to leave *A* and individual 3 were to join *A*:

Cultural Repertoire₂ [3, 4]

Cultural Repertoire₃ [7, 8].

Now, the cultural identity of group *A* would be:

Cultural Identity_A [[3, 7], [4, 8]].

Migration. During each time step, each individual attempts to migrate to another group with probability ν . However, to emigrate from one’s group, the prospective migrant must find a suitable group within its group’s search radius (not including its own group), one which it deems as culturally similar enough to allow for a culturally mediated migration event. If no suitable destination group can be found, the individual will remain where it is for the duration of that time step. Let *i* represent a prospective emigrant, and *I* the prospective emigrant’s current group. A randomly chosen individual *j*, from each group *J* within *I*’s search radius τ , serves as a representative of its group’s cultural identity. For all *j* and for all cultural traits *l*, *i* compares the variant *j_l* to *I_l*, the set of variants expressed by *I* at *l*. If the proportion of *l* in which *j_l* exists in *I_l* exceeds CST, then *i* adds *J* to its list of possible destination groups. When multiple groups occupy this list, *i* immigrates to one chosen randomly.

Death. Each time step, all individuals experience death with probability δ . When an individual dies, it is removed from the simulation immediately. The dead take their cultural repertoires and haplotypes with them, so that they are no longer included in groups’ cultural identities or in calculations of gene diversity found within groups, between groups, and in the metapopulation.

Conservative Model Assumptions. Our model assumptions are purposefully conservative. When they have an effect, it is of maintaining or increasing genetic variation within groups rather than between them. These conservative assumptions include: each agent can emigrate an unlimited number of times during its lifetime; learning occurs only once per lifetime (upon birth); social learning is bounded by the length of the search radius, but cultural transmission is unbiased within that radius; prospective migrants base their assessment of cultural similarity on just 1 individual sampled from each possible destination group; an agent chooses a mate randomly from within its group each time it has a chance to reproduce; group “extinction” occurs only if the last member emigrates or dies; there is no direct competition between groups, such as war; migrants choose their destination group randomly from the set of groups for which they calculate a degree of cultural similarity greater than CST; genetic loci undergo single stepwise mutation in one direction only so as to avoid back-mutations; and when a group expands into a cell that is already occupied, the current residents are immediately incorporated into the gene pool rather than dispatched or treated as pariahs. Each of these assumptions has the effect of maintaining or increasing genetic and/or cultural diversity within groups. Some of these assumptions also increase average gene and/or cultural diversity at the metapopulation level.

Obtaining the Source Code. Of course, the best description of any computer model is the program itself. Please contact the authors to obtain the Objective-C source code for ASP, which requires the Swarm object libraries.

1. Minar N et al. (1996) The swarm simulation system: A toolkit for building multi-agent systems. *SFI Work Pap* 96–06-042.

2. Grimm V et al. (2006) A standard protocol for describing individual-based and agent-based models. *Ecol Model* 198:115–126.

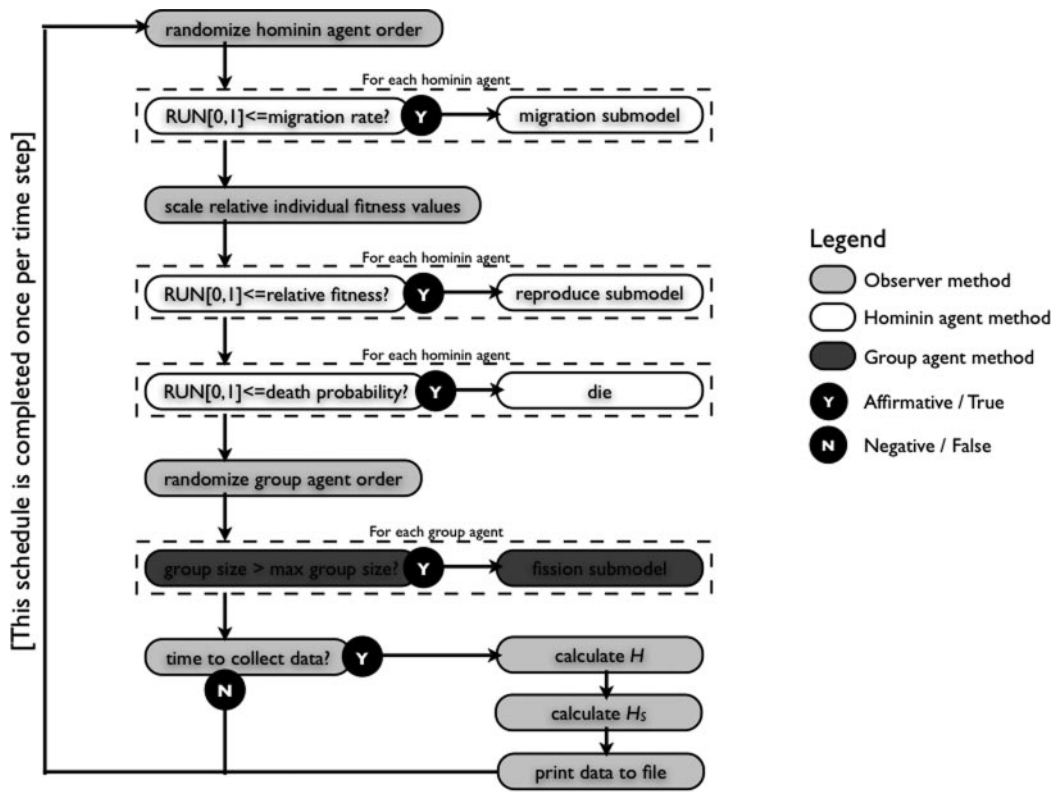


Fig. S1. Schematic diagram of a single time step in ASP. "RUN[0,1]" refers to a real number drawn randomly from a uniform distribution between 0 and 1.

Table S1. Parameter values used to initialize the simulations reported in the text

Parameters	Value(s)
Number of starting hominins	1,000
Minimum number of empty cells	100
Number of starting groups	50
Target population size	1,175
Maximum group size (φ)	50
Search radius (τ)	1
Death probability (δ)	0.15
Mutation rate (μ)	0.001
Proportion of neutral loci (η)	0.98
Haplotype length	1,000
Cultural repertoire length	100
Experiment duration (time steps)	10,000
Migration rate (ν)	0.05
Cultural similarity threshold (CST)	0, 0.7, 0.8, 0.9, 0.99
Innovation rate (ι)	0.0001, 0.001, 0.01
Proportion of neutral traits (ε)	0.3, 0.7
Evolution	0 (neutral), 1 (differential fitness)
Random number seeds	480207697, 387946597, 562954596, 706570560, 153768594, 107032492, 291582303, 847543645, 223575322, 515092991, 796702084, 795649783, 695459265, 165075607, 495867351, 294230701, 303285192, 546345748, 635758753, 199290515