Modeling the genealogy of a cultural trait

Elliot Aguilar [a,b,*], Stefano Ghirlanda [a,b,c]

[a] Biology Department, Graduate Center of the City University of New York, 365 5th Ave., New York, NY 10016, United States
[b] Centre for the Study of Cultural Evolution, Lillafrescativägen 7B, 106 91 Stockholm, Sweden
[c] Psychology Department, Brooklyn College, 2900 Bedford Ave., Brooklyn, NY 11217, United States

ARTICLE INFO

ABSTRACT

The mathematical study of genealogies has yielded important insights in population biology, such as the ability to estimate the time to the most recent common ancestor (MRCA) of a sample of genetic sequences or of a group of individuals. Here we introduce a model of cultural genealogies that is a step toward answering similar questions for cultural traits. In our model individuals can inherit from a variable, potentially large number of ancestors, rather than from a fixed, small number of ancestors (one or two) as is typical of genetic evolution. We first show that, given a sample of individuals, a cultural common ancestor does not necessarily exist. We then introduce a related concept: the most recent unique ancestor (MRUA), i.e., the most recent single individual who is the earliest cultural ancestor of the sample. We show that, under neutral evolution, the time to the MRUA can be staggeringly larger than the time to MRCA in a single ancestor model, except when the average number of learning opportunities per individuals is small. Our results point out that the properties of cultural genealogies may be very different from those of genetic genealogies, with potential implications for reconstructing the histories of cultural traits.

1. Introduction

One of the most intriguing questions in the study of cultural evolution is to what extent cultural data can inform us about the past. Researchers have used cultural datasets to estimate the dates of important historical and evolutionary events, such as the age of human language (Perreault and Mathew, 2012), the spread of ethno-linguistic groups (Gray et al., 2009), and the origin of the capacities for cultural transmission (Lind et al., 2013). In particular, phylogenetic methods from molecular systematics have been applied to cultural datasets to infer the ages of cultural "taxa", such as languages, and the relationships between them (Gray and Atkinson, 2003; Kitchen et al., 2010; Walker and Ribeiro, 2010; Reesink et al., 2009; Rogers et al., 2009).

These methods are a useful complement to archeological and historical evidence of cultural change, but rely on data from assemblages of many cultural traits typified at a population level, and thus investigate *macroevolutionary* cultural change. For example, these studies usually treat languages as characteristic sets of lexical, phonemic, and grammatical features shared by all members of a population. Individual cultural traits – e.g. words, techniques for making hand axes, oral literature, folksongs, childrearing practices, etc. – may have unique histories that are lost when considering only the larger systems they constitute at the macroevolutionary level. While we acknowledge the difficulty inherent in defining a 'single' cultural trait – indeed, many traits can be thought of as combinations or systems of component traits – we assert that whatever the definition, single traits spread due to the *microevolutionary* process of individual to individual transmission. In this paper, we develop a model of the history of a single cultural trait transmitted between individuals based on population size, mode of trait transmission, and current prevalence of the trait.

In order to investigate the history of a cultural trait we require a model of the genealogical process for learned traits. In population biology, the mathematical study of genealogies has flourished since the introduction of Kingman's "coalescent", a retrospective model of how lineages of gene copies merge in common ancestors (Kingman, 1982a,b,c). The coalescent predicts the statistical properties of genealogies and thereby enables the estimation of other aspects of population history such as ancestral population size. A central concept in genetic genealogies is the most recent common ancestor (MRCA), i.e., the most recent individual in the past whose gene copy is ancestral to all those in the present. The coalescent estimates that, for a population of size N, the expected time to MRCA

* Corresponding author at: Biology Department, Graduate Center of the City University of New York, 365 5th Ave., New York, NY 10016, United States.
E-mail addresses: elliot.g.aguilar@gmail.com (E. Aguilar),
drghirlanda@gmail.com (S. Ghirlanda).

is of the order of N generations (Hudson, 1991). Might retrospective models from population genetics shed light on the behavior of cultural trait genealogies, just as molecular systematics has done for cultural taxa? We argue that fundamental differences in the nature of cultural transmission make genetic genealogical models unsuited for culture. While genes are inherited uniparentally (or, in the case of recombination, biparentally), cultural traits can be inherited from multiple ancestors. For example, an oral story may be learned in repeated episodes from a number of sources.

Here we introduce a simple, neutral model of cultural genealogies in order to investigate the relationship between the individual-to-individual nature of cultural transmission and historical inference based on cultural data. Specifically, we model the history of learning events pertaining to a single cultural trait possessed by a group of individuals. Each individual has a variable, potentially large number of cultural ancestors—in contrast to models of genetic genealogies in which each individual has one or two ancestors. Individuals form their traits based on information inherited from their multiple ancestors—thus, a learning event may represent the partial or complete transmission of a trait. We first show that, for a sample of individuals drawn from a population, a cultural MRCA does not always exists. That is, the genealogical lines of the sampled individuals may fall into two or more disjoint sets, without ever intersecting. Alternately, we define a concept related to the MRCA but that identifies an individual that is guaranteed to have existed. We call this individual the *most recent unique ancestor* (MRUA), defined as the most recent single individual who is a source of cultural information for individuals in the sample. For any random sample of a population of fixed size and undergoing neutral cultural evolution, we show that the expected time to the MRUA grows hyper-exponentially with population size, and thus, above certain parameter values, is much larger than the time to MRCA in a uniparental model. This result highlights the need for more theory on historical inference based on cultural data.

2. Model

2.1. Defining the MRUA concept

The distinction between MRUA and MRCA exists because of differences between cultural and genetic inheritance. A single gene copy is inherited uniparentally, and has its origin in a single individual. This means that as we move backwards in time lineages of copies of a gene will eventually converge in the common ancestor. Even in the event of recombination, the gene copy has at most two parents and those lineages too will eventually merge in a common ancestor. By contrast, a person's cultural trait may have inputs from many sources. The original trait could have arisen in a single individual or among a group of individuals. In the latter case it is clear that no single MRCA exists. However, even if a trait has a single origin it may still be modified by subsequent inputs of information. The individuals who contribute this additional information may be "dead ends" in the genealogy, since their own learning lineages may never merge with the others in the sample. Fig. 1 depicts an example of a cultural genealogy as a directed graph. The nodes represent individuals in the genealogy, and arrows show ancestor–descendant relationships (arrows point from ancestor to descendant). At the top of the figure we have individual A, who possesses a cultural trait, let us say knowledge of a specific folksong. At some point in the past, A learned her song from B, C, and D. Prior to that, B and C learned it from E, thus E is their common ancestor. However, Individual D served as a learning model for A – perhaps contributing a new verse – without having learned anything from E. Thus, E is not a *common ancestor* for everyone in the genealogy, but she is still the earliest single individual who contributes to the song learned by A. For this reason, we refer to E as the *most recent unique ancestor*, or MRUA.
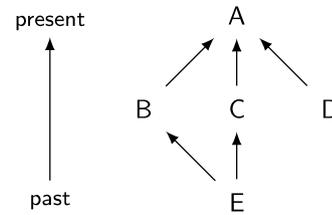

Fig. 1. Sample learning genealogy for a cultural trait (e.g. a folksong). While B, C, D, and E are all ancestral to A, only B and C share a common ancestor in E. D has no learning lineages that intersect with the other individuals in the genealogy.

In genetic evolution, origin and common ancestry coincide. As we have seen, in cultural evolution there may be no common ancestor, since: (1) individuals may independently invent a trait that already exists (e.g. Leibniz's and Newton's independent inventions of the calculus), and go on to serve as learning models; and (2) individuals may independently invent some information that gets incorporated into a trait even without possessing it themselves, as in the example of D above. Therefore, the origin of a trait for some group of individuals is represented by the MRUA, which plays a role analogous to the that of the MRCA in the neutral coalescent.

2.2. Learning model

To investigate the properties of cultural genealogies, we begin with a simple model of cultural transmission. We consider a fixed population of N individuals that evolves according to a continuous-time Markov process in which only two events can occur:

Replacement: A randomly selected individual is removed from the population and is replaced by a naive individual.

Learning: A randomly selected individual learns from another randomly selected individual.

The time between successive replacement events is exponentially distributed with rate r per individual, and the time between successive learning events is exponentially distributed with rate a per individual. Equivalently, we can say that replacement and learning events occur according to a Poisson process with rate $(a+r)N$. A given event is a replacement with probability $r/(a+r)$, and a learning event with probability $a/(a+r)$. The expected lifetime of individuals is $1/r$, and the expected number of learning events per lifetime is a/r.

This model allows individuals to acquire multiple cultural ancestors via multiple learning events in the course of a lifetime. We define a learning event as the transmission of some amount of information about a trait from one individual to another. We do not assume that the event causes the exact replication of a cultural trait, and we leave unspecified exactly how an individual forms her cultural trait from multiple inputs. While cultural transmission can be conceptualized in many ways (Henrich and Boyd, 2002), here we are concerned only with tracking genealogies of a single trait. In other words, regardless of the details of how cultural transmission occurs, we can consider anyone from whom an individual has learned as the individual's cultural ancestor. Note that in a multiple-ancestry model it is important to distinguish learning events from ancestors. Any number of learning events with the same ancestor, in fact, results in a single lineage in the genealogy (*see* Section 2.3).

2.3. Genealogical model

The aim of a theory of cultural genealogies is to infer the statistical properties of genealogies from knowledge of the process that generates them, which, in the present case, is the simple cultural transmission process introduced above. We are not tracking
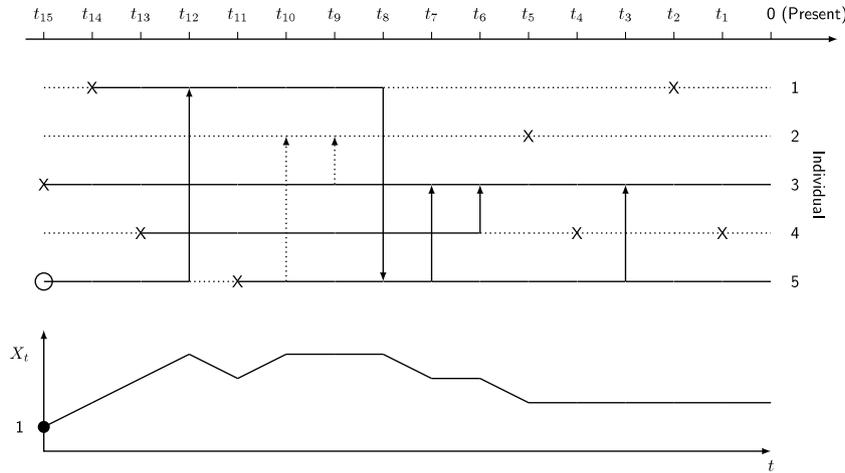
Fig. 2. Example of a cultural genealogy from a simulation of the model in the text, with $a = 1.5$, $r = 1$. Horizontal lines indicate individuals through time, with solid lines highlighting individuals within the sample being tracked. Arrows indicate learning events and ×'s indicate replacement. Times, t_j, represent the occurrence of the jth event (birth or replacement), shown here at equal intervals for convenience. Five individuals are followed through 15 successive events. The initial sample has been arbitrarily defined, for illustration, as individuals 3 and 5 at $t = 0$. Starting from the right side, the graph is considered backward in time. Any individual located at the base of an arrow incident on a sample member is added to the sample. Sample members who are replaced are removed from the sample. Individual i at time t_j is an ancestor of individual k at time t_l if it is possible to reach point (k, t_l) from (i, t_j) following forward arrows. The open circle indicates the most recent unique ancestor of the sample (MRUA, see text). Sample size is tracked in the plot below the genealogy.

the spread of a cultural variant, but modeling the history of learning events that have resulted in a group of individuals possessing a given trait. We imagine encountering a population in the present with no information about the history of learning relationships between its members, and we want to know how far back in time we would expect to find the MRUA of a sample of n individuals who share a trait in common.

To calculate the expected time to MRUA we define the random variable X_t as the number of ancestry lines existing at a time t before the present that lead to individuals in the initial sample. Starting from an initial sample size, $X_0 = n$, we define the time to MRUA, τ_n as the smallest time for which $X_t = 1$:

$$\tau_n \stackrel{\text{def}}{=} \inf_t \{X_t = 1 \mid X_0 = n\}. \tag{1}$$

In our model there are only two events that affect X_t:

1. An individual in the sample learns from an individual outside the sample. This leads to the branching of an ancestry line, increasing sample size by 1. These events occur with rate

   learning rate $\times$ Pr (learner in sample)
   $\times$ Pr (ancestor outside sample)

   $$= aN \times \frac{X_t}{N} \times \frac{N - X_t}{N - 1}. \tag{2}$$

2. An individual in the sample is replaced by a naïve individual. This leads to the interruption of an ancestry line, decreasing sample size by 1. These events occur with rate

   replacement rate $\times$ Pr (replacement occurs in the sample)

   $$= rN \times \frac{X_t}{N}. \tag{3}$$

Other events, such as replacement of an individual outside the sample or learning between individuals who are in the sample, do not affect X_t.

As a consequence of learning and replacement events, sample size changes according to a continuous time birth–death process whose states are in the subset of integers $\{0, N\}$ (though we are only concerned with the process until it reaches state $X_t = 1$), and with state-dependent birth and death rates (i.e. $X_t = k$), λ_k and μ_k,

given by Eqs. (2) and (3):

$$\lambda_k = ak \frac{N - k}{N - 1} \tag{4}$$

$$\mu_k = rk. \tag{5}$$

The time between events is exponentially distributed with parameter $\kappa_k = \lambda_k + \mu_k$. The time evolution of X_t governed by the rates given above is the *genealogical process*.

Fig. 2 shows the results from a simulation of the genealogical process for a population $N = 5$ (*see* Appendix D for simulation algorithm). The lower plot shows the time evolution of X_t over a series of events resulting in MRUA. The graph above, inspired by the percolation diagram introduced by Krone and Neuhauser (1997) to represent the coalescent process with selection, shows the simulated genealogy. Horizontal lines represent individuals, and time moves backwards as we move from right to left. Vertical arrows are learning events with the learner at the head of the arrow (i.e. information 'flows' in the direction of the arrow), while ×-marks indicate replacement of an individual. Beginning at the rightmost end of the figure, we consider an initial sample of individuals 3 and 5 who possess some cultural trait. Solid lines indicate lineages that are part of our sample, or events that affect sample size. Thus, X_t is equal to the number of solid horizontal lines at t. Moving left along the lineages and following the arrows in reverse, we arrive at the MRUA.

The genealogical process generated by our model bears a superficial resemblance to the dual process introduced by Krone and Neuhauser (1997) and a conceptual similarity to the recombination graph (Griffiths, 1991; Griffiths and Marjoram, 1997); all three are equivalent to a coalescing, branching process on a complete graph of order N. Unlike the result given below, which is explicit for finite population size, N, Krone and Neuhauser find an expected time expression in the limit $N \to \infty$, using diffusion limit forms of the state-dependent coalescence and branching rates. Similarly, the recombination graph uses diffusion scalings of the transition rates to solve for expected times in infinite populations. There is an important similarity between the recombination graph and our own process, which is that it allows for gene sequences to have upwards of two parents and thus, for finite populations, could be seen as a special case of our multi-ancestor model. However, as will be mentioned below, our process leads to results that differ from those of the recombination graph.
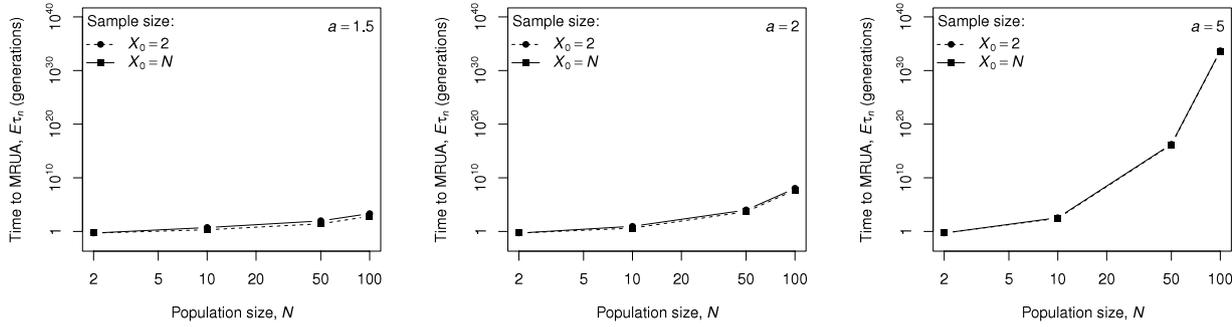
Fig. 3. Expected times to MRUA in samples of 2 and N individuals in populations of size N. Parameters values: $a = 1.5$ (left), $a = 2$ (center), $a = 5$ (right). Time is measured in units of expected lifetime ($r = 1$), see Section 2.2. The vertical scale is the same to facilitate comparison between the graphs.

3. Results

3.1. Model analysis

Let $X_t = k$; the process will remain at k for an expected time $1/\kappa_k$ before jumping to state $X_{t+s} = k + 1$ with probability λ_k/κ_k, or state $X_{t+s} = k - 1$ with probability μ_k/κ_k (s is the actual waiting time, exponentially distributed with mean $1/\kappa_k$). Therefore, the expected value of τ_k is given by the following second-order difference equation (Nåsell, 2011):

$$E[\tau_k] = \frac{1}{\kappa_k} + \frac{\lambda_k}{\kappa_k} E[\tau_{k+1}] + \frac{\mu_k}{\kappa_k} E[\tau_{k-1}] \tag{6}$$

for $2 \le k \le N - 1$, and with boundary conditions

$$E[\tau_1] = 0, \qquad E[\tau_N] = \frac{1}{\kappa_N} + E[\tau_{N-1}]. \tag{7}$$

The first condition reflects the fact that $n = 1$ is the state corresponding to the convergence of all sample lineages – excluding those that have terminated in "dead ends" – in the MRUA; the second condition reflects the fact that the maximum value of X_t is the population size, N. The solution to (6) with these boundary conditions is given by (*see* Appendix A),

$$E[\tau_n] = \frac{1}{r} \sum_{i=2}^{n} \sum_{j=i}^{N} \frac{1}{j} \left[\frac{a}{r(N-1)} \right]^{j-i} \frac{N^{\underline{j}}}{N^{\underline{i}}} \tag{8}$$

where $N^{\underline{j}}$ is the "descending factorial" or "falling power" (Graham et al., 1990), defined by $N^{\underline{j}} \overset{\text{def}}{=} N(N-1)\cdots(N-j+1)$. A similar calculation leads to the variance and higher moments (*see* Appendix B).

3.2. Time to the MRUA

Eq. (8), though exact, is not transparent. In Fig. 3 we show numerical calculations of times to MRUA for parameter values intended to represent cultural traits that are learned in only a few learning events (small a) or from a large number of events (larger a, expected for many human cultural traits). MRUA times grow faster than exponentially with population size, reaching staggering magnitudes even for a relatively modest $a = 5$. We can also see that a small increase from an average 1.5 to 2 learning events leads to a 10^4-fold increase in MRUA times. These properties stand in contrast with the time to the genetic MRCA, which is of the order of population size, N for large samples (Kingman, 1982c).

Fig. 3 also suggests that increasing a decreases the difference in MRUA times between different initial sample sizes. To clarify this result, Fig. 4 plots the ratio E_{τ_N}/E_{τ_2} for a range of population sizes. As the learning parameter a increases, the ratio approaches 1, and
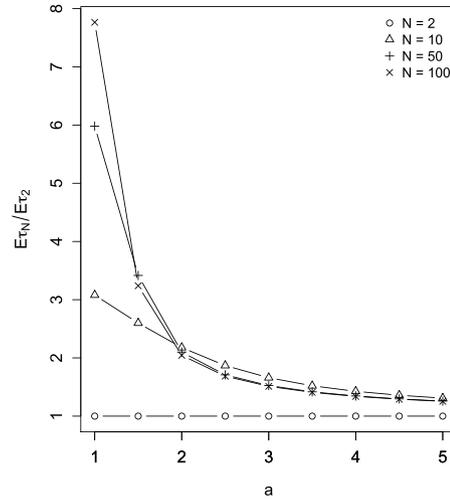

Fig. 4. Ratios of expected MRUA times, E_{τ_N}/E_{τ_2}, for $N = 2, 10, 50, 100$. In all cases, as a increases, the ratio approaches 1.

expected MRUA times do not differ between initial sample sizes. The learning parameter a is a feature of the trait being modeled; some traits will on average be learned in more or fewer learning episodes. Thus, for a trait that is acquired in many learning episodes (i.e. high value of a), we should not expect the MRUA to be any older because the trait is popular than we would if it were very rare.

To shed light on the dependence of MRUA times on population parameters, we approximate the stochastic process with a deterministic process obtained by considering *expected* changes in sample size. When the stochastic process is in state k the expected change in sample size is

$$E[\Delta k] = \frac{\lambda_k - \mu_k}{\lambda_k + \mu_k} = \frac{a(N-k) - r(N-1)}{a(N-k) + r(N-1)}. \tag{9}$$

The deterministic approximation is obtained by using this expected change to recursively update sample size. Fig. 5 shows that Eq. (9) accurately describes the mean evolution of the stochastic process, obtained from simulation. The deterministic approximation reaches an equilibrium value, k^*, which corresponds to the *quasi-stationary* equilibrium of the stochastic process and is independent of initial sample size, n. Setting $E(\Delta k) = 0$ and solving for k we get the equilibrium sample size

$$k^\star = N - \frac{r}{a}(N-1). \tag{10}$$

Thus the equilibrium value of the deterministic process decreases linearly with r/a. After a transient phase determined by the initial sample size, the stochastic process hovers around its *quasi-stationary* equilibrium until a fluctuation brings it down to the state $n = 1$. Higher values of $k^\star$ correspond to longer times to the MRUA
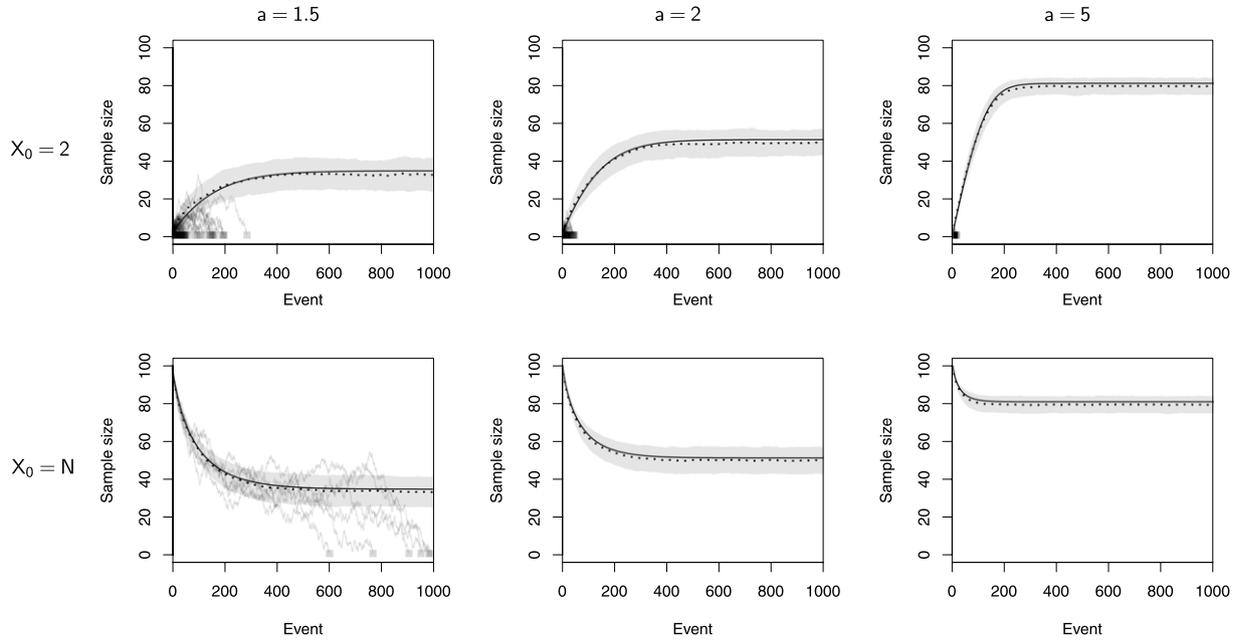
Fig. 5. Change of sample size with time in model simulations (mean: dotted lines; standard deviation: gray area) and in the deterministic approximation (solid lines, Eq. (9)) for three values of learning rate parameter $a \in \{1.5, 2, 5\}$ and two values of initial sample size, $X_0 = 2$ (above), $X_0 = 100$ (below), for a population of $N = 100$ individuals. Replacement rate is $r = 1$. All simulations that reach MRUA in the plotted time window are shown in light gray. One-thousand simulations were performed for each choice of parameters.

because larger fluctuations are necessary to reach $k = 1$. Numerical analysis indicates that time to the MRUA is an exponential function of $k^\star$, namely (see Fig. 6):

$$E[\tau_n] \simeq 10^{0.7\, k^\star \log a}. \tag{11}$$

The above equation does not take into account the variation of $E\tau_n$ with initial sample size, which is significant only for small learning rates, as mentioned in the previous paragraph (*see* Fig. 3). The deterministic approximation shows more clearly the role of a and r in determining time to the MRUA. Increasing a, the expected number of learning events, and thus the potential for multiple ancestry, increases the quasi-stationary equilibrium sample size, prolonging time to MRUA.

3.3. Model comparison

To further explore the implications of multiple ancestry we compared our model with a uniparental model of cultural transmission. The uniparental model we select is that introduced by Strimling et al. (2009) to investigate the accumulation of cultural traits in a fixed population. While the model was originally described in discrete time, we consider a continuous time version. Ignoring mutation and imperfect copying, the model proceeds in the following steps:

1. A randomly selected individual leaves the population.
2. A new individual replaces her and chooses another member of the population at random to be her sole cultural parent.

The times between events are exponentially distributed with parameter r, with $1/r$ being the average individual lifetime; for simplicity we set $r = 1$ for both models.

The model of Strimling et al. is similar to a Moran model of reproduction, but with an important exception: the individual that leaves the population cannot be chosen as a cultural parent by her replacement. All lineages will eventually converge in an MRCA because, as in genetic evolution, information is inherited uniparentally. Therefore, MRCA and MRUA are equivalent in the

Strimling et al. model, and we will only refer to MRUA in the following discussion. In continuous time, with $r = 1$, we arrive at the following expression for the time to MRUA for a sample of n individuals (see Appendix C)

$$E\tau_n = (N - 1) \sum_{j=2}^{n} \frac{1}{j(j-1)}. \tag{12}$$

This result is nearly identical to that of the standard neutral coalescent. Fig. 7 shows a comparison of MRUA times in the multiple ancestry and Strimling et al. models with increasing N and an initial sample, $X_0 = 2$. In our model, the time to MRUA grows hyper-exponentially with population size, typically, while it grows linearly in the Strimling et al. model. However, when learning opportunities are few (a smaller than a critical value that depends on population size), the expected times are actually longer for the Strimling et al. model than for our own. For example, for $a = 1$, meaning individuals have on average one learning event per lifetime, the multiple ancestry model times are shorter by an order of magnitude for populations larger than 10, for all values of initial sample size up to N. Given the general effect of multiple ancestry on extending MRUA times, it is interesting to see that in some cases the multiple ancestry model can actually lead to MRUA times that are shorter than in the single-ancestor model. This stands in contrast to the recombination graph mentioned above, whose minimum expected times are equivalent to the standard neutral coalescent.

4. Discussion

We constructed what is to our knowledge the first model of cultural transmission to explore the behavior of cultural genealogies. The major advantage of our model is its generality; we tracked the learning lineages without making any assumptions about the nature of transmission (e.g., discrete vs. continuous units) or about how cultural phenotypes emerge from learned information. The model has two key hypotheses: neutrality and multiple inheritance.

Neutrality, i.e., random choice of individuals for replacement and learning independent of their biological and cultural traits, is
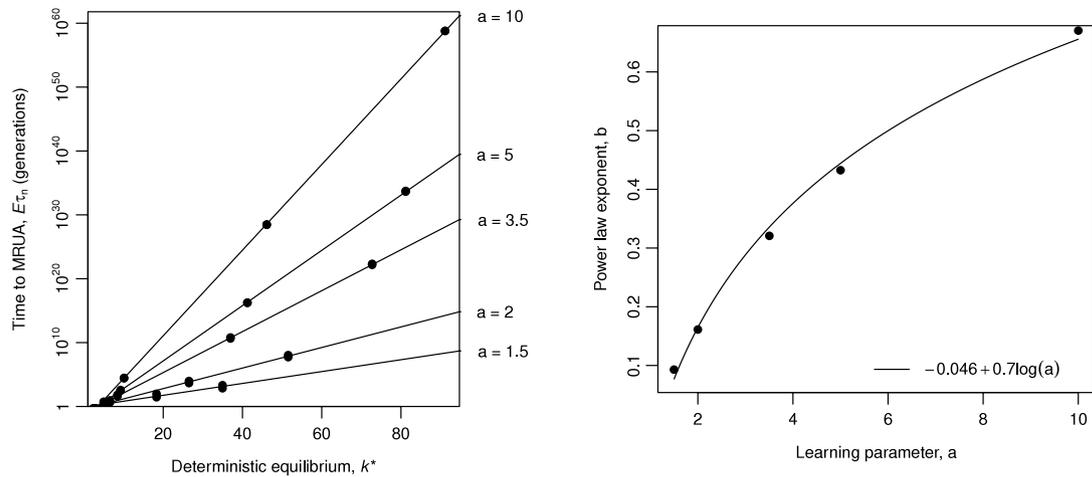
Fig. 6. Numerical determination of the approximate relationship in (11). The left panel shows that $E\tau_N$ is approximately exponential in $k^\star$, as defined in (10), $E\tau_n \simeq 10^{bk^\star}$. The right panel shows that b in the last expression is a logarithmic function of the learning rate parameter a. Numerical fits have been obtained with the `lm` function of the R software, version 3.0.0, applied to log-transformed values as appropriate (R Core Team, 2013).

not realistic in general. Cultural time series often appear to deviate from neutrality (Gureckis and Goldstone, 2009; Berger and Le Mens, 2009; Acerbi et al., 2012), though not invariably (Bentley, 2008). Many mechanisms could cause such deviations: some cultural variants may be acquired more easily (or discarded less easily), some individuals may be preferred as cultural models, cultural traits may influence survival and reproduction, and so on (Boyd and Richerson, 1985; Henrich, 2001; Strimling et al., 2009; Acerbi et al., 2012). Future research will have to assess how the cultural genealogical process is affected by these factors, and by others such as variable population size (e.g., a population expansion will appear, looking back in time, as a reduction in the number of possible ancestors, which should shorten times to the MRUA). However, by assuming neutrality we have been able to isolate the effect of the other main hypothesis: multiple ancestry.

Unlike neutrality, multiple ancestry is a fundamental aspect of cultural transmission (Boyd and Richerson, 1985; Cavalli Sforza and Feldman, 1981; Enquist et al., 2010) and thus a necessary ingredient of any model of cultural genealogies. We have shown that multiple ancestry leads to genealogies that behave very differently from those produced by haploid and diploid inheritance. First, the concept of a most recent common ancestor, familiar from population genetic theory, appears less useful in the study of cultural genealogies, because these may lack a common ancestor altogether. We proposed to overcome this difficulty by studying instead the "unique" ancestor of a sample of individuals—the most recent single individual who contributed cultural information to at least one of the lineages leading to the sample. This individual is guaranteed to have existed, and coincides with the most recent common ancestor when the latter exists. Our model demonstrated that multiple inheritance generally causes times to MRUA to be much longer than in a uniparental model.

MRUA and MRCA are analogous because they both identify an ancestral individual at the root of a genealogy. However, other definitions of "common ancestor" are possible. For example, a common ancestor can be defined as any individual who appears in the genealogies of all those sampled. This is the lay definition of common ancestry, according to which a grandmother is a common ancestor of all her grandchildren. Chang (1999) showed that the time to such an ancestor for a random sample of a large diploid population is of the order of $\log_2 N$ generations for large N, thus much shorter than the time to MRCA in uniparental models, which is of the order of N generations, and of course shorter than time to MRUA in our model. We leave it to future work to determine the expected time to "cultural grandmothers", though we suspect that
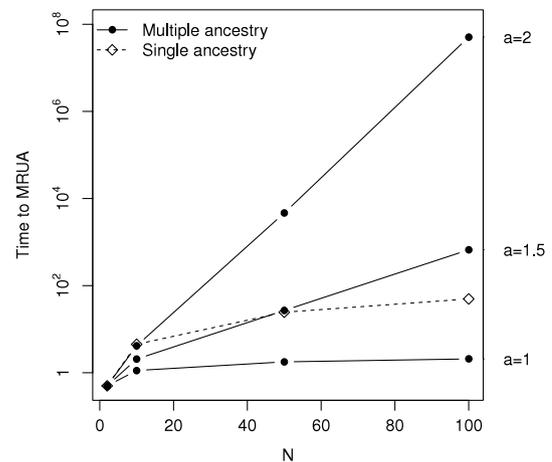

Fig. 7. Times to MRUA in the multiple ancestry model (solid lines) and single ancestor model (dashed line) for an initial sample size of 2.

the times will be even shorter than the diploid result, due to the branching nature of our process.

Our model may be applicable to questions of genealogical processes in cultural transmission, such as oral story transmission. The model incorporates the possibility of learning from multiple individuals, learning multiple times from the same individual, and independent contribution, which are all factors in the transmission of oral stories (Ong, 1982). Estimating the parameters of the model, "lifetime" of the active storyteller/learner, and rate of learning events, might lead to estimates of the time of MRUA for a story, which could be compared with independent historical evidence to validate the model. Experimental setups could also be used to investigate the time to the MRUA of stories as a function of parameter values, and to compare these times to the case of uniparental transmission. As there is a growing body of experimental work on oral transmission and cultural evolution (Eriksson and Coultas, 2012; Barrett and Nyhof, 2001), this is an especially promising direction for empirical application of our model.

In conclusion, our work represents a first step toward a theoretical foundation for historical inference based on cultural data. We investigated the history of an individual cultural trait by explicitly modeling the individual-to-individual transmission process with multiple ancestry. Many other properties of cultural genealogies remain to be explored, such as the effects of biased transmission, and the effect of historical fluctuations in the popularity of a trait.

Further work is needed to ascertain the impact of the unique features of cultural inheritance on the theory and practice of reconstructing the histories of cultural traits.

Appendix A. Expected time to MRUA

To solve Eq. (6), we first multiply through by κ_k and rearrange the terms to obtain,

$$\mu_k(E[\tau_k] - E[\tau_{k-1}]) - \lambda_k(E[\tau_{k+1}] - E[\tau_k]) = 1. \tag{A.1}$$

By Eq. (3.56) in Nåsell (2011) we have,

$$E[\tau_k] - E[\tau_{k-1}] = \frac{1}{\mu_1 \rho_k} \sum_{j=k}^{N} \pi_j, \tag{A.2}$$

where we have made use of the second boundary condition in (7) and,

$$\rho_k = \frac{\lambda_1 \lambda_2 \dots \lambda_{k-1}}{\mu_1 \mu_2 \dots \mu_{k-1}} \tag{A.3}$$

$$\pi_j = \frac{\lambda_1 \lambda_2 \dots \lambda_{j-1}}{\mu_2 \mu_3 \dots \mu_j}. \tag{A.4}$$

Imposing the boundary condition $E[\tau_1] = 0$ leads to,

$$E[\tau_k] = \frac{1}{\mu_1} \sum_{i=2}^{k} \sum_{j=i}^{N} \frac{\pi_j}{\rho_i} \tag{A.5}$$

where, for convenience, we have moved the factor $1/\rho_j$ into the innermost sum. Lastly, writing ρ_i and π_j in terms of birth and death rates (Eq. (4)) gives the solution in Eq. (8).

It is also easy to derive an expected time until sample size reaches a size m smaller than our initial sample size but larger than one. If we define $\tau_{n,m} = \inf_t\{X_t = m; X_0 = n, n > m\}$ we have

$$E[\tau_{n,m}] = \frac{1}{r} \sum_{i=m+1}^{n} \sum_{j=i}^{N} \frac{1}{j} \left[\frac{a}{r(N-1)} \right]^{j-i} \frac{N^{\underline{j}}}{N^{\underline{i}}}, \tag{A.6}$$

which we arrive at with the same argument as above, but with the revised boundary conditions

$$E[\tau_{m,m}] = 0, \qquad E[\tau_{N,m}] = \frac{1}{\kappa_N} + E[\tau_{N-1}]. \tag{A.7}$$

Appendix B. Variance in time to MRUA

We begin with the more general form of (6) (Norden, 1982),

$$E_i^r = \frac{r}{\lambda_i + \mu_i} E_i^r + \frac{\lambda_i}{\lambda_i + \mu_i} E_{i+1}^r + \frac{\mu_i}{\lambda_i + \mu_i} E_{i-1}^r \tag{B.1}$$

where $E_i^r = E(\tau_i^r)$, the *rth* moment of the random variable τ_i. For $r = 2$ we have boundary conditions,

$$E_1^2 = 0, \qquad E_1^1 = 0, \qquad E_N^1 = \frac{1}{\lambda_N + \mu_N} + E_{N-1}^1. \tag{B.2}$$

Using (B.2), we can derive the following equation for the second moment (see Eq. (6.7), Norden, 1982),

$$E_i^2 = 2 \sum_{j=2}^{N} E_j^1 \alpha_j \sum_{k=2}^{min(i,j)} \psi_k \tag{B.3}$$

where $\alpha_j = (\mu_j \psi_j)^{-1}$ and $\psi_k = \frac{\lambda_k \lambda_{k+1} \dots \lambda_N}{\mu_k \mu_{k+1} \dots \mu_N}$. Reversing the order of the sums and moving all terms to the innermost sum yields,

$$E_i^2 = 2 \sum_{k=2}^{i} \sum_{j=2}^{N} E_j^1 \alpha_j \psi_k \tag{B.4}$$

Substituting the definitions for α_l and ψ_k, we have the state dependent second moments,

$$E_i^2 = \frac{2}{r} \sum_{k=2}^{i} \sum_{j=2}^{N} E_j^1 \frac{1}{j} \left[\frac{a}{r(N-1)} \right]^{j-k} \frac{N^{\underline{j}}}{N^{\underline{k}}}. \tag{B.5}$$

Appendix C. Expected time to MRUA in Strimling et al. model

Common ancestry events occur when a replacement event has occurred within the sample and the new individual has chosen another sample member as her cultural parent. Replacement events occur with rate rN, where r is the reciprocal of average lifetime and N is the population size. Given that a replacement event has occurred, the probability that replacement and learning both take place within the sample is $j(j-1)/[N(N-1)]$. Therefore, a common ancestry event will occur in a sample of size j with rate,

$$\frac{rj(j-1)}{N-1} \tag{C.1}$$

and the expected time to this event will be $(N-1)/[rj(j-1)]$. Note that τ_n, the time to MRUA for an initial sample $X_0 = n$, is

$$\tau_n = t_1 + t_2 + \dots + t_{n-1}, \tag{C.2}$$

where t_i is the time between the ith and $(i-1)$th common ancestry event. Thus, $E[\tau_n] = E[t_1] + E[t_2] + \dots + E[\tau_{n-1}]$, and,

$$E[\tau_n] = \frac{N-1}{rn(n-1)} + \frac{N-1}{r(n-1)((n-1)-1)} + \dots + \frac{N-1}{2r}$$

$$= \sum_{j=2}^{n} \frac{(N-1)}{rj(j-1)}. \tag{C.3}$$

For $r = 1$ we have the result given in (12).

Appendix D. Algorithm for simulating the genealogical process

Below is the algorithm we used for simulating the genealogical process:

1. Let N = population size, r = replacement rate, a = learning rate.
2. Let k = initial sample size.
3. Let $c = 0$ (event counter).
4. Increment c by 1.
5. Let $\lambda = a * k * (N-k)/(N-1)$.
6. Let $\mu = r * k$.
7. Let $x = \lambda/(\lambda + \mu)$.
8. Let y = uniform random deviate in [0, 1].
9. If $y \leq x$, increment k by 1, else decrement k by 1.
10. If k is 1, print out c and stop, otherwise go to step 4.

This algorithm simulates events. If one wants to simulate time one can increment c by an exponential deviate with expected value of $(r + a) * N$, then c will measure time.

References

Acerbi, A., Ghirlanda, S., Enquist, M., 2012. The logic of fashion cycles. PLoS One 7 (3), e32541. http://dx.doi.org/10.1371/journal.pone.0032541.

Barrett, J.L., Nyhof, M.A., 2001. Spreading non-natural concepts: the role of intuitive conceptual structures in memory and transmission of cultural materials.. J. Cogn. Cult. 1, 69–100.

Bentley, R.A., 2008. Random drift versus selection in academic vocabulary: an evolutionary analysis of published keywords. PLOS One 3 (8), e3057.

Berger, J., Le Mens, G., 2009. How adoption spread affects the abandonment of cultural tastes. PNAS 106 (20), 8146–8150.

Boyd, R., Richerson, P.J., 1985. Culture and the Evolutionary Process. Chicago University Press, Chicago.

Cavalli Sforza, L., Feldman, M., 1981. Cultural Transmission and Evolution. Princeton University Press, Princeton, NJ.

Chang, J.T., 1999. Recent common ancestors of all present-day individuals. Adv. in Appl. Probab. 31, 1027–1038.

Enquist, M., Strimling, P., Eriksson, K., Laland, K.N., Sjostrand, J., 2010. One cultural parent makes no culture. Animal Behav. 79, 1353–1362.

Eriksson, K., Coultas, J., 2012. The advantage of multiple cultural parents in the cultural transmission of stories. Evol. Hum. Behav. 33, 251–259.

Graham, R.L., Knuth, D.E., Patashnik, O., 1990. Concrete Mathematics, fourth printing. Addison-Wesley Publishing Company, p. 1989.

Gray, R.D., Atkinson, Q.D., 2003. Language-tree divergence times support the anatolian theory of Indo–European origin. Nature 426, 435–439.

Gray, R.D., Drummond, A.J., Greenhill, S.J., 2009. Language phylogenies reveal expansion pulses and pauses in Pacific settlement. Science 323, 479–483.

Griffiths, R.C., 1991. The two-locus ancestral graph. In: Selected Proceedings of the Sheffield Symposium on Applied Probability. Institute of Mathematical Statistics, Hayward, CA, pp. 100–117.

Griffiths, R.C., Marjoram, P., 1997. An ancestral recombination graph. In: Donnelly, P., Tavaré, S. (Eds.), Progress in Population Genetics and Human Evolution. Springer Verlag, pp. 257–270.

Gureckis, T.M., Goldstone, R.L., 2009. How you named your child: understanding the relationship between individual decision making and collective outcomes. Top. Cogn. Sci. 1, 651–674.

Henrich, J., 2001. Cultural transmission and the diffusion of innovations: adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change. Am. Anthropol. 103 (4), 992–1013.

Henrich, J., Boyd, R., 2002. On modeling cognition and culture: why replicators are not necessary for cultural evolution. J. Cogn. Cult. 2 (2), 87–112.

Hudson, R.R., 1991. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. 7, 144.

Kingman, J.F.C., 1982a. On the genealogy of large populations. J. Appl. Probab. 19A, 27–43.

Kingman, J.F.C., 1982b. The coalescent. Stochastic Process. Appl. 13, 235–248.

Kingman, J.F.C., 1982c. Exchangeability and the evolution of large popualtions, in: G. Koch, F. Spizzichino (Eds.), Exchangeability in Probability and Statistics. Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6th–9th April, 1981, in honour of Professor Bruno De Finetti.

Kitchen, A., Ehret, C., Assefa, S., Mulligan, C.J., 2010. Bayesian phylogenetic analysis of semitic languages identifies an early bronze age origin of semitic in the near east. Proc. Biol. Sci. 276, 2703–2710.

Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theor. Popul. Biol. 51 (3), 210–237.

Lind, J., Lindenfors, P., Ghirlanda, S., Liden, K., Enquist, M., 2013. Dating human cultural capacity using phylogenetic principles. Sci. Rep. 3.

Nåsell,, 2011. Extinction and Quasi-stationarity in the Stochastic Logistic SIS Model. In: Springer Lecture Notes in Mathematics, Springer-Verlag, Berlin and Heidelberg.

Norden, R.H., 1982. On the distribution of the time to extinction in the stochastic logistic population model. Adv. Appl. Probab. 14 (4), 687–708.

Ong, W.J., 1982. Orality and Literacy. Routledge, New York, NY.

Perreault, C., Mathew, S., 2012. Dating the origin of language using phonemic diversity. PLoS ONE 7 (4).

R Core Team 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, URL http://www.R-project.org.

Reesink, G., Sinfer, R., Dunn, M., 2009. Explaining the linguistic diversity of Sahul using population models. PLoS Biol. 7 (11).

Rogers, D.S., Feldman, M.W., Ehrlich, P.R., 2009. Inferring population histories using cultural data. Proc. Roy. Soc. Lond. Ser. B.

Strimling, P., Eriksson, K., Enquist, M., 2009. Repeated learning makes cultural evolution unique. Proc. Natl. Acad. Sci. USA 13870–13874.

Walker, R.S., Ribeiro, L.A., 2010. Bayesian phylogeography of the Arawak expansion in lowland South America. Proc. Biol. Sci. 278, 2562–2567.

Further reading

Boyd, R., Richerson, P.J., Borgerhoff Mulder, M., Durham, W.H., 1997. Are cultural phylogenies possible? In: Weingart, P., Richerson, P.J., Mitchell, S.D., Maasen, S. (Eds.), Human by Nature. Between Biology and the Social Sciences. Lawrence Erlbaum Associates, Mahwah, NJ.

Mace, R., Jordan, F.M., 2011. Macro-evolutionary studies of cultural diversity: a review of empirical studies of cultural transmission and cultural adaptation. Philos. Trans. Roy. Soc. B (366), 402–411.