Published in final edited form as:

J Theor Biol. 2009 November 21; 261(2): 238–247. doi:10.1016/j.jtbi.2009.08.006.

Inbreeding, Pedigree Size, and the Most Recent Common Ancestor of Humanity

Joseph Lachance

Graduate Program in Genetics, Department of Ecology and Evolution, Stony Brook University,
Stony Brook, NY 11794-5222

Abstract

How many generations ago did the common ancestor of all present-day individuals live, and how does inbreeding affect this estimate? The number of ancestors within family trees determines the timing of the most recent common ancestor of humanity. However, mating is often non-random and inbreeding is ubiquitous in natural populations. Rates of pedigree growth are found for multiple types of inbreeding. This data is then combined with models of global population structure to estimate biparental coalescence times. When pedigrees for regular systems of mating are constructed, the growth rates of inbred populations contain Fibonacci n -step constants. The timing of the most recent common ancestor depends on global population structure, the mean rate of pedigree growth, mean fitness, and current population size. Inbreeding reduces the number of ancestors in a pedigree, pushing back global common ancestry times. These results are consistent with the remarkable findings of previous studies: all humanity shares common ancestry in the recent past.

Keywords

biparental coalescence; genealogy; inbreeding; MRCA; theory

1. Introduction

All modern humans are ultimately related and the most recent common ancestor (MRCA) of humanity lived in the recent past. The exact timing of the MRCA depends on whether gene lineages or family trees are considered. Substantial numbers of individuals can trace their heredity to the likes of Niall of the Nine Hostages and Genghis Khan (Moore *et al.*, 2006; Zerjal *et al.*, 2003). On a grander scale, global coalescence times have been found for mtDNA and non-recombining Y chromosomal DNA lineages (Ayala, 1995; Cann *et al.*, 1987; Thomson *et al.*, 2000). However, the so-called “mitochondrial Eve” and “Y-chromosome Adam” need not be the most recent common ancestors of humanity. Relatedness does not require individuals to share an unbroken matrilineal or patrilineal lineage. Instead, two individuals are related if they share at least one direct ancestor (i.e. the same individual appears in the pedigrees of both individuals). The organismal MRCA is defined here as the most recent individual that is in the family tree of every present-day individual. This biparental definition captures the colloquial

Joseph.Lachance@sunysb.edu, Phone: 631-632-8588, Fax: 631-632-9797.

Mailing address: Joseph Lachance, Graduate Program in Genetics, Department of Ecology and Evolution, Stony Brook University,
Stony Brook, NY 11794-5222

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

meaning of common ancestry. As family trees of present-day individuals are traced backwards in time, the likelihood of common ancestry increases. The number of ancestors in the pedigree of a single present-day individual can be used to calculate biparental coalescence times for an entire population (Chang, 1999). These methods yield estimates of global coalescence times as low as 33 generations for panmictic populations (Chang, 1999) and 76 generations for subdivided populations (Rohde *et al.*, 2004).

Organismal lineages coalesce much faster than gene genealogies. Population size t generations ago is defined as N_t , with present-day population size equal to N_0 . Gene trees coalesce on the order of $2N_0$ generations, while organismal lineages in randomly mating populations coalesce on the order of $\log_2 N_0$ generations (Chang, 1999). This difference in time scales is because genes are inherited uniparentally and organismal ancestry is biparental (every individual has a mother and a father). However, the genetic contribution of the organismal MRCA to present-day individuals can be quite small and two individuals need not inherit the same genes from a shared ancestor (Hein, 2004; Matsen and Evans, 2008). Note that biparental coalescence times are much less variable than uniparental coalescence times (Chang, 1999). Additionally, the existence of a MRCA does not imply that only a single pair of individuals were alive: other individuals existed at the time of the MRCA (Ayala, 1995).

Population structure, such as inbreeding, influences the timing of the MRCA of humanity. Mating is rarely panmictic and inbreeding is ubiquitous in natural populations (Hedrick and Kalinowski, 2000; Keller and Waller, 2002). Regional estimates of consanguinity (second cousin or closer mating) range from less than 1% to greater than 50% (Bittles, 2001). Inbreeding is defined here as positive assortative mating with respect to heredity, and does not refer to incidental matings between close relatives in small populations. Inbreeding causes the number of direct ancestors in a pedigree to differ from 2^t (where t is the number of generations in the past). For example, the progeny of first cousin matings have six, rather than eight, great-grandparents. As inbreeding affects the number of ancestors in a pedigree, it results in modified organismal coalescence times. Previous studies of organismal coalescence times assumed random mating within demes and did not explicitly consider the effects of inbreeding.

In this paper, rates of pedigree growth are found for multiple types of inbreeding. This information is then combined with observed levels of inbreeding to estimate the timing of the most recent common ancestor of humanity. By incorporating inbreeding, estimates of T_{MRCA} become more realistic. Inbreeding results in increased biparental coalescence times.

2. Model

A population of diploid individuals with discrete generations is modeled. Population sizes are finite, but large enough that sex ratios do not differ appreciably from 1:1. Inbreeding is the sole exception to random mating within demes. Thus, the Wright-Fisher model of population genetics is modified to include inbreeding and biparental inheritance. Within demes, individuals preferentially mate with relatives. A proportion of matings involve siblings, a proportion of matings involve first cousins, a proportion of matings involve second cousins, etc. This type of inbreeding is independent of population size. The global population is composed of a number of demes connected by migration. Under this formulation global population structure can be represented via an evolutionary graph (as per Rohde *et al.*, 2004).

Levels of inbreeding are assumed to be uncorrelated within families (i.e. an individual's probability of inbreeding is independent of the type of mating of his or her parents). If inbred pairings are clustered within certain families, those families will have smaller pedigrees. This phenomenon is a ramification of individuals being double-counted within pedigrees. To fully describe such a process would require a large transition matrix (where each row and column

represents a particular level of inbreeding). Consequently, mathematical tractability is retained by assuming that mating patterns are uncorrelated in families.

Organismal coalescence times hinge upon the number of ancestors (A_t) relative to population size (N_t) t generations ago (Jobling *et al.*, 2003). A_t is defined as the number of ancestors a single present-day individual has at time t . Note that A_t and N_t refer to the number of individuals within a single deme, rather than the global number of individuals. Beginning with a single present-day individual, lineages are traced backward in time to obtain rates of pedigree growth for inbred populations. This biparental approach takes advantage of the fact that every individual must have two parents, while every individual need not have two offspring. Once rates of pedigree growth are found, forward time approaches are used to calculate within-deme and global estimates of T_{MRCA} .

3. Methods

Pedigree construction and recursion equations for the number of ancestors

Previous studies compute global coalescence times from the number of ancestors in a pedigree (Chang, 1999; Rohde *et al.*, 2004). This approach can be extended to situations where the number of direct ancestors of each individual differs from two, such as species that reproduce both sexually and asexually (Donnelly *et al.*, 1999; Hein *et al.*, 2005). Similarly, inbreeding causes the rate of pedigree growth to differ from two. How fast do pedigrees grow for different types of inbreeding? This question is answered by creating pedigrees where every mating pair shares the same level of relatedness. These pedigrees grow deterministically at a rate determined by the type of inbreeding. For example, pedigrees where every mating involves first cousins grow more slowly than pedigrees where each mating involves second cousins.

Each generation ancestors are added using a two-step algorithm. First, parents are generated for every individual at the current pedigree depth. Secondly, each instance of inbreeding involves the removal of a pair of putative ancestors and the closing of hereditary loops. The type of inbreeding determines the size and shape of hereditary loops. For example, first cousin mating results in shared grandparents and hereditary loops span that two generations. This algorithm is iterated backwards in time, resulting in multiple generations of ancestors. The greater the level of inbreeding, the more slowly pedigrees expand in size. Inbred pedigrees are characterized by repeating motifs. Let g denote the number of generations separating an inbreeding pair from their most recent common ancestor ($g = 2$ for first cousins). Pedigrees for regular systems of mating are constructed, where every individual within a pedigree shares the same, uniform level of inbreeding (Figure 1). This enables the rate of pedigree growth for a particular type of inbreeding to be found.

The recursive nature of uniform pedigrees allows the number of individuals at a particular depth of a pedigree to be computed from other levels of the family tree. These recursion equations recapitulate the pedigree construction algorithm. The number of ancestors t generations ago is equal to two times the number of ancestors $t-1$ generations ago, minus the number of ancestors $t-g$ generations ago.

$$A_t = 2A_{t-1} - A_{t-g} \quad (1)$$

Iteration of Equation 1 gives the exact number of ancestors t generations ago for different types of inbreeding (Table 1). Looking backwards in time, the ratio of the number of ancestors in consecutive generations quickly converges to the mating-specific rate of pedigree growth (r_g). For regular systems of mating with $t \gg 0$:

$$r_g = \frac{A_t}{A_{t-1}} \quad (2)$$

Algebraic manipulation of Equations 1 and 2 (see Appendix A) indicates that mating-specific rates of pedigree growth are equal to the largest root of:

$$(r_g)^{g-1}(r_g - 2) + 1 = 0 \quad (3)$$

Values of r_g in Table 1 are calculated numerically from Equation 3.

Pedigrees grow more slowly if matings span generations. For example, matings might involve older males and younger females. In this scenario the number of generations within a pedigree differs from the number of “generations” of absolute time that have transpired. Uniform pedigrees with trans-generation mating can be constructed by modifying step one of the pedigree construction algorithm. Here, each individual has one parent that is a single “generation” older and second parent that is two “generations” older. Trans-generation mating causes pedigrees to become stretched (see diagonal lines in Figure 1). The number of ancestors t generations ago for a trans-generation outbred pedigree follows below. Note that this recursion equation generates terms of the Fibonacci sequence.

$$A_t = A_{t-1} + A_{t-2} \quad (4)$$

Time-reversibility of pedigree growth rates for regular systems of mating

Although pedigrees for regular levels of inbreeding are generated via a time-backward algorithm, subsequent proofs and computer simulations require time-forward growth rates. With a few restrictions, it is possible to construct regular inbred pedigrees in a time-forward manner. First, each individual is required to have two offspring (one female, one male). Second, a single inbreeding event occurs g generations in the future for each individual. For example, two of an individual’s eight great-grandchildren mate in second cousin pedigrees ($g=3$, see Figure 2). Over multiple generations, this results in fewer descendants relative to outbred expectations. Third, pedigrees are constructed so that lines do not cross. An example of a pedigree created using these rules is shown in Figure 2. Let τ denote the number of generations in the future and D_τ denote the number of descendants at time τ . Quantifying the above rules, the number of descendants τ generations in the future is equal to two times the number of descendants at time $\tau-1$ minus the number of descendants at time $\tau-g$.

$$D_\tau = 2D_{\tau-1} - D_{\tau-g} \quad (5)$$

Note that this recursion equation has the same form as Equation 1. This indicates that the number of relatives in regular inbred pedigrees grows at the same rate whether one looks forward or backward in time.

Mean rate of pedigree growth

Real populations contain a mixture of different types of inbreeding. The mean rate of pedigree growth for a population ($\bar{r}$) will be used to calculate biparental coalescence times. In large populations $\bar{r}$ depends on the proportion of each type of mating and mating-specific rates of

pedigree growth. This is because pedigree sizes are large and we have assumed that levels of inbreeding are uncorrelated within families. Let p_g denote the expected proportion of matings that involve a particular type of inbreeding. Realized proportions of each type of mating follow a multinomial distribution. Looking backwards in time, the number of ancestors in a pedigree grows geometrically. In only a few generations the number of direct ancestors becomes quite large. The Law of Large Numbers indicates that if A_t is large, the realized proportion of matings that are a particular type converge to expectations. Consequently, the mean rate of pedigree growth is equal to the weighted mean of mating-specific rates of pedigree growth.

$$\bar{r} = \sum_{g=0}^{\infty} p_g r_g \quad (6)$$

The mean rate of pedigree growth can be calculated by plugging p_g and r_g values from real world populations into Equation 6. One particularly comprehensive data set involves marriage records from rural Sweden (Bittles and Egerbladh, 2005). Data from this study covers the years 1720–1899 and includes types of inbreeding out to 6th cousin mating (Table 2). The proportion of outbred trans-generation matings was estimated by comparing the relative proportions of same-generation inbreeding to trans-generation inbreeding.

T_{MRCA} depends on the ratio of A_t to N_t

Biparental coalescence times depend on the relative sizes of ancestral pedigrees and past populations. Consider a single individual who lived t generations ago and a single present-day individual. The probability that they are related is:

$$P(\text{past and present individuals are related}) = \frac{A_t}{N_t} \quad (7)$$

For a past individual to be a common ancestor, they must be related to all N_0 individuals living in the present.

$$P(\text{specific past individual is not a common ancestor}) = 1 - \left(\frac{A_t}{N_t} \right)^{N_0} \quad (8)$$

Each of the N_t individuals living in the past is a potential common ancestor, and the probability that each is a common ancestor is assumed to be independent. The probability that a common ancestor existed t generations ago is the complement of the probability that a common ancestor did not exist t generations ago.

$$P(\text{a common ancestor exists } t \text{ generations ago}) = 1 - \left(1 - \left(\frac{A_t}{N_t} \right)^{N_0} \right)^{N_t} \quad (9)$$

However, the existence of a common ancestor t generations ago does not imply that a common ancestor first appeared t generations ago. The cumulative probability that no biparental coalescence has occurred in $t - 1$ generations is:

$$P(\text{no common ancestor in } t-1 \text{ generations}) = \prod_{i=1}^{t-1} \left(1 - \left(\frac{A_t}{N_t} \right)^{N_0} \right)^{N_i} \quad (10)$$

If there has been no biparental coalescence in $t-1$ generations but a common ancestor existed t generations ago, then the MRCA existed t generations ago.

$$P(\text{MRCA existed } t \text{ generations ago}) = \left[1 - \left(1 - \left(\frac{A_t}{N_t} \right)^{N_0} \right)^{N_t} \right] \times \left[\prod_{i=1}^{t-1} \left(1 - \left(\frac{A_t}{N_t} \right)^{N_0} \right)^{N_i} \right] \quad (11)$$

Exponents in Equation 11 involve population size (which is assumed to be large). The probability that there exists a MRCA approaches unity when the ratio of A_t to N_t is close to one. When the ratio of A_t to N_t is much less than one, the probability goes to zero. As a result, biparental coalescence times are determined by the ratio of A_t to N_t .

The number of ancestors t generations ago is approximately equal to the mean rate of pedigree growth raised to the t power.

$$A_t \approx \bar{r}^t \quad (12)$$

However, Equation 12 is inexact for two reasons. First, the number of ancestors in a pedigree must be an integer. It takes $g+1$ generations for inbreeding to modify the size of a pedigree (see Figure 1). Even if a pedigree grows geometrically at a rate of 1.6180, the first generation in the past must include two parents. Thus, Equation 12 underestimates A_t when t is small. Secondly, the number of ancestors cannot exceed population size, and as A_t approaches N_t double counting of putative ancestors occurs. Note that this double counting is due to finite population size rather than inbreeding per-se. The eventual inclusion of related individuals in pedigrees imposes a ceiling on the number of ancestors (Ohno, 1996; Shoumatoff, 1985). Thus, Equation 12 overestimates A_t when the number of ancestors t generations ago is similar in magnitude to N_t . Despite these caveats, simulations described in the next section indicate that Equation 12 can be used to derive reasonable estimates of T_{MRCA} .

Estimating T_{MRCA} within a single deme

Equations for T_{MRCA} rely upon the fact that a common ancestor exists when the number of ancestors is close to population size (Chang, 1999). Chang found that biparental coalescence times for outbred populations are given by

$$T_{MRCA, \text{outbred}} = \log_2 N_t \quad (13)$$

Wiuf and Hein argue that Chang's reasoning can be extended to situations where family trees do not double every generation (Donnelly *et al.*, 1999; Hein *et al.*, 2005). Inbreeding is one such scenario, suggesting that the base 2 logarithm of Equation 13 can be replaced by a different logarithm for inbred populations. Here, an equation for the biparental coalescence time of a single, inbred population is derived and tested via multiple approaches. First, the number of ancestors relative to population size yields an estimate of T_{MRCA} . Computer simulations

subsequently verify this approximation. In addition, Chang's 1999 proof is extended to cases where pedigrees do not double every generation in (see Appendix B).

An approximation of T_{MRCA} can be found by setting the number of ancestors equal to population size ($A_t = N_t$). This estimate of T_{MRCA} is affected by changes in population size. Growing populations yield shorter coalescence times than constant-size populations, while shrinking populations yield longer coalescence times than constant-size populations (Kuhner *et al.*, 1998). Let $\bar{w}$ equal the mean fitness of a population, where $\bar{w} = 1.06$ corresponds to a population that grows by 6% every generation. Current population size is a function of population size in the past, mean fitness and time.

$$N_0 = N_t \bar{w}^t \quad (14)$$

Algebraic manipulation of Equation 13 yields:

$$N_t = N_0 \bar{w}^{-t} \quad (15)$$

Equation 12 gives the number of ancestors t generations ago, and Equation 15 gives population size t generations ago. By definition, $t = T_{MRCA}$ at the time of the most recent common ancestor. Setting A_t equal to N_t and rearranging terms yields:

$$\bar{w}^{T_{MRCA}} = N_0 \quad (16)$$

Taking the base $\bar{w}$ logarithm of both sides of the above equation gives an approximation of T_{MRCA} .

$$T_{MRCA} = \log_{\bar{w}} N_0 \quad (17)$$

The validity of Equation 17 is verified by Monte Carlo simulations (see Table 3). These time-forward MATLAB (Mathworks, 2005) simulations begin with a single individual (labeled I) at time t . The number of descendants over time is modeled as a Galton-Watson process, and the number of offspring per individual is assumed to be a Poisson random variable with a mean of $\bar{w}$. Parentage of offspring in subsequent generations is assigned with the restriction that each individual can have at most two parents that are connected to the putative common ancestor (i.e. double counting is allowed). This process is iterated until the lineage of individual I dies out or the entire population in the current generation is related to individual I . The number of generations for the entire population to share common ancestry is recorded. As each of the N_t individuals living at time t is a potential common ancestor, the simulation is run N_t times. The minimum number of generations for these N_t runs is subsequently selected as the T_{MRCA} . Computer code is available in the Supplemental Material.

Estimating global T_{MRCA} for inbred populations

In a previous study (Rohde *et al.*, 2004), evolutionary graphs were used to compute global coalescence times and this method is repeated here. Graphs involve a collection of nodes (demes) that are connected via edges (with connections indicating that migration can occur between a pair of demes). The radius (R) of a graph is the length of the longest path from the center of a graph to the perimeter. Looking forward in time, a common ancestor first appears in a central deme. It takes $\log_{\bar{w}} N_0$ generations for every individual in the initial deme to be

related to the common ancestor. As this is occurring, individuals from the source deme migrate to adjacent demes and establish lineages. Common ancestry spreads outwards via migration to encompass the entire graph. This takes $R+1$ pulses, and Equation 17 gives the duration of each pulse. However, individuals related to the common ancestor can emigrate before every individual in a source deme is related to the common ancestor. This is modeled by treating one of the pulses of common ancestry as a partial pulse. The parameter Δ quantifies the effect of migratory head starts, and ranges from zero to one. Migratory head starts depend on both the connectivity of central nodes and the size of graphs. If there are multiple paths of length R from the center, the extent of head starts is reduced (as shorter global coalescence times require migratory head starts along every path). Let v be the number of nodes neighboring the center of a graph that lie along paths of length R to the periphery. As described in a previous study (Rohde et al., 2004), Δ can be approximated by $(v-1)/v$. A graph with a large radius also has more opportunities for migratory head starts. Consequently, Δ is inversely proportional to the radius of a graph: $\Delta \propto 1/R$. Since both connectedness and graph size can be rate-limiting factors, Δ is defined as:

$$\Delta = \max\left(\frac{v-1}{v}, \frac{1}{R}\right) \quad (18)$$

Computer simulations (see below) assess the suitability of Equation 18. Population sizes change over time, and every deme is assumed to have the same population size. Looking forward in time, the proportion of individuals in a single deme that are related to the common ancestor increases at a rate of $\bar{r}$ each generation. Pulse times for populations follow from Equation 17, and migration rates are assumed to be one migrant per generation ($Nm = 1$). Molecular data indicates that rates of human gene flow are at least this high (Santos et al., 1997). After common ancestry fills the central node, it expands via a number of pulses equal to the radius of the graph (i.e. it takes $R + \Delta$ pulses for the entire population to become related). Considering outbred populations, Rohde *et al.* found

$$T_{\text{global MRCA, outbred}} = (R + \Delta) \log_2 N_0 \quad (19)$$

Rohde et al. demonstrated via probabilistic analysis that Equation 19 provides a good estimate of global coalescence times for large N (see supplemental information of Rohde et al., 2004). Note that Equations 19 and 20 are heuristic approximations that are justified by subsequent computer simulations. Incorporating inbreeding and changing population size yields the following equation for the global coalescence time:

$$T_{\text{global MRCA}} = (R + \Delta) \log_{rw} N_0 \quad (20)$$

As per Rohde *et al.*, global population structure is represented by simple 10-node graph ($R = 3$, $\Delta = 0.333$). Here, nodes refer to sub-Saharan Africa, North Africa, Eurasia, Northeast Asia, North America, Greenland, South America, Indonesia, Australia, and Oceania (Figure 4). Real-world population structures are much more complex than this graph. However, use of a simple 10-node graph allows comparisons with previous studies.

Forward time Monte Carlo simulations allow the validity of Equation 20 to be assessed. Computer simulations were coded in MATLAB (Mathworks, 2005). Each simulation run involves deterministic growth of family trees within demes and stochastic migration between demes. These simulations quantify the effects of graph structure and the number of migrants

per generation (Table 4). Simulations were run 1000 times for each set of parameter values (R , Δ , Nm , $\bar{r}$, $\bar{w}$, and N_t). See Supplemental Material for MATLAB code.

4. Results

Inbred pedigree growth rates are Fibonacci n -step constants

Inbred pedigrees for regular systems of mating are shown in Figure 1. The number of ancestors at each level of a family tree and mating-specific rates of pedigree growth (r_g) are listed Table 1. Pedigrees in which every mating involves siblings contain two individuals each generation. Qualitative differences exist between first cousin and second cousin pedigrees. First cousin matings result in pedigrees that increase linearly as a function of time and the second cousin matings result in pedigrees that increase geometrically as a function of time. The more outbred the type of mating, the closer the rate of pedigree growth is to two. Rates of pedigree growth for second, third, fourth, and fifth cousin matings are 1.6180, 1.8393, 1.9276 and 1.9659, respectively.

Some forms of inbreeding (uncle-niece, aunt-nephew, first cousins once removed, etc.) involve individuals that belong to different generations. In these cases two separate processes reduce the number of ancestors in a pedigree: inbreeding and trans-generation mating. When every mating involves unrelated individuals separated by one generation, the number of ancestors follows the Fibonacci sequence and the rate of pedigree growth is 1.6180 (see Figure 1 and Table 1). When individuals mate later in life, pedigrees are vertically stretched, and the number of ancestors grows a reduced rate. Conversely, when individuals mate early in life the number of ancestors grows at an increased rate. Ancestral pedigrees grow at reduced rates when a single recent ancestor, as opposed to two recent ancestors, is shared (i.e. half-sibling vs. full sibling mating). When only a single recent ancestor is shared, rates of pedigree growth are one level more outbred than in Figure 3A. For example, half-first cousin pedigrees grow at the rate of full-second cousin pedigrees (1.6180 in both cases).

Intriguingly, the golden ratio (ϕ , 1.6180) appears in pedigrees where every mating involves second cousins. While the presence of this number in the growth rate an inbred pedigree may seem unexpected, ϕ is often associated with recursive patterns. The ratio of consecutive terms of a Fibonacci sequence converges to ϕ , as does the rate of pedigree growth for X chromosome and haplodiploid pedigrees (Crow and Kimura, 1970; Livio, 2002). A generalized form of Fibonacci numbers involves summing n subsequent terms to generate the next term of a sequence. The ratio of subsequent terms converges to the Fibonacci n -step constants. The tribonacci ($n = 3$), tetranacci ($n = 4$), and pentanacci ($n = 5$) constants are equal to 1.8393, 1.9276, and 1.9659, respectively (Vajda, 1989). Inspection of Table 1 indicates that rates of growth for third, fourth, and fifth cousin pedigrees are Fibonacci n -step constants.

Global coalescence times

What are realistic levels of inbreeding? Historical data (Bittles and Egerbladh, 2005) yields $\bar{r} = 1.827$ for a population in rural Sweden (Table 2). A large proportion of matings in this population are trans-generation. While other natural populations are expected to have different levels of inbreeding, this data indicates that ancestral pedigrees cannot be assumed to double in size each generation.

The effect of inbreeding on coalescence times is greater in large populations. This is because it takes longer for A_t to approach N_t in large populations, and pedigree size changes are compounded over multiple generations. $\bar{r}$ is much less than two when inbreeding is prevalent and/or a high proportion of matings are trans-generation. Both of these conditions occur in natural populations, suggesting that previous studies underestimate the actual T_{MRCA} .

Approximately 206 million individuals lived 1500 years ago (Cavalli-Sforza *et al.*, 1994). Since then, the global population size of humanity has multiplied 32-fold. Population increases of 6% per generation are consistent with generation times of 25 years ($\bar{w} = 1.06$).

All humanity shares common ancestry in the last few thousand years. Equation 20 gives T_{MRCA} as a function of global population structure, the mean rate of pedigree growth, mean fitness, and current population size. Global coalescence times for outbred populations are 90.2 generations. Incorporating inbreeding ($\bar{F} = 1.827$) yields biparental coalescence times of 102.5 generations. Thus, inbreeding increases T_{MRCA} by 13.69% (Table 4). The radius of an evolutionary graph largely determines biparental coalescence times. Hub-and-spoke graphs have much shorter coalescence times than linear graphs. High migration rates result in reduced coalescence times, while low migration rates result in increased coalescence times. However, doubling the number of migrants per generation does not drastically change coalescence times. MATLAB simulations closely match analytic approximations.

Relative T_{MRCA} for inbred and outbred populations

Using Equation 20, a general relationship for the relative coalescence times of inbred and outbred populations is obtained.

$$\frac{T_{inbred}}{T_{outbred}} = \frac{(R+\Delta)\log_{\bar{r}\bar{w}}N_0}{(R+\Delta)\log_{2\bar{w}}N_0} = \log_{\bar{r}\bar{w}}2\bar{w} \quad (21)$$

The ratio of inbred to outbred coalescence times is 1.1501 for constant-sized populations ($\bar{F} = 1.827$, $\bar{w} = 1.00$), and 1.1369 for growing populations ($\bar{F} = 1.827$, $\bar{w} = 1.06$). Equation 21 indicates that the relative T_{MRCA} of inbred populations is independent of deme size. In addition, graph size does not affect this ratio.

5. Discussion

Fibonacci constants and the decay of heterozygosity

The Fibonacci sequence also appears in equations for the decay of heterozygosity under regular systems of mating (Jennings, 1914; Wright, 1921). Heterozygosity decays geometrically in inbred populations (Crow and Kimura, 1970), with the parameter λ_g indicating the proportion of heterozygosity retained each generation (Figure 3B). Note that these regular systems of mating in Figure 3B involve multiple instances of inbreeding (e.g. double-first cousin and quadruple-second cousin matings), while regular systems of mating in Figure 3C involve single instances of inbreeding (e.g. single-first cousin and single-second cousin matings). Interestingly, values of r_g and λ_g are related. Rates of pedigree growth are twice the proportion of heterozygosity that is retained, but with an offset of two generations (Figure 3A). For example, matings between second cousins result in pedigrees that grow at a rate of 1.6180, two times the proportion of heterozygosity retained each generation by sibling mating (0.8090).

$$r_g = 2\lambda_{g-2} \quad (22)$$

T_{MRCA} for a single deme

Equation 17 is a reasonable approximation for the time until the most recent common ancestor of a single deme (see Table 3). As confirmed by computer simulations, inbred populations have longer coalescence times than outbred populations. In both cases, it takes only a few generations

for the entire population to share common ancestry. Discrepancies between analytic approximations and computer simulations arise because Equation 12 underestimates the number of ancestors when t is small, and it overestimates the number of ancestors when t is large.

Some parameter values fail to result in biparental coalescence. When the product of $\bar{r}$ and $\bar{w}$ is less than one, Equation 17 yields negative coalescence times. Looking backwards in time, this refers to a scenario where population size “expands” faster than an individual’s pedigree grows. However, populations where $\bar{r}\bar{w} < 1$ are highly unlikely to be observed. This is because $\bar{r}$ tends to be slightly less than two, and $\bar{w}$ tends to be close to one.

Inbred populations exhibit greater homozygosity, and recessive lethality leads to increased mortality (Bittles and Neel, 1994; Charlesworth and Charlesworth, 1999). Populations with inbreeding depression have longer coalescence times than populations where such selection is absent, as there is a reduction in present-day individuals relative to the number of ancestors. Note, however, that greater reproductive success has been observed for 3rd and 4th cousins than outbred pairs in Iceland (Helgason *et al.*, 2008).

It is notable Equations 17 and 20 do not incorporate Wright’s inbreeding coefficient (f), where f is defined to be as the probability that two alleles in an individual are identical by descent (Crow and Kimura, 1970). This is because multiple systems of mating can have the same value of f (Nordborg and Krone, 2002; Pollak, 1987; Slatkin, 1991). For example, populations where every mating involves first cousins possess the same inbreeding coefficient as populations where 25% of matings involve siblings and 75% of matings are outbred ($f = 0.0625$). These two scenarios lead to different numbers of ancestors as a function of time. This indicates that multiple types of inbreeding must explicitly be considered when calculating the time until the MRCA of an inbred population (instead of using a summary statistic like f).

Global T_{MRCA}

Biparental coalescence times are remarkably recent (within the last 2500 years). This occurs regardless of the amount of inbreeding. It is likely that Figure 4 underestimates the true complexity of global population structure. However, increasing the size of this graph 4-fold still results in common ancestors who lived in a post-agricultural world. Increasing the number of demes also has the side effect of decreasing within-deme coalescence times. This is because population sizes of individual demes are inversely proportional to the number of demes. Population structure affects both uniparental and biparental coalescence times, and gene trees coalescence much slower than organismal lineages. A migrant has a good chance to become one of the organismal common ancestors of their new deme (so long as they manage to establish a foothold and leave a significant number of great grandchildren). However, the probability that a migrant will become the mtDNA or Y-chromosome common ancestor of a new deme is low ($1/N_{females,t}$ or $1/N_{males,t}$). Because of differences in coalescence times, the timing of the organismal MRCA effectively sets the extreme lower bounds for gene-specific coalescent times.

The net effect of inbreeding is an increase in global coalescence times. This is a direct consequence of smaller rates of pedigree growth for inbred populations. Inbreeding’s effect on global coalescence times is greatest in graphs composed of a relatively few number of large nodes, and is minimized in graphs composed of a large number of small nodes. Somewhat paradoxically, randomly selected pairs of individuals are more likely to be closely related in inbred populations. Randomly selected individuals in an inbred population exhibit greater variance in the degree of relatedness than randomly selected individuals in a panmictic population. When inbreeding is present, subsets of a population share recent common ancestry but coalescence times for the entire population are lengthened.

It is worth noting that the global coalescence times in this paper are crude estimates (as global levels of inbreeding are likely to differ from Swedish estimates and fine-scale population structure is ignored). In addition, high variance in male reproductive success exists in real-world populations (Segurel *et al.*, 2008). Sex-specific variance in reproductive success requires numbers of male and female ancestors to be tracked separately. This scenario causes there to be a reduction in effective population size, leading to a reduction in coalescence times. If levels of inbreeding change over time, generation-specific values of $\bar{r}$ must be calculated. The number of ancestors t generations ago is equal to the product of generation-specific pedigree growth rates. This causes the overall rate of pedigree growth to be equal to the geometric mean of generation-specific rates of pedigree growth.

6. Conclusion

Ultimately, global coalescence times hinge upon within-deme population structure and between-deme patterns of migration. The number of ancestors in a family tree increases very quickly even when there is significant inbreeding. While inbreeding within local populations pushes the global T_{MRCA} back a significant number of years, the qualitative conclusions of previous studies hold. Present-day individuals share common ancestors that lived in the relatively recent past. Exceptions involve isolated demes. However, the majority of humanity is connected via migration. Coalescence times in this paper are most sensitive to the radius of the population graph (i.e. how far the central deme is from peripheral demes), and the number of individuals per deme at the time of the MRCA. Future estimates of biparental coalescence times will benefit from the inclusion of inbreeding, finer resolution evolutionary graphs, and more accurate migratory data. In the words of Alex Haley: “When you start about family, about lineage and ancestry, you are talking about every person on earth” (Marmon, 1977).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

I thank members of Stony Brook University’s Department of Ecology and Evolution, A. Bittles, J. Crow, J. Flowers, L. Jung, A. Onstine, P. Ralph, J. True, S. Yeh, and R. Yukilevich for stimulating discussions and constructive criticism during the preparation of this manuscript. Additional thanks are directed towards my own recent ancestors. This work was supported by an NIH Predoctoral Training Grant (5 T32 GM007964-24).

References

- Ayala FJ. The Myth of Eve: Molecular Biology and Human Origins. *Science* 1995;270:1930–6. [PubMed: 8533083]
- Bittles AH. Consanguinity and its relevance to clinical genetics. *Clinical Genetics* 2001;60:89–98. [PubMed: 11553039]
- Bittles AH, Neel JV. The costs of human inbreeding and their implications for variations at the DNA level. *Nat Genet* 1994;8:117–21. [PubMed: 7842008]
- Bittles AH, Egerbladh I. The Influence of Past Endogamy and Consanguinity on Genetic Disorders in Northern Sweden. *Annals of Human Genetics* 2005;69:549–558. [PubMed: 16138913]
- Cann RL, Stoneking M, Wilson AC. Mitochondrial DNA and human evolution. *Nature* 1987;325:31–6. [PubMed: 3025745]
- Cavalli-Sforza, LL.; Menozzi, P.; Piazza, A. *The History and Geography of Human Genes*. Princeton University Press; Princeton, N.J: 1994.
- Chang JT. Recent common ancestors of all present-day individuals. *Adv Appl Probab* 1999;31:1002–1026.

- Charlesworth B, Charlesworth D. The genetic basis of inbreeding depression. *Genet Res* 1999;74:329–40. [PubMed: 10689809]
- Crow, JF.; Kimura, M. *An Introduction to Population Genetics Theory*. Harper and Row; New York: 1970.
- Donnelly P, Wiuf C, Hein J, Slatkin M, Ewens WJ, Kingman JFC. Discussion: recent common ancestors of all present-day individuals. *Adv Appl Probab* 1999;31:1027–1035.
- Hedrick PW, Kalinowski S. Inbreeding depression and conservation biology. *Ann Rev Ecol Syst* 2000;31:139–162.
- Hein J. Human evolution: pedigrees for all humanity. *Nature* 2004;431:518–9. [PubMed: 15457241]
- Hein, J.; Schierup, MH.; Wiuf, C. *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press; Oxford: 2005.
- Helgason A, Palsson S, Gudbjartsson DF, Kristjansson T, Stefansson K. An association between the kinship and fertility of human couples. *Science* 2008;319:813–6. [PubMed: 18258915]
- Jennings HS. Formulae for the results of inbreeding. *Am Nat* 1914;48:693–696.
- Jobling, MA.; Hurles, M.; Tyler-Smith, C. *Human Evolutionary Genetics: Origins, Peoples and Disease*. Garland Science; New York: 2003.
- Keller LF, Waller DM. Inbreeding effects in wild populations. *Trends Ecol Evol* 2002;17:230–241.
- Kuhner MK, Yamato J, Felsenstein J. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* 1998;149:429–34. [PubMed: 9584114]
- Livio, M. *The golden ratio: the story of phi, the world's most astonishing number*. Broadway Books; New York: 2002.
- Marmon W. Haley's Rx: Talk, Write, Reunite. *Time* 1977;109:72–75.
- Mathworks, MATLAB 7, Mathworks, Natick, MA 2005.
- Matsen FA, Evans SN. To what extent does genealogical ancestry imply genetic ancestry? *Theor Popul Biol* 2008;74:182–90. [PubMed: 18634815]
- Moore LT, McEvoy B, Cape E, Simms K, Bradley DG. A Y-Chromosome Signature of Hegemony in Gaelic Ireland. *Am J Hum Genet* 2006;78:334–8. [PubMed: 16358217]
- Nordborg, M.; Krone, S. Separation of time scales and convergence to the coalescent in structured populations. In: Slatkin, M.; Veuille, M., editors. *Modern Developments in Theoretical Population Genetics*. Oxford University press; Oxford: 2002. p. 194-232.
- Ohno S. The Malthusian parameter of ascents: what prevents the exponential increase of one's ancestors? *Proc Natl Acad Sci USA* 1996;93:15276–8. [PubMed: 8986801]
- Pollak E. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* 1987;117:353–60. [PubMed: 3666446]
- Rohde DL, Olson S, Chang JT. Modeling the recent common ancestry of all living humans. *Nature* 2004;431:562–6. [PubMed: 15457259]
- Santos EJ, Epplen JT, Epplen C. Extensive gene flow in human populations as revealed by protein and microsatellite DNA markers. *Hum Hered* 1997;47:165–72. [PubMed: 9156328]
- Segurel L, Martinez-Cruz B, Quintana-Murci L, Balaesque P, Georges M, Hegay T, Aldashev A, Nasyrova F, Jobling MA, Heyer E, Vitalis R. Sex-specific genetic structure and social organization in Central Asia: insights from a multi-locus study. *PLoS Genet* 2008;4:e1000200. [PubMed: 18818760]
- Shoumatoff, A. *The Mountain of Names*. Simon & Shuster; New York: 1985.
- Slatkin M. Inbreeding coefficients and coalescence times. *Genet Res* 1991;58:167–75. [PubMed: 1765264]
- Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc Natl Acad Sci* 2000;97:7360–7365. [PubMed: 10861004]
- Vajda, S. *Fibonacci and Lucas Numbers and the Golden Section: Theory and Application*. Ellis Horwood Ltd; Chichester: 1989.
- Wright S. Systems of mating. II. The effects of inbreeding on the genetic composition of a population. *Genetics* 1921;6:124–143. [PubMed: 17245959]

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C. The genetic legacy of the Mongols. *Am J Hum Genet* 2003;72:717–21. [PubMed: 12592608]

Appendix A. Derivation of Equation 3

Each term of Equation 1 can be divided by A_{t-1} :

$$\frac{A_t}{A_{t-1}} = \frac{2A_{t-1}}{A_{t-1}} - \frac{A_{t-g}}{A_{t-1}} \quad (\text{A.1})$$

Substituting Equation 2 gives:

$$r_g = 2 - \frac{A_{t-g}}{A_{t-1}} \quad (\text{A.2})$$

A_{t-g} divided by A_{t-1} is equal to $r_g^{-(g-1)}$ giving:

$$r_g = 2 - r_g^{-(g-1)} \quad (\text{A.3})$$

After some algebra:

$$(r_g)^{g-1}(r_g - 2) + 1 = 0 \quad (\text{A.4})$$

Appendix B. Extension of Chang's 1999 Proof

The following appendix sketches how the proof of Theorem 1 in (Chang, 1999) can be extended to situations where pedigrees do not double every generation. Note that these arguments are only heuristic, as a complete extension of Chang's proof is beyond the scope of this paper. Chang's five-stage proof uses base 2 logarithms because outbred pedigrees double in size every generation. A number of differences arise when the base 2 logarithm of Chang's proof is replaced with a different logarithm. However, the central logic of this proof remains unchanged. Stage-specific details are listed below.

For consistency with (Chang, 1999), the notation in this section differs from the main body of the paper. Population size at the time of the MRCA is represented by n and t is the number of generations since the MRCA. Equation 15 allows coalescence times in Appendix B to be translated into the notation used in the main body of this paper:

$$\log_{\bar{r}} n = \log_{\bar{r}\bar{w}} N_o \quad (\text{A.5})$$

$\bar{r}$ is assumed to be between 1.6180 and 2, and per-generation changes in population size are assumed to be small (i.e. $\bar{w}$ is close to one). As will be shown below, it takes approximately $\log_{\bar{r}} n$ generations for an entire population to become related to the most recent common ancestor. Thus, biparental coalescence times follow $T_{MRCA} = \log_{\bar{r}\bar{w}} N_o$ when pedigrees grow at a rate different than two and population size is not constant ($\bar{r} \neq 2.00$ and $\bar{w} \neq 1.00$).

Stage 1: Establishment of a lineage

Start with a single individual labeled I . The number of descendants of individual I can be modeled as a Galton-Watson process. Here, the number of descendants of every individual is distributed as a Poisson random variable with a mean of $\bar{r}\bar{w}$. Note that successful establishment of a lineage is more likely in growing populations (there is a reduced chance of having zero descendants in a given generation). How long does it take for an individual to establish a lineage with at least $(\log_{\bar{r}} n)^2$ descendants (i.e. enough descendants that the lineage that is unlikely to die out)? This takes only a few generations, as the expectation of the process described above is $(\bar{r}\bar{w})^t$ (which quickly approaches $(\log_{\bar{r}} n)^2$ when $\bar{r}\bar{w} \geq 1.5$). From Lemma 3 and Lemma 6 of (Chang, 1999), Stage 1 takes $o \log_{\bar{r}} n$ generations, which is negligible compared to $\log_{\bar{r}} n$.

Stage 2: Geometric increase in the number of descendants without double counting

During Stage 2, the number of descendants of individual I increases geometrically to at least n^β (where β is between 0 and 1). n^β is assumed to be a small enough fraction of n that we do not need to consider individuals that have two parents that are related to individual I (i.e. there is no double counting). After Stage 1, there are sufficiently many descendants for realized pedigree growth rates to converge to $\bar{r}$. This means that the number of descendants related to individual I is multiplied by $\bar{r}$ each subsequent generation. As a result, an approximation for the time spent in Stage 2 is given by:

$$T_{\text{Stage2}} \approx \log_{\bar{r}} n^\beta = \beta \log_{\bar{r}} n \quad (\text{A.6})$$

Stage 3: Half of a population becomes related to a common ancestor

Here, the number of individuals related to the MRCA increases from n^β to $n/2$. Double counting is allowed to occur, and during Stage 3 the fraction of the population related to individual I (g_t) increases at a rate of $\bar{r} - g_t$.

$$g_{t+1} = g_t(\bar{r} - g_t) \quad (\text{A.7})$$

As $\bar{r} \geq 1.618$ and $g_t \leq 0.5$, $\bar{r} - g_t \geq 1.118$. This means that during Stage 3 the proportion of descendants related to the MRCA is multiplied by at least 1.118 each generation. The time it takes for the number of descendants of the MRCA to increase from n^β to $n/2$ is given by:

$$T_{\text{Stage3}} \leq \log_{\bar{r}-g_t} \frac{n/2}{n^\beta} = (1 - \beta) \log_{\bar{r}-g_t} n - \log_{\bar{r}-g_t} 2 \quad (\text{A.8})$$

Note that $\log_{\bar{r}-g_t} \bar{r} < 5$. Converting Equation A.8 to a base $\bar{r}$ logarithm:

$$T_{\text{Stage3}} < 5[(1 - \beta) \log_{\bar{r}} n - \log_{\bar{r}} 2] \quad (\text{A.9})$$

Recall that β is between 0 and 1. This indicates that Stage 3 will take at most $\sim 5 \log_{\bar{r}} n$ generations. Setting β close to 1, Stage 3 takes an arbitrarily small fraction of $\log_{\bar{r}} n$ generations.

Stage 4: The vast majority of the population becomes related to a common ancestor

This stage is unchanged from Chang's proof. Here, the perspective switches from the fraction of the population that is related to individual I to the fraction that is unrelated to individual I (defined as B_t). Since individuals that are unrelated to individual I must have two parents that are also unrelated:

$$(B_{t+1}|B_t, B_{t-1}, \dots) \sim \frac{1}{n} \text{Bin}(n, B_t^2) \quad (\text{A.10})$$

Note that this equation is the same as Equation 7 in (Chang, 1999). From Equation A.10, B_t is expected to square each generation. The net result of Stage 4 is that B_t is reduced from $1/2$ to $n^{-\alpha}$ (where α is between $1/2$ and $2/3$). Via Bernstein's inequality, Stage 4 is completed in order $\log_2 \log_2 n$ generations (i.e. much shorter than $\log_{\bar{F}} n$ generations).

Stage 5: The entire population becomes related to a common ancestor

The final stage takes a single generation, and is unchanged from Chang's proof. If most of the population is related to individual I (i.e. $B_t \leq n^{-\alpha}$, the outcome of Stage 4), it is highly unlikely that individuals in subsequent generations will have two parents that are unrelated to individual I .

Upper and lower bound of T_{MRCA}

The upper and lower bounds of the time until the MRCA follow (Chang, 1999). The important point here is that the majority of time is spent in Stages 2 and 3. Since Stages 2 and 3 take order $\log_{\bar{F}} n$ generations, the coalescence times take approximately $\log_{\bar{F}} n$ generations (i.e. $\log_{\bar{F}N_0} n$ generations). Note that when n is small, the relative amount of time spent in other stages will be greater.

Figure 1.

Uniform pedigrees for different types of inbreeding. Males are represented by squares, females with circles. The proband is indicated with an arrow. Diagonal lines correspond to trans-generation matings. A. Sibling mating, no growth of pedigree size. B. 1st cousin mating, linear growth of pedigree size. C. 2nd cousin mating, geometric growth of pedigree size. D. Half-sibling mating, linear growth of pedigree size. E. Half-1st cousin mating, geometric growth of pedigree size. F. Uncle-niece mating, no growth of pedigree size. G. Trans-generation outbred mating, geometric growth of pedigree size.

Figure 2.

Pedigree growth rates for regular systems of mating are the same forward and backwards in time. All matings involve second cousins ($g=3$). The proband is indicated with an arrow, and individuals that are not direct descendants of the proband are labeled gray. Note that the number of direct ancestors and descendants follows the same series (2, 4, 8, 14, 24...).

A

g	λ_g	r_g
0 (selfing)	.5000	1
1 (sibling)	.8090	1
2	.9197	1
3	.9638	1.6180
4	.9785	1.8393
5	.9918	1.9276
6	.9960	1.9659

B **C** **Figure 3.**

Fibonacci n -step constants, decay of heterozygosity, and rates of pedigree growth. A. Proportion of heterozygosity retained (λ_g) and rates of pedigree growth (r_g) for multiple levels of inbreeding (g). Shared colors indicate values that are multiples of the same Fibonacci n -step constant. B. Decay of heterozygosity due to inbreeding under regular systems of mating (Crow and Kimura, 1970). C. Number of ancestors in an inbred population.

Figure 4.

Global population structure viewed as an evolutionary graph. As per “Modeling the recent common ancestry of all living humans” (Rohde *et al.*, 2004), the graph has 10 nodes, a radius of 3, and a diameter of 5. This graph has two centers (Eurasia and Northeast Asia), and the parameter Δ is minimized if a common ancestor first appears in Eurasia.

Table 1
Number of ancestors as a function of time for different types of mating.

t	outbred	sibling	1 st cousin	2 nd cousin	3 rd cousin	4 th cousin	5 th cousin	half-sibling	half-1 st cousin	uncle-niece	trans-generation
1	2	2	2	2	2	2	2	2	2	1	1
2	4	2	4	4	4	4	4	3	4	2	2
3	8	2	6	8	8	8	8	4	7	2	3
4	16	2	8	14	16	16	16	5	12	2	5
5	32	2	10	24	30	32	32	6	20	2	8
6	64	2	12	40	56	62	64	7	33	2	13
7	128	2	14	66	104	120	126	8	54	2	21
8	256	2	16	108	192	232	248	9	88	2	34
9	512	2	18	176	354	448	488	10	143	2	55
10	1024	2	20	286	652	864	960	11	232	2	89
g	∞	1	2	3	4	5	6	1*	2*	1 & 2	∞
r _g	2	1	1	1.6180	1.8393	1.9276	1.9659	1	1.6180	1	1.6180

The parameter t is the number of generations before the present, g is the number of generations separating an inbred couple from their most recent shared ancestor, and r_g is the mating-specific rate of pedigree growth.

* indicates one, rather than two, shared recent ancestors.

Table 2Estimation of $\bar{r}$ from empirical data.

Relationship	Proportion	r_g
1 st cousin	0.02049	1
2 nd cousin	0.02241	1.618
3 rd cousin	0.02910	1.8393
4 th cousin	0.03832	1.9276
5 th cousin	0.02254	1.9659
6 th cousin	0.00137	1.9836
Same-generation outbred	0.51088	2
Trans-generation outbred	0.28105	1.618
Trans-generation inbred	0.07384	1.618
Total	1.00000	1.827

Modified from Table 1 of “The Influence of Past Endogamy and Consanguinity on Genetic Disorders in Northern Sweden” (Bittles & Egerbladh, 2005). Data comes from marriage records in Skellefteå between 1720–1899. The proportion of outbred trans-generation matings was estimated by comparing the relative proportions of same generation inbred matings relative to trans-generation inbred matings.

Table 3Simulations and analytic approximations of T_{MRCA} for a single population.

Scenario	Simulation	Analytic
Outbred	5.7 (0.48)	6.64
Outbred, large N_t ($N_t=1000$)	9.9 (0.32)	9.97
Outbred, growing population ($\bar{w}=1.06$)	5.9 (0.74)	6.64
Outbred, large N_t , growing population ($N_t=1000$, $\bar{w}=1.06$)	10.8 (0.42)	9.97
Inbred ($\bar{F}=1.827$)	7.0 (0.82)	7.64
Inbred, large N_t ($\bar{F}=1.827$, $N_t=1000$)	12.7 (0.48)	11.46
Inbred, growing population ($\bar{F}=1.827$, $\bar{w}=1.06$)	7.6 (0.52)	7.64
Inbred, large N_t , growing population ($\bar{F}=1.827$, $N_t=1000$, $\bar{w}=1.06$)	15.0 (1.49)	11.46

Coalescence times are in generations with standard deviations of simulations in parentheses. Unless otherwise specified, parameter values are: $\bar{F}=2$, $N_t=100$, $\bar{w}=1.00$. Present-day population sizes (N_0) for growing populations are: 147 (outbred, $N_t=100$), 1787 (outbred, $N_t=1000$), 156 (inbred, $N_t=100$), 1950 (outbred, $N_t=1000$). Analytic approximations use Equation 17. Each simulation was run 10 times for each parameter set. See supplemental material for MATLAB code.

Table 4

Simulations and analytic approximations of global common ancestry times.

Scenario	Simulation	Analytic
Outbred	89.5 (2.6)	90.2
Outbred, high migration ($Nm = 2$)	85.0 (2.1)	< 90.2
Outbred, low migration ($Nm = 0.5$)	93.5 (2.6)	> 90.2
Outbred, hub-and-spoke graph ($R = 1$, $\Delta = 0.889$)	43.9 (0.4)	51.1
Outbred, linear graph ($R = 5$, $\Delta = 0.2$)	138.8 (3.6)	140.7
Outbred, small linear graph ($R = 1$, $\Delta = 1$, $w = 1.00$)	54.6 (1.4)	58.7
Outbred, medium linear graph ($R = 5$, $\Delta = 0.2$, $w = 1.00$)	157.6 (3.0)	152.5
Outbred, large linear graph ($R = 9$, $\Delta = 0.111$, $w = 1.00$)	263.3 (4.4)	267.2
Inbred ($\bar{F} = 1.827$)	103.4 (2.5)	102.5
Inbred, high migration ($\bar{F} = 1.827$, $Nm = 2$)	98.4 (2.5)	< 102.5
Inbred, low migration ($\bar{F} = 1.827$, $Nm = 0.5$)	108.8 (3.0)	> 102.5
Inbred, hub-and-spoke graph ($\bar{F} = 1.827$, $R = 1$, $\Delta = 0.889$)	50.8 (0.6)	58.1
Inbred, linear graph ($\bar{F} = 1.827$, $R = 5$, $\Delta = 0.2$)	162.2 (4.7)	159.9

Coalescence times are in generations with standard deviations of simulations in parentheses. Unless otherwise specified, parameter values are: $R = 3$, $\Delta = 0.333$, $Nm = 1$, $\bar{F} = 2$, $w = 1.06$. Analytic approximations use Equation 20. In each simulation migration between populations is stochastic, while growth of family trees within populations is deterministic. Initial deme sizes: $N_t = 3.50 \times 10^6$ when $w = 1.06$, and $N_t = 6.75 \times 10^8$ when $w = 1.00$. Each simulation was run 1000 times for each parameter set. See supplemental material for MATLAB code.