https://www.scirp.org/journal/ns Natural Science, 2022, Vol. 14, (No. 8), pp: 305-321

Pairwise Shared Genealogical Ancestry in Structured Populations

Philip M. Service

Department of Biological Sciences, Northern Arizona University, Flagstaff, AZ, USA

Correspondence to: Philip M. Service, Philip.Service@nau.edu

Keywords: Population Structure, Pairwise Shared Ancestry, Genealogy, Migration, Most Recent Common Ancestor, Humans

Received: July 5, 2022 **Accepted:** August 7, 2022 **Published:** August 10, 2022

ABSTRACT

Simulation was used to investigate the effects of population structure and migration on metrics of pairwise shared ancestry. Random and hierarchical structures, or migration geometries, were examined. Compared to panmictic populations, progress to all qualitative metrics of pairwise ancestry is delayed in structured populations. However, unless migration is very low, the time required is generally less than triple and often less than twice that required in a panmictic population of the same total size. Population structure also increases, to a similar degree, the time required for a population-wide most recent common ancestor (MRCA). As a result, the relationships between various qualitative metrics of pairwise shared ancestry and MRCA time are relatively unaffected by population structure. For example, the mean time to most recent shared ancestor (MRSA) with global sampling of pairs is 40% - 50% of the MRCA time for almost all simulated structures and migration levels. Quantitative pairwise genealogical overlap is strongly affected by population structure. With global sampling, pairwise quantitative overlap never approaches 1.0, as it does in panmictic populations; and instead eventually becomes stationary at much lower values. Possible implications of the present results for human pairwise shared ancestry are discussed. For globally sampled pairs, the longest time to most recent shared ancestor (MRSA) for humans is suggested to be approximately 2100 years before the present. If generation time is 30 years, then all humans are 69th, or closer, cousins. For people with recent European ancestry, the MRSA time may be only half as long, about 1000 years.

1. INTRODUCTION

In populations with biparental reproduction, the number of ancestors of present-day individuals doubles with each additional past generation. Thus, G generations ago, an individual will have 2^G ances-

tors. For even moderate values of G, the number of ancestors can be greater than the past population size. For example, every present-day individual has approximately 1.1 trillion 40th generation ancestors. There are two necessary consequences of exponential increase in ancestor number. First, distant ancestors will occur many times in the genealogical tree of a present-day individual: that is, for each ancestor, there will be more than one path of descent linking that ancestor to the individual in question. Second, present day individuals will share ancestors with high probability. For example, at some time in the past (and earlier), all present-day individuals will have one (or more) ancestor(s) in common. Particular attention has been given to the most recent common ancestor of the entire population, the MRCA [1-3]. Furthermore, it is virtually certain, at least in panmictic populations, that the most recent shared ancestor (MRSA) of any pair of present-day individuals will have lived more recently than the population-wide MRCA [4].

In a random-mating population of constant size N, the MRCA will have lived very nearly $\log_2 N$ generations in the past [1]. The number of common ancestors increases with additional past generations. Eventually a generation is reached in which all individuals then alive are either ancestors of no one in the present, or ancestors of everyone. That past generation is the generation of most recent identical ancestry (MRIA)—all present-day individuals share an identical set of ancestors. For a constant-size, random-mating population, the MRIA generation will have lived approximately 2 $\log_2 N$ generations in the past [1]. In order to investigate the effects of population subdivision and migration, for which the assumption of random mating is no longer appropriate, simulations provide a flexible, general framework [2, 3]. In the case of randomly generated migration geometries with as many as 500 subpopulations, some generalizations can be made [3]. The number of migrants per generation relative to the number of subpopulations seems to be more important than the total number of migrants. Provided that the number of migrants in each generation is at least twice the number of subpopulations, MRCA and MRIA times increase by less (often much less) than 2.5x compared to a panmictic population of the same total size. That can be true even when the overall migration rate is as low as 0.5% per generation, and when migration between the vast majority of subpopulation pairs is indirect, requiring multiple steps over several generations [3].

Pairwise shared ancestry has received less attention than population-wide common ancestry (MRCA and MRIA). Nonetheless, in the case of a random-mating, constant size population of size N, important results can be summarized [4]. On average, the most recent shared ancestor (MRSA) for random pairs of individuals will have lived slightly more than 0.5 $\log_2 N$ generations in the past: in other words, about half as long ago as the MRCA of the entire population. There is a ~100% probability that a random pair of individuals will share at least one ancestor who lived about 0.7 $\log_2 N$ generations previously, or more recently. Put another way, every individual in the population is related to every other individual by one or more ancestors who lived no longer than 0.7 $\log_2 N$ generations ago, although different pairs will have different ancestors in common. I define ~100% probability of pairwise shared ancestry as *universal pairwise shared ancestry* (UPSA), or universal "cousin-ness". Random pairs will share ≥ 99% of their ancestors who lived about 1.4 $\log_2 N$ generations in the past, considerably more recently than the generation of identical ancestry for the entire population.

The purpose of the present paper is to extend the analysis of pairwise shared ancestry to structured populations. Migration is a prerequisite for shared ancestry when the two individuals of a pair are sampled from different subpopulations. The term "migration" suggests movement between geographically separated (allopatric) subpopulations. However, for purposes of the present analysis, interbreeding between distinct groups that inhabit the same location (sympatric "subpopulations") is equivalent to allopatry with migration. Intermarriage between different sympatric groups might be particularly relevant to humans, for whom sympatric groupings can occur along many dimensions, including race, ethnicity, and religion, among others. In what follows, I will generally use the term "migration", but it should be remembered that "intermarriage" or "interbreeding" are equivalent terms.

The principal question of interest is: how do population subdivision and migration affect metrics of pairwise shared ancestry? I suggest two strategies for comparing the present results to those of earlier studies. First, I ask whether the scaling between panmictic and structured populations is similar for metrics of pairwise shared ancestry and for metrics of population-wide common ancestry. Does the MRSA time, for

example, generally increase by a factor of 2.5 or less in structured vs. panmictic populations, as is the case for the MRCA time? Second, I ask whether the temporal relationships between pairwise and population-wide metrics are similar in structured and random-mating populations. For example, is the UPSA time still about 70% of the MRCA time, as it is in panmictic populations?

Ancestry can be treated as a qualitative or quantitative trait. In the qualitative sense, ancestry is binary (0 or 1)—a member of a previous generation either is (1) or is not (0) an ancestor of a present-day individual in question. The MRCA, MRIA, MRSA, and UPSA metrics, for example, are all qualitative. However, as noted above, sufficiently distant ancestors will occur many times in the genealogy of an individual. Quantitative metrics of shared ancestry compare the frequencies of shared ancestors in genealogical trees [5]. For example, the more similar the frequencies of shared ancestors, the greater the quantitative genealogical overlap, or similarity, of a pair of individuals. In panmictic populations, qualitative and quantitative metrics of pairwise overlap are remarkably similar [4]. For example, random pairs will have quantitative overlap ≥ 0.99 for ancestors who lived only about 1.5 $\log_2 N$ generations in the past—just slightly longer ago than the 1.4 $\log_2 N$ generations required for a similar degree of qualitative overlap.

In general, this study will show that population structure affects qualitative metrics of pairwise shared ancestry as one might expect, based on previous comparisons of MRCA and MRIA times between panmictic and structured populations [3]: population structure delays, but does not prevent, qualitative shared ancestry. On the other hand, the effect of population structure on quantitative metrics of shared ancestry can be profound. The implications of these results for pairwise shared ancestry in the current global human population will be discussed.

2. METHODS

2.1. General Simulation Procedure

The simulation procedure was the same as previously [3, 4]. Programs were written in C language and run on an iMac (Model 18,3; 2017; 4.2 GHz Core i7; 64GB RAM; MacOS 12.3). Total population size was constant, reproduction was monogamous, and generations were non-overlapping. Total population size was 20,000 (qualitative ancestry), or 16,000 (quantitative ancestry). Each analysis is based on 100 or 50 replicate simulations. Each simulation began at Generation 0 and proceeded forward. In each descendant generation, the Generation 0 ancestors of the population were examined for metrics of shared ancestry, which are described below. The population was divided into S subpopulations with mean size m. For (S - 1) subpopulations, sizes were chosen randomly from a uniform distribution with maximum 1.5 m and minimum 0.5 m. The size of the one remaining subpopulation was then fixed by the total population size and sum of the (S - 1) subpopulations. Subpopulation sizes were re-randomized for each replicate. Migration occurred after reproduction in each generation. Migrants represented "excess" reproduction in their source populations. They replaced random individuals in the destination population. There was no guarantee that a migrant successfully reproduced in its destination subpopulation.

Two types of migration geometries were investigated: 1) "random"—for which migration paths between subpopulations were generated randomly; and 2) "hierarchical"—for which subpopulations were assumed to be geographically nested ("states" within "countries", "countries" within "continents", etc.) and migration was greater between "near" subpopulations and less between "distant" subpopulations. A previous study found that random and hierarchical geometries yield generally similar results with regard to MRCA and MRIA times [3]. Consequently, only one instance of a hierarchical geometry was simulated for the present study.

2.1.1. Random Migration Geometries

Migration geometries were created by randomly connecting subpopulations with one-step, one-way migration paths, P [3]. The total number of paths was kS, where k is the one-step path coefficient (2 or 4 for these simulations). On average, each subpopulation sent emigrants to k subpopulations and received immigrants from k subpopulations. If migration between a pair of subpopulations was reciprocal, it was by

chance. With relatively small values of k and large values of S, it is possible that some subpopulations will be isolated, and that Generation 0 members cannot become global (population-wide) common ancestors [3]. Therefore, migration paths were periodically re-randomized during each replicate simulation. Migrants were randomly assigned to the 1-step migration paths. Unless otherwise noted, migration paths were re-randomized after every four generations, and migrants were re-randomized over paths every other generation. There was no guarantee that every path was populated by at least one migrant: in fact, in some cases, there were fewer migrants than migration paths. Although periodic re-randomization of paths and migrants was necessary for practical reasons, it might also more faithfully represent the situation in natural populations, where migration patterns might not be expected to persist unchanged over many generations.

2.1.2. Hierarchical Migration Geometry

If migration is considered in a geographical context, hierarchical or nested migration geometries may be appropriate. In such geometries, migration rates depend inversely upon "distance" between subpopulations. The number of possible geometries is limitless, depending on the number of levels in the hierarchy, the number of units at each level, and migration rates within and between levels. I chose to simulate a single geometry in order to obtain a general sense of the effect of hierarchical models on metrics of pairwise shared ancestry.

A three-level nested model was simulated, consisting of five "continents", each with eight "countries", and each country with five "states" or regions (=200 states or subpopulations). States within countries might also be thought of as culturally-defined subpopulations. Per-generation migration rates were: 5% between states within countries, 1% between countries on the same continent, and 0.4% between continents. For $N = 20{,}000$, there were 1280 migrants per generation (6.4%), which consisted of 1000 migrants between states within countries, 200 between countries on the same continent, and 80 between continents. One-step migration paths were randomly generated and re-randomized every four generations. Migrants were randomly assigned to paths and re-randomized every two generations. The number of one-step paths between states on different continents was 20. Thus, at any time, 40 of the 200 states were origins or destinations of intercontinental migrants. In general, each continent exchanged migrants directly with every other continent. The number of migration paths between countries on the same continent was 80. In most cases, that was probably sufficient to enable one-step migration between all countries within continents. However, the per-generation number of migrants between countries within continents was only 40, so only half the migration paths between countries were "populated" at any one time. One-step migration between all states within a country was assumed and most, if not all, paths had migrants in each generation.

Provided that subpopulation numbers and overall migration rates are similar, there is no systematic difference between hierarchically and randomly structured populations with regard to effects on MRCA and MRIA times [3]. However, hierarchical population structure enables hierarchical sampling. Metrics of pairwise shared ancestry (see below) were, therefore, assessed by sampling pairs of individuals globally, within continents, within countries, and within states. The expectation is that pairs of individuals sampled from the same country, for example, will be more closely related than pairs sampled from the same continent without regard for country. Similar logic applies to comparisons between other levels of the hierarchy. A separate set of 50 replicate simulations was done for each level of sampling.

2.2. Global Common Ancestry

Although not the focus of this paper, population-wide measures of common ancestry provide useful reference points for metrics of pairwise shared ancestry. Considerable attention has been given to the times to the most recent common ancestor (MRCA) and the most recent generation of identical ancestors (MRIA) [1-3].

As noted above, for a panmictic, constant population of size N, the MRCA will have lived very nearly $\log_2 N$ generations previously, and the MRIA generation will have occurred about twice as long ago [1, 3]. It is worth noting that these metrics are based on qualitative ancestry: that is, an individual in an earlier

generation either is, or is not, an ancestor of a present-day individual. Population structure (subdivision) increases the time to the MRCA and MRIA generations [2, 3]. My previous paper [3] had an explicitly forward-looking perspective: Generation 0 was taken to be the present, and common ancestry was framed by asking how long until a present-day individual became the first global common ancestor (FGCA) of the future population; and how long until all members of Generation 0 with surviving lineages became ancestors of every individual in some future generation? I referred to the latter as complete global common ancestry (CGCA).

Although not necessarily identical, the FGCA is, for all practical purposes, equivalent to the MRCA. As noted elsewhere [3], when an individual of Generation 0 becomes the FGCA of the population, say by Generation 15, it is not an absolute requirement that that individual also be the MRCA of that generation. For example, it is possible that the actual MRCA is a Generation 1 descendant of the FGCA. However, simulations show that this must be a rare occurrence because MRCA and FGCA times do not differ more than expected by chance (data not shown). Similar considerations apply to the equivalence of the MRIA and CGCA generations. Given that the present simulations track the fate only of Generation 0 members, the metrics of population-wide common ancestry that are reported here are technically the FGCA and CGCA. However, because the MRCA terminology is more familiar, and because there is no detectable difference in the number of generations involved, I will use MRCA and MRIA for the rest of this paper to refer to landmarks of population-wide common ancestry. If any bias is introduced, it is in the direction of making times to most recent global common ancestry slightly longer than they might actually be. It should be noted that similar considerations apply to the concept of "most recent pairwise shared ancestor" as used in this paper.

2.3. Qualitative Pairwise Shared Ancestry

For each replicate and each generation, random pairs of individuals were examined for pairwise shared qualitative ancestry. Unless otherwise noted, pairs were chosen without regard for subpopulation, and in most cases, the two members of each pair would have been sampled from different subpopulations. There are three metrics of interest that can be estimated by this sampling procedure:

1) What is the probability that a random pair of individuals will share one or more common ancestors from Generation 0, and how does that probability change with time, population structure, and migration rate?

2) What is the time to most recent pairwise shared ancestor (MRSA), and how is that affected by population structure and migration?

3) What is the qualitative overlap (similarity) in ancestry between the two members of a pair, and how is that affected by time, population structure, and migration rate?

With regard to (1), emphasis will be placed on the number of generations required for ≥50% and ≥99.9% probability that random pairs share common Generation 0 ancestors. The latter metric corresponds to the situation in which (almost) every individual in the population is related to every other individual, though not by the same ancestors; and is the operational definition of universal pairwise shared ancestry (UPSA).

With regard to (2), because only Generation 0 ancestry is retained, estimated mean times to most recent pairwise shared ancestry (MRSA) are indirect. As generations progress, the total number of sampled pairs that share Generation 0 ancestors increases: in general $X_{t+1} \geq X_t$, where t and $t+1$ refer to successive generations, and X_t is the number of sampled pairs in Generation t that share Generation 0 ancestors. The data of interest, however, are the number of pairs that have just obtained a shared ancestor as a result of the most recent generation of the simulation. That is estimated as $(X_{t+1} - X_t)$. The same sampled pairs (5000 per generation) were used for (1) and (2).

A separate sample of 5000 pairs from each generation was used to estimate qualitative pairwise genealogical overlap (3). The index of qualitative overlap is the same as used previously [4]. Let individual A have N_A different ancestors from generation 0, and individual B, N_B different ancestors. Let A and B share

N_{AB} ancestors from Generation 0. Then the qualitative genealogical overlap due to shared ancestors who lived G generations in the past is $(N_{AB}/N_A + N_{AB}/N_B)/2$. The range of values for this index is 0 - 1. Two metrics will be presented: the number of generations required for ≥0.5 similarity, and the number of generations required for ≥0.99 overlap. The latter corresponds to the situation where both members of a pair have nearly identical qualitative common ancestry.

Because reproduction was monogamous, shared ancestry necessarily involves pairs of ancestors, or couples. For brevity and clarity, however, I will use phrases such as "most recent shared ancestor", with the understanding that "ancestor" means "ancestor pair". This applies also, of course, to the MRCA of the population.

2.4. Quantitative Pairwise Shared Ancestry

With biparental reproduction, an individual will have 2^G ancestors G generations in the past. In a finite population, the number of possible unique ancestors cannot be greater than the total population size. Consequently, sufficiently distant ancestors must appear multiple times in an individual's genealogy; each instance corresponding to a different pathway of descent through intermediate ancestors. When ancestry is shared between individuals, the frequencies of those shared ancestors provide the data for pairwise quantitative genealogical similarity, or overlap. The metric $q^{(\alpha,\beta)}(G)$ was introduced by Derrida *et al.* [5], and is the overlap between the trees of individuals α and β at generation G in the past. The range of values is 0 - 1. This metric was calculated for a random sample of 5000 pairs in each generation, chosen without regard to subpopulation. The samples were not the same as those used for qualitative overlap.

In a previous paper [4], I introduced the quantity C_{Gij}, which is the number of times Generation 0 member i occurs in the tree of individual j in generation G. If the C_{Gij} for ancestor i are summed over all individuals j in the population in generation G, and the sum divided by 2^G, the resulting metric, $C_{Gi.}$, is the quantitative representation of ancestor i across all individuals in the population in generation G. It can be thought of as the *coefficient of quantitative ancestry* of Generation 0 member i. The scaling factor $1/2^G$ means that the expected value of $C_{Gi.}$ is 1.0 for each ancestor i in every generation, and the sum of $C_{Gi.}$ over all Generation 0 members = N (for a constant-size population). Rather than sum over all individuals in the population, C_{Gij} can be summed separately over the individuals of each subpopulation, s. The resulting values of $C_{Gi.s}$ for a given ancestor i can then be compared across subpopulations. It is of interest to know, for example, if they converge over time as a result of migration. (If the quantity C_{Gij} is divided by 2^G, it becomes $w_\gamma^{(\alpha)}(G)$, which is the "*weight* of an ancestor γ in the tree of individual α at generation G in the past", as defined by Derrida *et al.* [5]. Those weights are used to calculate the pairwise quantitative overlap, $q^{(\alpha,\beta)}(G)$.)

A second metric of quantitative genealogical similarity is the coefficient of variation of C_{Gij} for ancestor i across individuals j in generation G, CV_{Gi} [4]. CV_{Gi} is calculated from the variance of C_{Gij} (scaled by $1/2^G$), using $C_{Gi.}/N$ as the mean. The mean value of CV_{Gi} was calculated in each generation for a sample of 1600 Generation 0 members, excluding from the calculation those members who had no descendants. This is not, strictly speaking, a pairwise measure of genealogical similarity. However, when all individuals have similar frequencies of Generation 0 ancestors, the mean value of CV_{Gi} can approach zero after enough generations, at least in panmictic populations [4].

3. RESULTS

3.1. Panmictic Population

A panmictic population provides a basis for evaluating the effects of population structure and migration on shared ancestry. Results for the metrics used in this paper are shown in **Table 1** (data from [4]). In addition to the number of generations required to reach various metrics of shared ancestry, a second measure is also provided: number of generations relative to the MRCA time. That relative measure will also be given for structured populations, in which case time relative to MRCA refers to the MRCA of the

structured population. It will be of interest to know whether the relationships between various metrics of shared ancestry in structured populations are similar to those in panmictic populations. For example, in panmictic populations, the MRSA time is about 50% of the MRCA time (Table 1). Does a similar relationship hold in structured populations?

3.2. Qualitative Pairwise Shared Ancestry—Random Population Structure

3.2.1. Number of Migrants (M) Increases with Number of Subpopulations (S)

This set of simulations evaluates the effects of subpopulation number, S, and path coefficient, k. The total number of migrants, M, per generation was equal to the number of one-way, one-step migration paths, P (=kS). Thus, the number of migrants per generation increased with number of subpopulations, which under some conditions may be a reasonable assumption [3]. The path coefficient, k, was either 2 or 4. Times to various metrics of shared ancestry relative to own MRCA are little affected by subpopulation number (Figure 1(a), Figure 1(b)). Also, these relative times are essentially the same as for a panmictic population. For example, the time required for the probability of pairwise shared ancestry to be ≥0.999 is about 70% of the MRCA time (P(SA) ≥ 0.999, Figure 1(a), Figure 1(b)). This is slightly longer than the 60% required for the same statistic in a panmictic population (Table 1). On the other hand, the mean MRSA time in structured populations is consistently somewhat less than 50% of the MRCA time, and thus slightly less than the relative time required in a panmictic population (Table 1).

As expected, time required for shared ancestry is greater in structured than in panmictic populations, and the increase is greater when migration paths are sparser: k = 2 (Figure 1(c)) vs. k = 4 (Figure 1(d)). Notably, however, the increased time is generally less than 2.5x, when k = 2; and generally between 1.5x and 2.0x when k = 4. The main exception is the 3.0x increase in time for the probability of pairwise shared ancestry to be ≥0.999 when k = 2 (Figure 1(c)). The decrease in pairwise metrics with increasing number of subpopulations (Figure 1(c), Figure 1(d)) is probably a result of the fact that total number of migrants per generation increased with number of subpopulations. A similar effect has been seen previously for MRCA and MRIA times [3].

Table 1. Metrics of shared ancestry in panmictic populations [4]. Means of 100 replicates.

	Generations	Relative to MRCA
N = 20,000		
MRCA	15.00	-
MRIA	27.73	1.85
MRSA	7.52	0.50
P(Pairwise shared ancestor) ≥ 0.5	8	0.53
P(Pairwise shared ancestor) ≥ 0.999	9	0.60
Qualitative overlap ≥ 0.5	14	0.93
Qualitative overlap ≥ 0.99	19	1.27
N = 16,000		
MRCA	14.83	-
MRIA	27.49	1.85
Quantitative overlap ≥ 0.5	13	0.88
Quantitative overlap ≥ 0.99	20	1.35
Mean CV_{Gi} ≤ 0.10	22	1.48

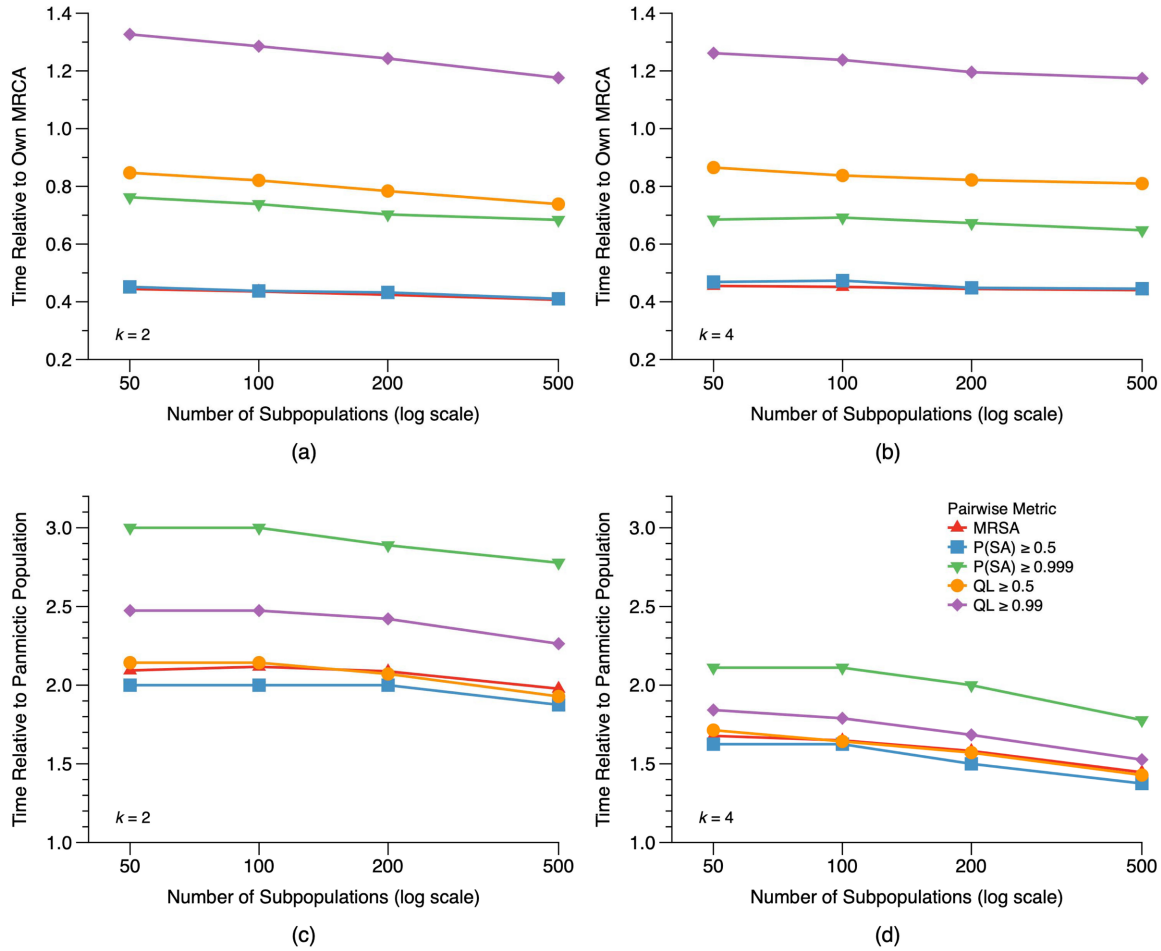

Figure 1. Time required for metrics of qualitative pairwise shared ancestry as a function of subpopulation number, *S*. Number of migrants, *M*, per generation = *kS*. (a) and (b) time relative to own MRCA; (c) and (d) time relative to panmictic population. (a) and (c) $k = 2$; (b) and (d) $k = 4$. P(SA) is the probability that a random pair of individuals share an ancestor; QL is pairwise qualitative overlap for Generation 0 ancestors. $N = 20{,}000$. Means of 100 replicates for each subpopulation number.

3.2.2. Fixed Number of Migrants

To discern the effects of subpopulation number without confounding effects due to changes in number of migrants, simulations were performed with 400 and 1000 total migrants per generation, representing 2% and 5% migration rates, respectively, with path coefficient, $k = 4$. Provided that the number of migrants per generation is at least twice the number of subpopulations, times to metrics of shared ancestry are consistently less than 2.5x those for a panmictic population of the same total size—often much less (Figure 2(c), Figure 2(d)). Pairwise shared ancestry metrics relative to own MRCA time (Figure 2(a), Figure 2(b)) display a pattern similar to that seen in Figure 1(a), Figure 1(b): dependence on subpopulation number is weak or possibly absent, but there is some suggestion of slight decrease with increasing number of subpopulations in the case of qualitative genealogical overlap. Pairwise metrics relative to MRCA are again similar to those in a panmictic population. For example, the probability of shared Generation 0 ancestry is ≥0.5 (P(SA) ≥ 0.5) when time relative to the MRCA is between 0.4 and 0.5 for all simulations shown in Figure 2(a), Figure 2(b). The comparable statistic for an equal-size panmictic population is 0.53 (Table 1).

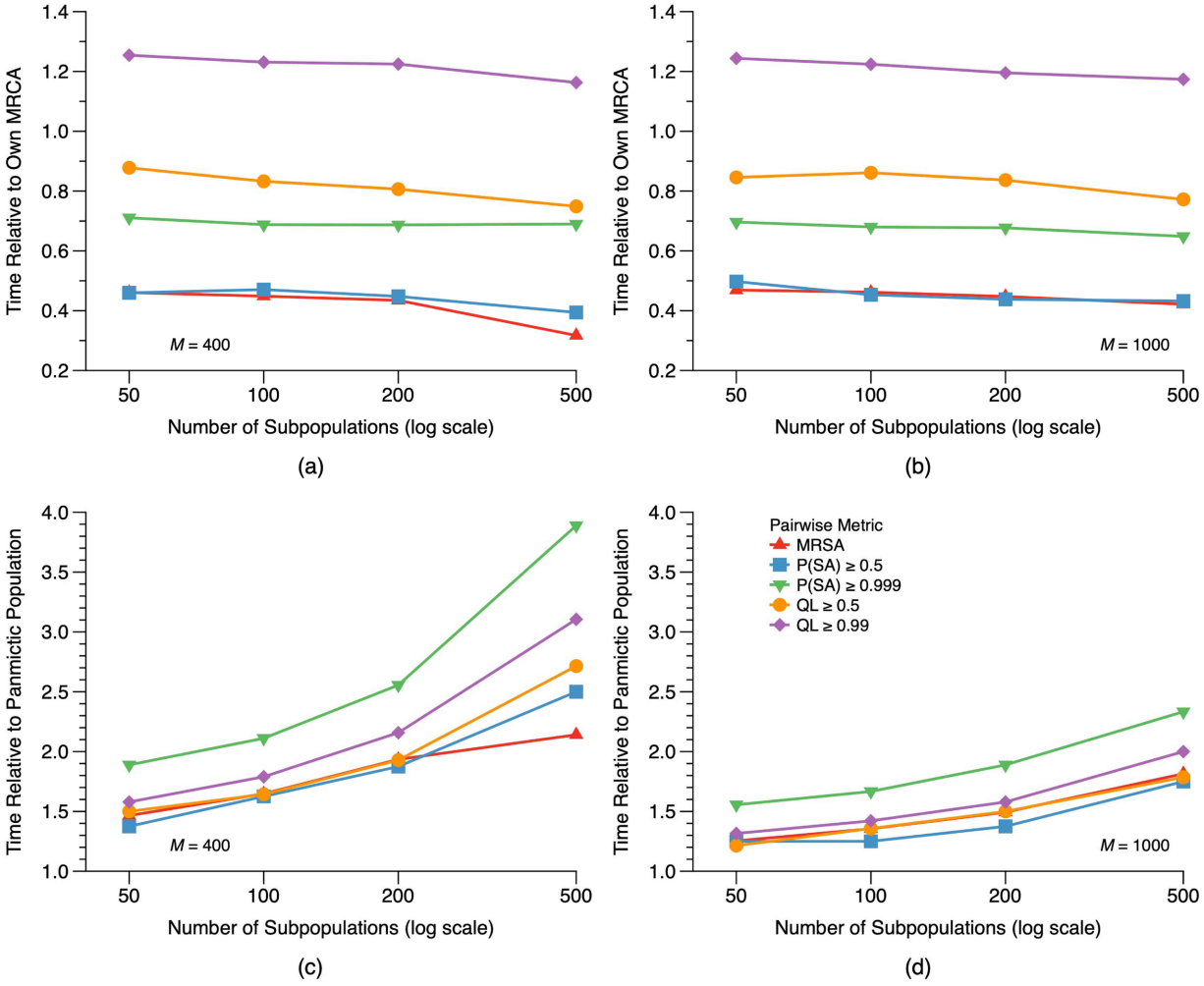

Figure 2. Time to pairwise shared ancestry as a function of number of subpopulations when total migration, *M*, is constant. (a) and (c) *M* = 400; (b) and (d) *M* = 1000. (a) and (b) time relative to own MRCA; (c) and (d) time relative to panmictic population. *k* = 4 in all cases. See Figure 1 for explanation of legend. *N* = 20,000. Means of 100 replicates for each subpopulation number.

3.2.3. Fixed Number of Subpopulations

The third set of simulations investigated the effect of changing number of migrants, *M*, per generation while keeping the number of subpopulations constant at 100, again with path coefficient, $k = 4$. When expressed as time relative to the MRCA, metrics of pairwise ancestry are relatively insensitive to changes in the number of migrants per generation (**Figure 3(a)**). For example, the time required for the probability of pairwise shared ancestry to be ≥0.5 (P(SA ≥ 0.5) varies only between 43% and 46% of the MRCA time. These times are slightly shorter than the 53% for a panmictic population of the same size (**Table 1**). As before, provided the number of migrants per generation is at least twice the number of subpopulations, the time required to reach pairwise metrics of shared ancestry is generally less than 2.5x the time required in a panmictic population, and often much less (**Figure 3(b)**).

3.3. Quantitative Pairwise Shared Ancestry—Random Population Structure

3.3.1. Fixed Number of Migrants

Results for quantitative overlap and CV_{Gi} when number of migrants per generation is fixed are shown

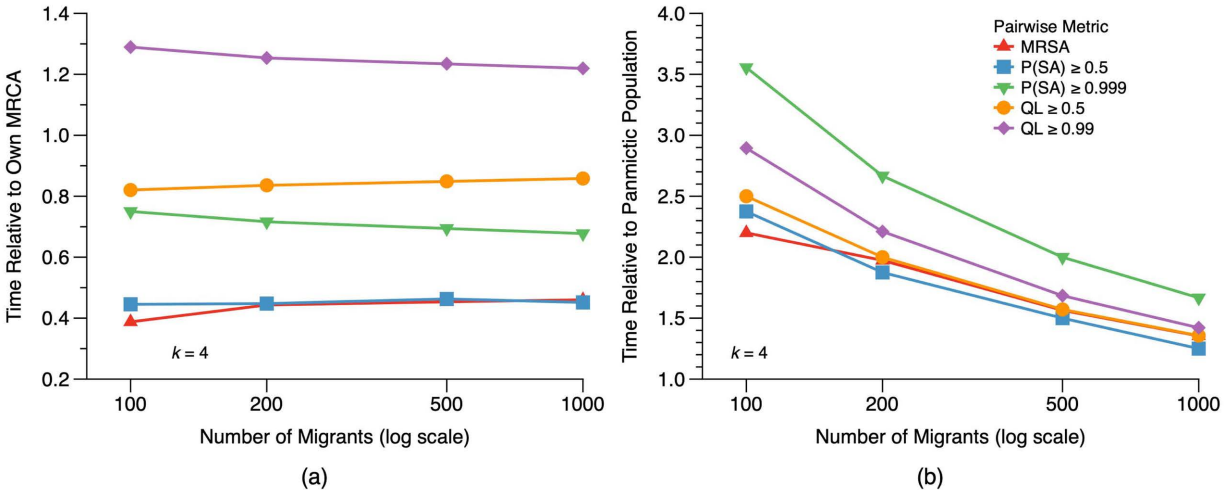

Figure 3. Time to pairwise shared ancestry as a function of number of migrants per generation when number of subpopulations, $S = 100$. $k = 4$. (a) time relative to own MRCA; (b) time relative to panmictic population. $N = 20{,}000$. Means of 100 replicates for each migrant number.

in Figure 4. Population structure has a pronounced effect on both metrics. Quantitative overlap never approaches 1.0, and CV_{Gi} never approaches zero, as they do in panmictic populations (Table 1). This result is not due to an insufficient number of generations—confirmed by allowing five replicates to proceed for 200 generations (data not shown). Rather, quantitative overlap stabilizes at values well below 1.0 somewhat after the MRIA appears, by which time qualitative overlap is necessarily 1.0. When the number of migrants per generation is relatively large ($M = 1000$) and the number of subpopulations small ($S = 50$), the stationary value for pairwise quantitative overlap is about 0.54 (Figure 4(b)). With fewer migrants ($M = 400$) and more subpopulations ($S = 100$), the stationary quantitative overlap is slightly less than 0.1 (Figure 4(a)).

Similarly, the mean coefficient of variation in the representation of Generation 0 members in the trees of later generation individuals becomes stationary at values often well above zero. In the case of $M = 1000$ and $S = 50$, CV_{Gi} stabilizes at about 1.0 after about 40 generations (Figure 4(d)). In the case of $M = 400$ and $S = 100$, CV_{Gi} becomes stationary at about 3.3 after about 50 generations (Figure 4(c)). In both cases, the number of generations required to become stationary is about the same as required for quantitative overlap (Figure 4(a), Figure 4(b)). When $M = 1000$ and $S = 50$, the mean MRIA time is 34.6 generations: when $M = 400$ and $S = 100$, the MRIA time is 46.0 generations.

Simulations were also carried out for the cases where the number of migrants per generation was equal to the number of one-step paths and increased with subpopulation number ($M = P = kS$) and where the number of subpopulations was held constant while the number of migrants was allowed to vary. Qualitatively, those simulations do not change the picture already presented, so the results are not shown. However, with $k = 4$, $S = 500$, and $M = 2000$, the stationary value of quantitative overlap is about 0.8, and the stationary CV_{Gi} is about 0.57. These results suggest that when subpopulations are densely connected and migration rates are very high, quantitative pairwise metrics can at least approximate those of panmictic populations.

3.3.2. Quantitative Representation of Generation 0 Ancestors by Subpopulation

To gain a better understanding of why quantitative overlap and CV_{Gi} become stationary in structured populations, it may be useful to look at the contributions of individual Generation 0 members to the trees of later generations, and to do so on a subpopulation-specific basis. Results for two representative Generation 0 individuals are shown (Figure 5). In both cases, total population size was 16,000, there were 50 subpopulations and the one-step path coefficient, k, was 4. The metrics shown are C_{Gi}, which is the coefficient of quantitative ancestry of Generation 0 individual i summed over the entire population in Genera-

tion G, and $C_{Gi.s}$, which is the coefficient of quantitative ancestry of that same individual summed separately over descendants in each of the 50 subpopulations. The sum of $C_{Gi.s}$ over subpopulations equals $C_{Gi.}$. Because $C_{Gi.s}$ is a sum, it's magnitude is affected by the number of individuals in each subpopulation. In order to avoid any confounding effects of variation in subpopulation size, equal subpopulations were used for these simulations.

The reason why quantitative overlap never approaches 1.0 and the coefficient of variation never approaches 0, as they do in panmictic populations (**Table 1**), is now clear. The subpopulation-specific coefficients of quantitative ancestry, $C_{Gi.s}$, for a given Generation 0 ancestor never converge completely. Instead, they become stationary at different values (**Figure 5(b)** and **Figure 5(d)**). Also, the population-wide coefficient of quantitative ancestry for a given Generation 0 ancestor, $C_{Gi.}$, becomes stationary in the same time frame (blue lines in **Figure 5(a)** and **Figure 5(c)**). This latter behavior mimics what happens in panmictic populations [4, 5], although population structure means that it takes longer.

When migration rates are higher, we might expect the $C_{Gi.s}$ values to converge more closely before becoming stationary. This appears to be the case (compare **Figure 5(a)** and **Figure 5(c)**) and is consistent with higher stationary values of pairwise overlap (**Figure 4(b)** vs. **Figure 4(a)**) and lower coefficients of variation (**Figure 4(d)** vs. **Figure 4(c)**).

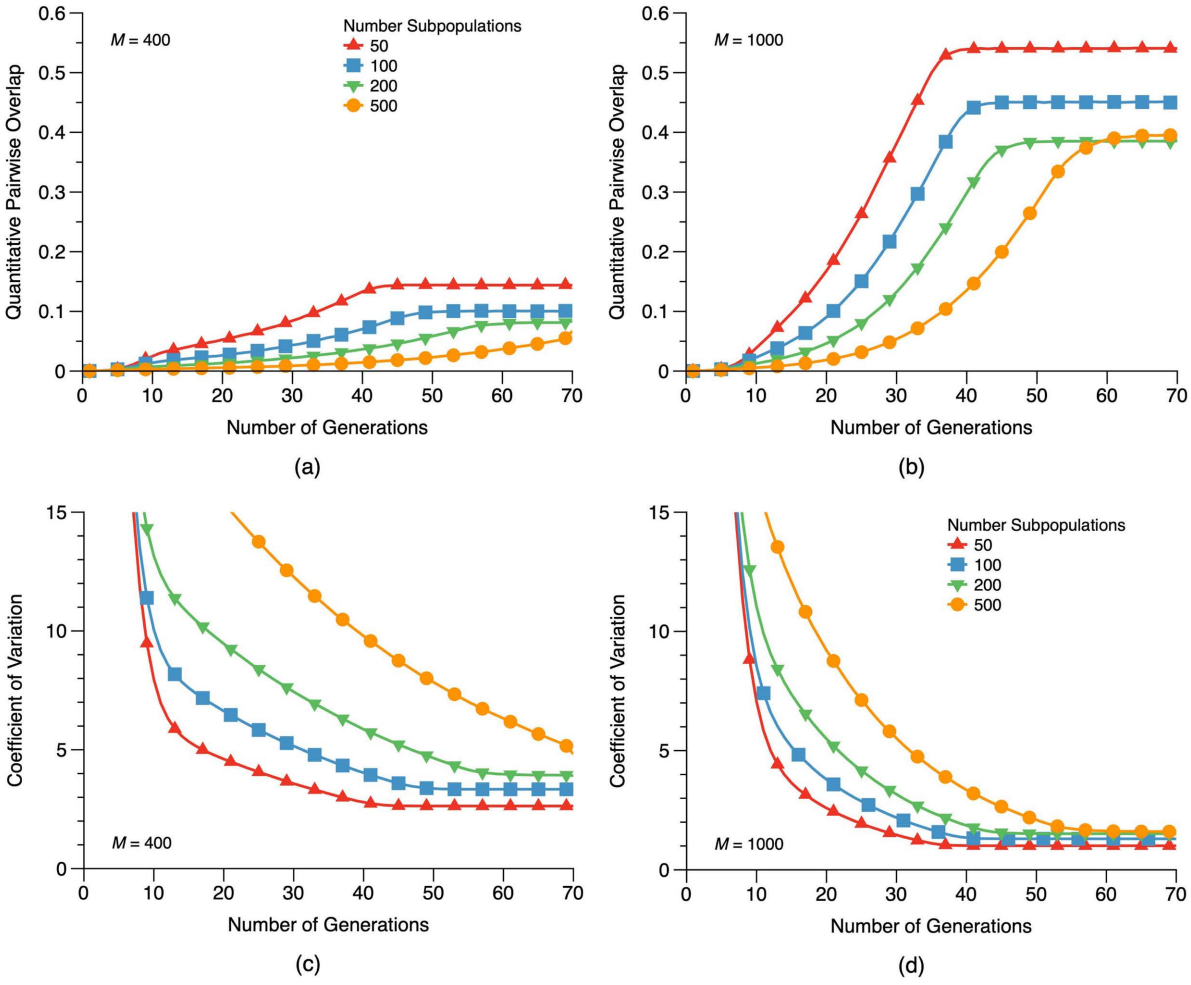

Figure 4. Time to pairwise shared ancestry as a function of number of subpopulations when number of migrants per generation, M, is constant. $k = 4$. (a) and (c) $M = 400$; (b) and (d) $M = 1000$. (a) and (b) quantitative pairwise overlap; (c) and (d) coefficient of variation, CV_{Gi}. $N = 16{,}000$. Means of 50 replicates.

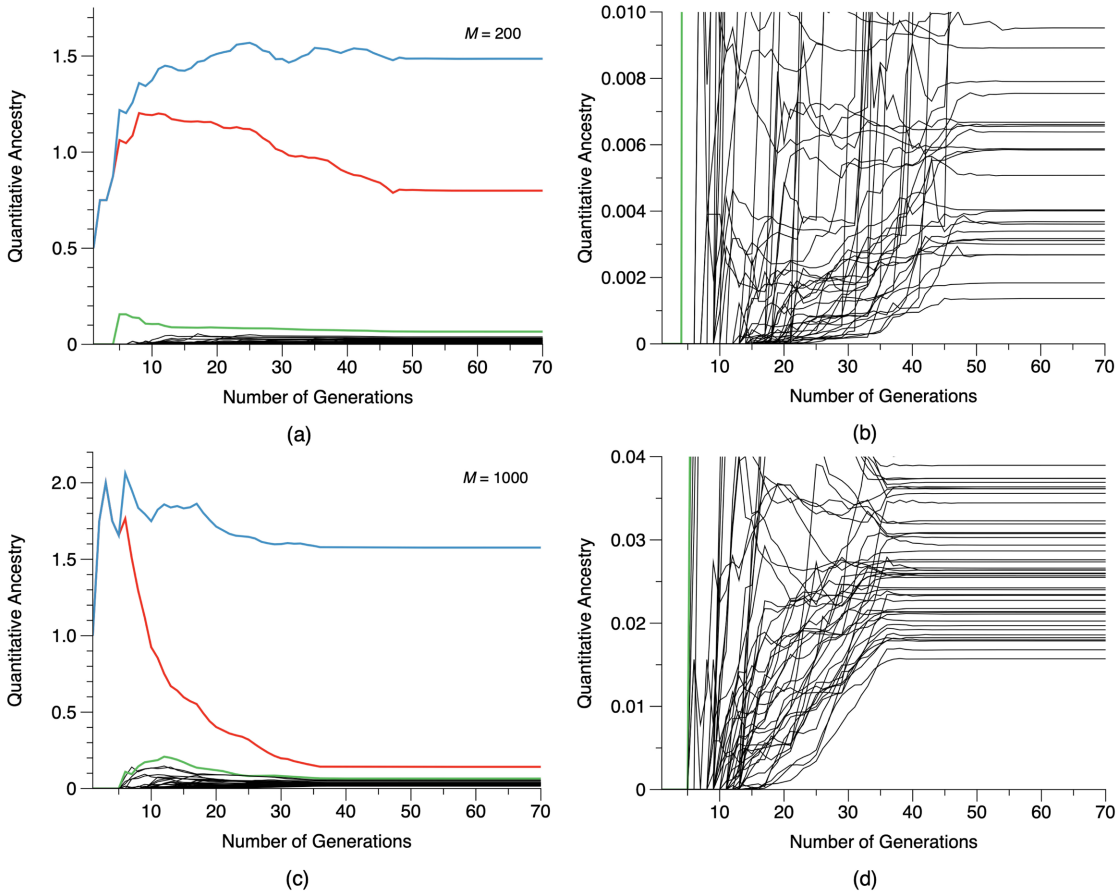

Figure 5. **Coefficient of quantitative ancestry for two Generation 0 ancestors, by subpopulation. (a) and (b) are plots for same ancestor, as are (c) and (d); the right-hand plot in each case expands the vertical axis in order to increase resolution. The blue lines are the population-wide coefficient, $C_{Gi.}$, for each ancestor. Red lines are the ancestry coefficient for each ancestor in its natal subpopulation. Green lines serve as reference points that are present in both plots for each ancestor. The remaining 48 black lines (not all distinct) are the $C_{Gi.s}$ values for the Generation 0 ancestor in the remaining 48 subpopulations. $N = 16{,}000$, $k = 4$, $S = 50$, $M = 200$ (a and b), $M = 1000$ (c and d).**

3.4. Hierarchical Population Structure

3.4.1. Qualitative Metrics of Shared Ancestry—Global Sampling

Qualitative metrics of common and pairwise shared ancestry for the hierarchical model are summarized in Table 2. The pairwise metrics are based on global sampling—the members of each pair were chosen randomly from the entire population. In addition to the actual number of generations required to achieve each metric, results for all metrics are shown relative to the comparable metric for a panmictic population of the same size (Table 1). Also, the pairwise metrics are shown relative to the MRCA time of the hierarchical model itself ("own MRCA"). Both sets of relativized metrics are very similar to values for random geometries. For example, the MRSA time for the hierarchical model is 1.9x and the time to qualitative overlap ≥ 0.99 is 2.21x that of a panmictic population. The corresponding results for random geometries are shown in Figure 1(c) and Figure 1(d); Figure 2(c) and Figure 2(d); and Figure 3(b). Similarly, the MRSA time is 0.45 and the time to qualitative overlap ≥ 0.99 is 1.32 relative to own MRCA. The comparable statistics for random geometries are in Figure 1(a), Figure 1(b); Figure 2(a), Figure 2(b); and Figure 3(a).

Table 2. Pairwise shared ancestry with hierarchical population structure and global sampling. Means of 50 replicates for pairwise metrics, and 200 replicates for MRCA and MRIA times. N = 20,000. P(SA) is probability of shared ancestry.

Metric	Generations	Relative to Panmictic Population (Table 1)	Relative to Own MRCA
MRCA	31.80	2.12	
MRIA	56.95	2.05	
MRSA	14.31	1.90	0.45
P(SA) ≥ 0.5	15	1.88	0.47
P(SA) ≥ 0.999	23	2.56	0.72
Qualitative overlap ≥ 0.5	27	1.93	0.85
Qualitative overlap ≥ 0.99	42	2.21	1.32

This is just one of many possible hierarchical migration models, and one should be cautious about generalizing. That said, this particular model behaves similarly to random models with respect to population-wide measures of common ancestry; and with respect to pairwise metrics of shared ancestry, so long as pairs are sampled globally, as they were for random models. The MRCA and MRIA times are slightly more than twice those of a panmictic population of the same size. This result is entirely consistent with previous results for a variety of random and hierarchal models [3].

3.4.2. Qualitative Metrics of Shared Ancestry—Hierarchical Sampling

As would be expected, sampling at progressively deeper levels of the nested population structure results in progressively shorter times to achieve various metrics of qualitative shared ancestry (Table 3, Figure 6). The one exception is that the time required for a high degree of qualitative genealogical overlap (≥ 0.99) is almost unaffected by sampling method (Figure 6(b)): the required number of generations decreases from 42 for global sampling to 40 for within-state sampling. Qualitative overlap behaves as one might expect initially, increasing in a logistic fashion and increasing most quickly in states and least quickly globally. However, all three nested sample types begin to deviate from logistic growth once overlap is greater than 0.5 - 0.6. The divergence is presumably a result of cumulative effects of persistent immigration, as the Generation 0 ancestors of immigrants will tend to differ from those of natives.

Relative MRSA times (and other metrics) conform to the expected pattern: global > within continent > within country > within state. Beyond that, there is little to say. For example, quantitative relationships between nested sample metrics are likely to be specific to this model instance. In this particular case, the within-country MRSA time is about 60% of the within-continent MRSA time (Table 3). This result is not generalizable because it is likely to depend on the number of nested levels and on migration rates within and between levels. For the same reason, there is no value in comparing the within-country MRSA time, for example, to the global MRCA time.

3.4.3. Quantitative Metrics of Shared Ancestry—Hierarchical Sampling

Quantitative genealogical overlap is greatest for within-state samples and least for global samples, as expected (Figure 7). Overlap approaches 1.0 for pairs sampled from the same state; but is only about 0.1, and apparently at equilibrium within about 70 generations, for globally-sampled pairs. The result for global pairs is similar to that for the random migration models (Figure 4(a), Figure 4(b)). For reasons already explained, the actual values of quantitative overlap are specific to this model and not, therefore, generalizable.

The coefficient of variation in quantitative ancestry due to Generation 0 individual i in later generations, CV_{Gi}, is not based on pairwise sampling. As such, it is not amenable to nested sampling and is not considered in this context.

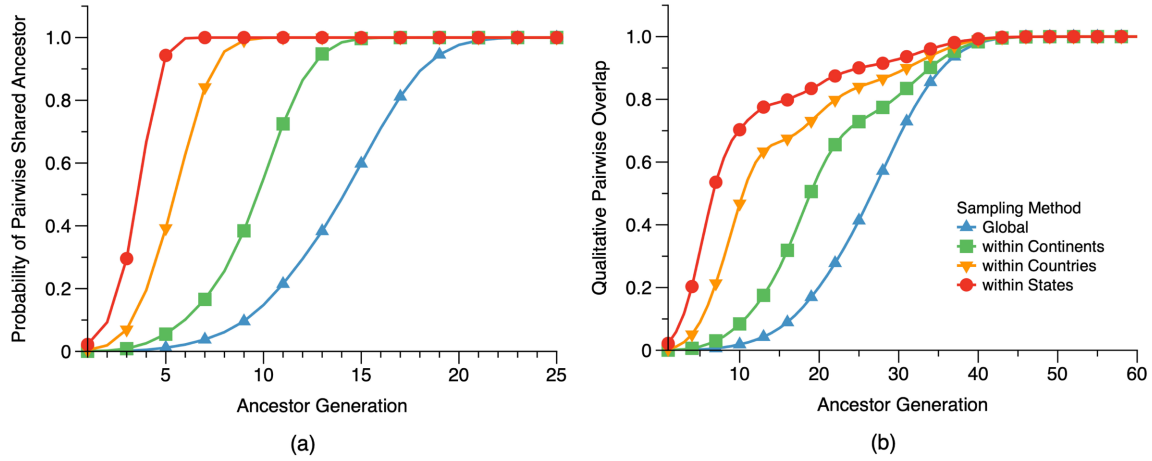

Figure 6. Pairwise shared ancestry with hierarchical sampling. (a) Probability of shared ancestry; (b) Qualitative pairwise overlap. $N = 20{,}000$, means of 50 replicates for each sample type. Note different horizontal axis scales.

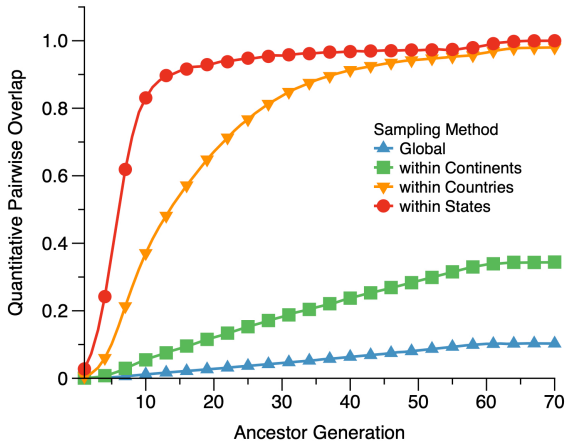

Figure 7. Quantitative pairwise overlap with hierarchical sampling. $N = 16{,}000$, means of 50 replicates for each sample type.

Table 3. Mean time to most recent pairwise shared ancestor (MRSA) with nested sampling. Means of 50 replicate simulations.

Sample Type	MRSA (generations)
Global	14.31
Within Continents	9.93
Within Countries	5.91
Within States	3.98

4. DISCUSSION

4.1. Comparisons to Panmictic Populations

Qualitative measures of pairwise shared ancestry are affected by population structure in essentially

the same way as are MRCA and MRIA times [3]. In most cases, times to pairwise metrics are less than triple what they would be in a panmictic population of the same size. When the number of migrants per generation is appreciably larger than the number of subpopulations and/or each subpopulation exchanges migrants with several other subpopulations ($k > 2$), shared ancestry times may be less than twice those for a random mating population of the same total size (**Figure 1(d)**, **Figure 2(c)**, **Figure 2(d)**, **Figure 3(b)**). Longer times (> 3x) to shared ancestry are likely when the number of migrants per generation is less than twice the number of subpopulations and/or migration routes are sparse ($k \leq 2$) (**Figure 1(c)**, **Figure 2(c)**, **Figure 3(b)**). The pairwise metric that is most strongly dilated by population structure is time to universal shared ancestry (UPSA), defined here as the time required for the probability of pairwise shared ancestry to be ≥ 0.999 (**Figure 1(c)**, **Figure 1(d)**, **Figure 2(c)**, **Figure 2(d)**, **Figure 3(b)**).

Quantitative ancestry is fundamentally altered by population structure. In a panmictic population, pairwise quantitative genealogical overlap is approximately 1.0 well before the MRIA time [4]. In structured populations with global sampling, quantitative overlap never approaches 1.0, and may become stationary at much lower values, depending upon number of subpopulations and migrants, and density of one-step migration paths (**Figure 4(a)**, **Figure 4(b)** and **Figure 7**). However, with hierarchical population structure, pairs sampled within deeply nested levels can achieve high values of quantitative overlap. For example, pairs sampled from the same state or same country achieve overlap values approaching 1.0, given enough time; while overlap for pairs sampled globally becomes stationary at about 0.10 (**Figure 7**). Precise results will be different for other hierarchical models, but in general we expect that quantitative overlap will be low for pairs sampled globally, and may be relatively high for pairs sampled from deeply nested levels.

4.2. Comparisons to Own MRCA

Absolute time (generations) to various metrics of pairwise shared ancestry is affected by population structure, as described in the preceding section. In contrast, those same metrics are remarkably insensitive to population structure when divided by the MRCA time of the same population (own MRCA). For example, MRSA time relative to MRCA is between 0.4 and 0.5 for all but one or two of the simulations reported here, and for those exceptions, the value is slightly less than 0.4 (**Figure 1(a)**, **Figure 1(b)**, **Figure 2(a)**, **Figure 2(b)**, **Figure 3(a)**; **Table 2**). The comparable value for a panmictic population of the same size is 0.5 (**Table 1**), suggesting that MRCA time is more strongly affected by population structure than is MRSA time. The time required for the probability of shared ancestry to be ≥0.5 is essentially the same as the MRSA time in all simulations, when both metrics are expressed relative to own MRCA. The time required for universal pairwise shared ancestry (probability ≥ 0.999), relative to own MRCA, is between 0.65 and 0.76 for all simulations reported here (**Figure 1(a)**, **Figure 1(b)**, **Figure 2(a)**, **Figure 2(b)**, **Figure 3(a)**; **Table 2**). The same metric for a panmictic population is 0.6, again suggesting that the UPSA time is more strongly dilated by population structure than are other metrics, or than is MRCA time. Similar statements can be made about the relative times to qualitative pairwise overlap ≥ 0.5 and ≥0.99—they are only very weakly affected by population structure.

4.3. Considerations for the Global Human Population

The insensitivity of qualitative pairwise metrics to specific details of population structure and migration, provided that they are based on global sampling and that they are expressed relative to own MRCA, suggests that the present results can be cautiously generalized to other contexts. Models of the global human population argue that the MRCA of present-day humans lived approximately 3400 to 2300 years ago, depending on the type of model and assumptions about historic migration rates [2]. For the sake of further discussion, I will take 3000 years as the global human MRCA time. The present results then suggest that the MRSA of globally-sampled random pairs of humans will have lived between 1200 (0.4x) and 1500 (0.5x) years ago, on average. Similarly, the probability that a random pair of individuals will share an ancestor who lived between 1200 and 1500 years ago, or more recently, is ≥0.5. The time to universal pair-

wise shared ancestry, UPSA, might be 2100 (0.7x) years before the present. That is, with probability ≥ 0.999, all pairs of humans share an ancestor who lived within the past 2100 years. If that is correct, and assuming a generation time of 30 years, then all humans are 69th, or closer, cousins. This argument extends to the remaining qualitative pairwise metrics. With factors of 0.85 and 1.25 (**Figure 1(a)**, **Figure 1(b)**, **Figure 2(a)**, **Figure 2(b)**, **Figure 3(a)**; **Table 2**), times to ≥0.5 and ≥0.99 qualitative overlap are 2550 and 3750, respectively, years before the present.

The Case of Charlemagne

Humphrys suggests that that all people "in the West" (*i.e.*, people with European ancestry) "probably" descend from the Holy Roman Emperor Charlemagne (747-814 CE) [6]. The argument had three supports. First, Charlemagne has many proven living descendants (assuming accurate paternity in intervening generations). Second, if Charlemagne has *any* living descendants, he must have a very large number of them. His tree of descendants must contain on the order of 2^{40}, very possibly more, pathways to presently living people [3]. This calculation assumes a 30-year generation time and that, therefore, Charlemagne died approximately 40 generations ago. It should be noted that 2^{40} is more than a hundred times greater than the current global human population, about 2^{33}. Third, as these and other simulations show, even small amounts of migration and intermarriage are sufficient to ensure shared ancestry among people occupying different social classes or countries [2, 3]. It should also be noted that there is nothing special about Charlemagne other than that he has documented present-day descendants. Everyone else who lived in Europe at that time, man or woman, who also has *any* currently living descendants is just as likely to be a common ancestor of all Europeans.

If Charlemagne is a common ancestor of Europeans, what do these simulations suggest about pairwise shared ancestry among Europeans? For brevity, I will use the term "European" to mean everyone who has *any* ancestors who lived in Europe within the past two or three centuries, regardless of where they currently live and regardless of how they identify themselves racially or ethnically. If Charlemagne is a common ancestor of all Europeans, then random pairs of Europeans must share ancestors who lived more recently than Charlemagne with probability ≥ 0.999. In other words, the UPSA time of Europeans must be less than 1200 years ago. That is so because when universal pairwise ancestry occurs, all pairs do not share the same ancestor. More time is required for a common pairwise ancestor: it would be the MRCA time of Europeans.

Is it reasonable that the European UPSA time is less than 1200 years? Europeans are effectively a nested subpopulation of all humans. I have shown that nested samples result in shorter times to various metrics of qualitative pairwise shared ancestry (**Table 3**). If the global human UPSA time is approximately 2100 years, as suggested above, then we certainly expect the European UPSA time to be less. It seems reasonable that it might be only half as long, 1050 years, but the current simulations can only be suggestive on this point. If true, it would mean that all Europeans are 34th, or closer, cousins.

5. CONCLUSION

The general features of pairwise shared ancestry in panmictic populations [4] also apply to structured populations when the latter are sampled globally. In particular, pairwise shared ancestry proceeds much more quickly than population-wide common ancestry. To be sure, population structure produces longer times to all qualitative metrics of pairwise and common ancestry. However, all metrics are affected to a roughly similar degree. The result is that in both panmictic and structured populations, the most recent shared ancestor (MRSA) of random pairs, for example, will have lived about half as long ago as the most recent common ancestor (MRCA) of the entire population. And, in both cases, universal pairwise shared ancestry (UPSA) will be true for ancestors who lived about 70% as long ago as the MRCA, or more recently. The chief exception to this generalization of similarity between panmictic and structured populations involves quantitative pairwise overlap. In panmictic populations, quantitative overlap approaches very close to 1.0 by about 1.5x the MRCA time. In structured populations with global sampling, however, quan-

titative overlap becomes perpetually stationary at values often much less than 1.0. Furthermore, considering ancestors who lived before the MRIA generation, the genealogical trees of any pair of present-day individuals are likely to be quantitatively very different, but must also be qualitatively identical.

ACKNOWLEDGEMENTS

I thank Kiisa Nishikawa for many helpful comments on this article.

FUNDING

This research did not receive any funding.

CONFLICTS OF INTEREST

The author declares no conflicts of interest regarding the publication of this paper.

REFERENCES

1. Chang, J.T. (1999) Recent Common Ancestors of All Present-Day Individuals. *Advances in Applied Probability*, **31**, 1002-1026. https://doi.org/10.1239/aap/1029955256
2. Rohde, D.L.T., Olson, S. and Chang, J.T. (2004) Modelling the Recent Common Ancestry of All Living Humans. *Nature*, **431**, 562-566. https://doi.org/10.1038/nature02842
3. Service, P.M. (2021) The Future Common Ancestry of All Present-Day Humans. *Natural Science*, **13**, 117-132. https://doi.org/10.4236/ns.2021.134011
4. Service, P.M. (2022) Pairwise Shared Ancestry in Random-mating Constant-Size Populations. *Natural Science*, **14**, 193-202. https://doi.org/10.4236/ns.2022.145019
5. Derrida, B., Manrubia, S.C. and Zanette, D.H. (2000) On the Genealogy of a Population of Biparental Individuals. *Journal of Theoretical Biology*, **203**, 303-315. https://doi.org/10.1006/jtbi.2000.1095
6. Humphrys, M. Royal Descents of Famous People. https://humphrysfamilytree.com/famous.descents.html