

# Pairwise Shared Ancestry in Random-Mating Constant-Size Populations

Philip M. Service

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

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

**Keywords:** Pairwise Shared Ancestry, Genealogical Overlap, Quantitative Ancestry, Most Recent Common Ancestor, Number of Cousins

**Received:** April 5, 2022

**Accepted:** May 21, 2022

**Published:** May 24, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## ABSTRACT

In a panmictic population of constant size  $N$ , random pairs of individuals will have a most recent shared ancestor who lived slightly more than  $0.5 \log_2 N$  generations previously, on average. The probability that a random pair of individuals will share at least one ancestor who lived  $0.5 \log_2 N$  generations ago, or more recently, is about 50%. Those individuals, if they do share an ancestor from that generation, would be cousins of degree  $(0.5 \log_2 N) - 1$ . Shared ancestry from progressively earlier generations increases rapidly until there is universal pairwise shared ancestry. At that point, every individual has one or more ancestors in common with every other individual in the population, although different pairs may share different ancestors. Those ancestors lived approximately  $0.7 \log_2 N$  generations in the past, or more recently. Qualitatively, the ancestries of random pairs have about 50% similarity for ancestors who lived about  $0.9 \log_2 N$  generations before the present. That is, about half of the ancestors from that generation belonging to one member of the pair are present also in the genealogy of the other member. Qualitative pairwise similarity increases to more than 99% for ancestors who lived about  $1.4 \log_2 N$  generations in the past. Similar results apply to a metric of quantitative pairwise genealogical overlap.

## 1. INTRODUCTION

Considerable attention has been given to the topic of population-wide common ancestry: in particular to the question of how many generations ago did the common ancestor of the present population live? In the case of an undivided, random-mating population of constant size  $N$ , the answer can be derived analytically. With bi-parental reproduction, the most recent genealogical common ancestor (MRCA) of all present-day individuals will have lived very nearly  $\log_2 N$  generations previously [1]. For example, if the population size is one billion, the time to the MRCA will be about 30 generations. The number of present-day common ancestors increases with progressively earlier generations, until a generation is

reached from which all present-day individuals share the exact same set of ancestors. That is the generation of most recent identical ancestry (MRIA), and it will have occurred about  $2 \log_2 N$  generations in the past [1]. In the case of subdivided (non-random-mating) populations, the MRCA and MRMA times can be estimated by simulation for various degrees of population structure and migration (or intermarriage) [2, 3].

Other aspects of genealogical relatedness, beyond the MRCA and MRMA, appear to have received less attention. For example, what is the time to the most recent shared ancestor of a random pair of individuals? Or, equivalently, how closely related are random pairs? How many currently living relatives, due to shared ancestors from a specified earlier generation, is any individual likely to have? In other words, if we focus on ancestors who lived  $G$  generations ago, how many present-day cousins of degree  $(G - 1)$  are expected? How many ancestors are random pairs of present-day individuals likely to share from each earlier generation? That is, what is the pairwise degree of genealogical overlap for ancestors in previous generations? Lastly, how many generations in the past must one look to find that every present-day individual is related to every other individual in the population?

Most analyses consider ancestry in qualitative, binary terms (0 or 1). An individual in the past is (1) or is not (0) ancestor of a present-day individual in question. The MRCA and MRMA, for example, are qualitative metrics. However, with biparental reproduction, number of ancestors doubles with each additional generation in the past. For example, an individual will have  $2^{30}$ , or more than one billion, 30<sup>th</sup>-generation ancestors. Clearly, the number of unique ancestors cannot exceed the past population size. Therefore, sufficiently distant ancestors will occur multiple times in the genealogy of an individual, and shared ancestry can be treated as a quantitative, as well as qualitative, variable.

For simplicity, I will consider only undivided—that is, random mating—populations. The results will provide a starting point for future investigations of subdivided populations. These simulations demonstrate that a high degree of pairwise relatedness is attributable to ancestors who lived much more recently than the most recent common ancestor of the entire population. Similarly, metrics of qualitative and quantitative pairwise genealogical overlap approach maximum possible values due to ancestors who lived considerably more recently than the generation of identical ancestors. Lastly, there is good evidence that the present results can be extrapolated to populations larger than those simulated here.

## 2. METHODS AND RESULTS

### 2.1. General Simulation Procedure

The basic simulation procedure is the same as previously [3], except that the population is undivided and reproduction is monogamous. To summarize, population size is constant and generations are non-overlapping. Each simulation begins at Generation 0 and proceeds forward for a predetermined number of generations. The only information recorded for individuals in subsequent generations is their Generation 0 ancestors. For most analyses, it was necessary only to have qualitative (0 or 1) information about ancestors. The exception was the analysis of quantitative genealogical overlap [4]. Sib mating was permitted and presumably occurred at the frequency expected by chance. Because reproduction is monogamous, shared ancestry must necessarily involve pairs of ancestors, or couples. For brevity and clarity, however, I will use phrases such as “at least one shared ancestor”, with the understanding that “ancestor” actually means “ancestor pair”. All pairwise genealogical comparisons were made by selecting 1000 random pairs of individuals from the population in each generation. The simulations that are shown in **Figure 2** involved sampling single (“focal”) individuals from the population, and comparing the Generation 0 ancestors of that individual to the ancestors of all other members of the population. For those simulations, the sample size of “focal” individuals was 10% of the population or 500, whichever was less. Unless otherwise noted, all results are based on 100 replicate simulations for each population size.

### 2.2. Time to Pairwise Most Recent Shared Ancestor

On average, the most recent shared ancestor of random pairs of individuals lived slightly more than  $0.5 \log_2 N$  generations previously (**Table 1**). There is some suggestion that the mean time to shared ances-

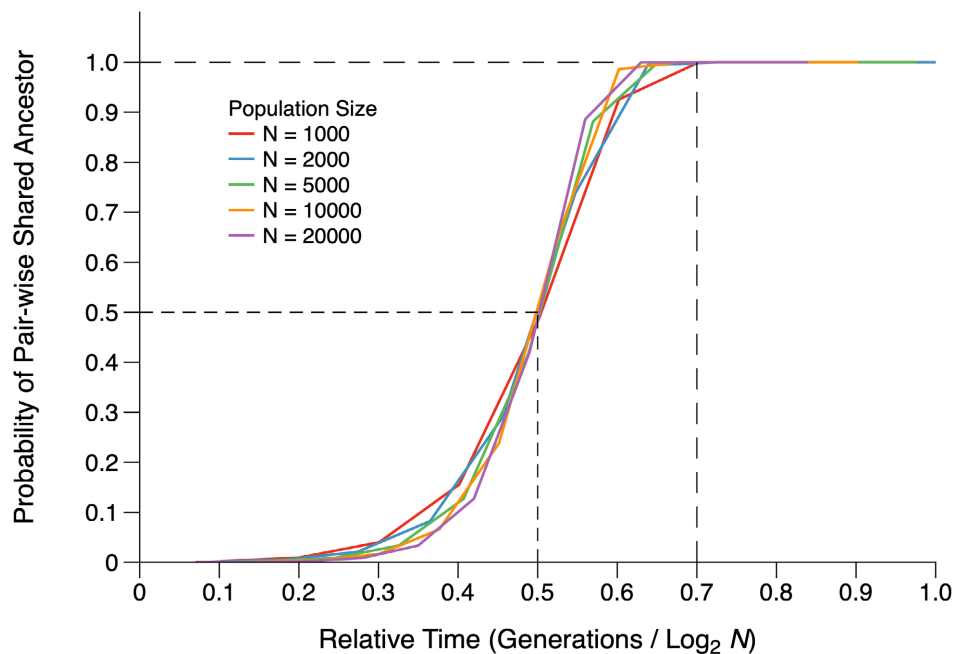
stry as a fraction of  $\log_2 N$  decreases with population size, although the effect is very small for the population sizes simulated.

### 2.3. The Probability of Shared Ancestry

The same set of simulations permits determination of the probability that random pairs of individuals will share at least one ancestor of specified degree. For all population sizes that were examined, the probability that two individuals will share an ancestor who lived  $0.5 \log_2 N$  generations ago, or more recently, is approximately 0.5 (Figure 1). For  $N = 20,000$ ,  $0.5 \log_2 N$  is about 7. Thus, the probability that a random pair of individuals drawn from a population of 20,000 will share a 7<sup>th</sup>-generation (or more recent) ancestor is about 50%. Seventh generation corresponds to fifth-great grandparent, and individuals who share such an ancestor are 6<sup>th</sup> cousins. The probability of shared ancestry increases very rapidly if additional generations are considered: for ancestors who lived  $0.6 \log_2 N$  generations in the past, or more recently, the probability of shared ancestry is greater than 90% for all population sizes that were simulated; and the probability of shared ancestry reaches 100% if we include ancestors who lived about  $0.7 \log_2 N$  generations previously. That is, there is universal pairwise shared ancestry: every individual in the population is related to every other individual by shared ancestors who lived  $0.7 \log_2 N$ , or fewer, generations in the past.

**Table 1.** Mean time to pairwise most recent shared ancestor.

Population Size, $N$	Generations	Generations/ $\log_2 N$
1000	5.38	0.540
2000	5.87	0.535
5000	6.53	0.531
10,000	7.03	0.529
20,000	7.52	0.526



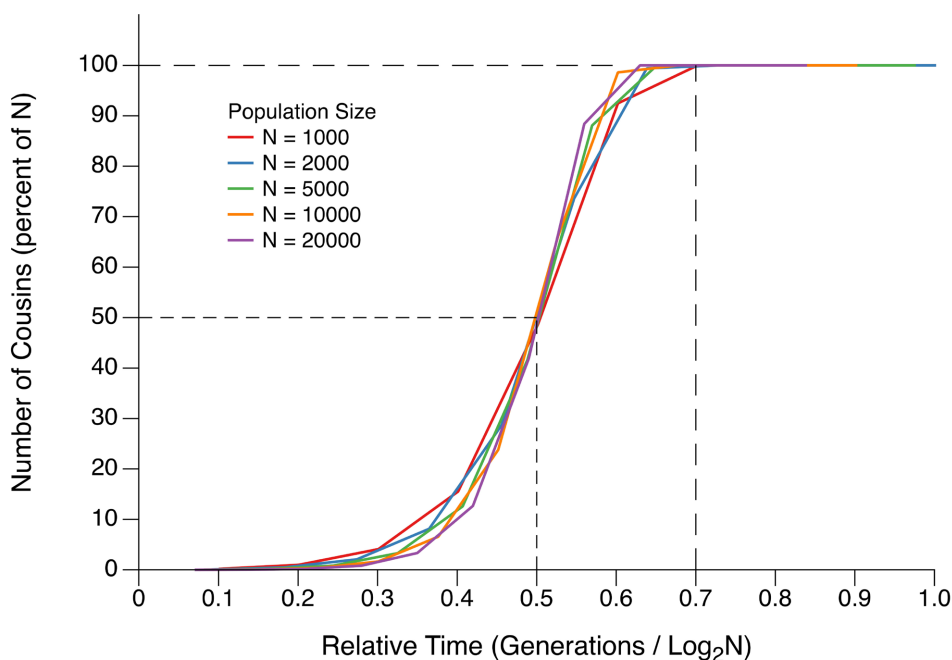
**Figure 1.** Probability of shared pairwise ancestry as a function of relative time to ancestor generation.

An important feature of **Figure 1** is that the time scale (x-axis) is not generations, but generations relative to  $\log_2 N$ . The fact that the curves lie on top of one another shows that the probability of pairwise shared ancestry scales uniformly with  $\log_2 N$ , albeit with some spread at the inflection points.

This analysis can be extended to consideration of shared ancestry in samples of  $S$  individuals. Consider a sample of  $S = 10$  from a population  $N = 20,000$ . From above, the probability that any pair of individuals do not share an ancestor who lived 7 generations ago (or more recently) is approximately 0.5. For  $S = 10$ , there are 45 pairwise comparisons. Thus, the probability that a sample of 10 individuals contains no 6<sup>th</sup>, or less distantly related, cousins is approximately  $0.5^{45}$ , or about  $2.8 \times 10^{-14}$  (assuming independence). On the other hand, for  $N = 20,000$ , the probability that all individuals in a sample of any size will be related to one another as 9<sup>th</sup> (or closer) cousins is very nearly 100%, as will be verified below.

## 2.4. Number of Relatives of Specified Degree

A random individual will be related to other individuals in the population by shared ancestors of various degrees. This idea can also be expressed as: what proportion of the population will be an individual's  $k^{\text{th}}$  degree, or closer, cousins? Results are shown in **Figure 2**. For example, a random individual will be related to about 50% of all other individuals in the population by shared ancestors who lived  $0.5 \log_2 N$  generations previously, or more recently. If  $N = 20,000$ , those relatives are 6<sup>th</sup> degree, or closer, cousins. **Figure 2** is almost identical to **Figure 1**, even though they depict the results of independent, and procedurally different sets of simulations. However, that is to be expected. For example, if an individual has a 0.5 probability of being related to another individual drawn randomly from the population (**Figure 1**), then we expect that the same individual will be related by similar degree to about 50% of the population (**Figure 2**). In short, the y-axis labels of the two figures are interchangeable. Considering ancestors who lived  $0.7 \log_2 N$  generations ago, or more recently, every individual in the population is related to every other individual (**Figure 2**), although not necessarily by the same ancestors—in other words, there is universal pairwise shared ancestry. Hence, the assertion in the previous section that all members of a sample of any size from a population of  $N = 20,000$  will be related to each other by ancestors who lived 10 or fewer generations ago ( $10 \approx 0.7 \log_2 20,000$ ).



**Figure 2.** Number of cousins as function of relative time to ancestor generation. (50 replicates for  $N = 20,000$ ).

## 2.5. Qualitative Genealogical Overlap

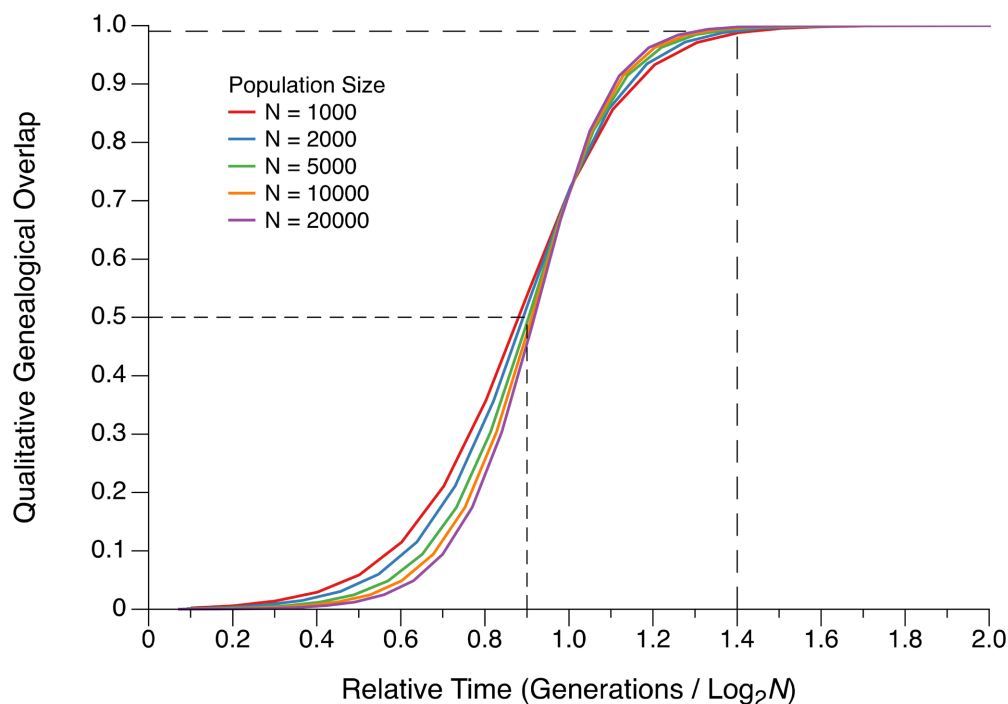
Pairwise qualitative genealogical overlap is the proportion of Generation 0 ancestors that are shared by pairs of individuals in subsequent generations. It is calculated as follows. Let individual  $A$  have  $N_A$  different ancestors from generation 0, and individual  $B$ ,  $N_B$  different ancestors. 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 - 1.0.

About  $0.9 \log_2 N$  generations are required for 0.5 overlap (Figure 3). This is somewhat less than the time required for a population-wide MRCA. An overlap of 0.5 means that pairs of individuals share half of their Generation 0 ancestors, on average. By about  $1.3 - 1.5 \log_2 N$  generations, qualitative pairwise overlap is  $>0.99$ : larger populations require less relative time. Qualitative overlap is necessarily 1.0 for all pairs when Generation 0 is an identical ancestry generation: the average time required for that is about  $2 \log_2 N$  generations.

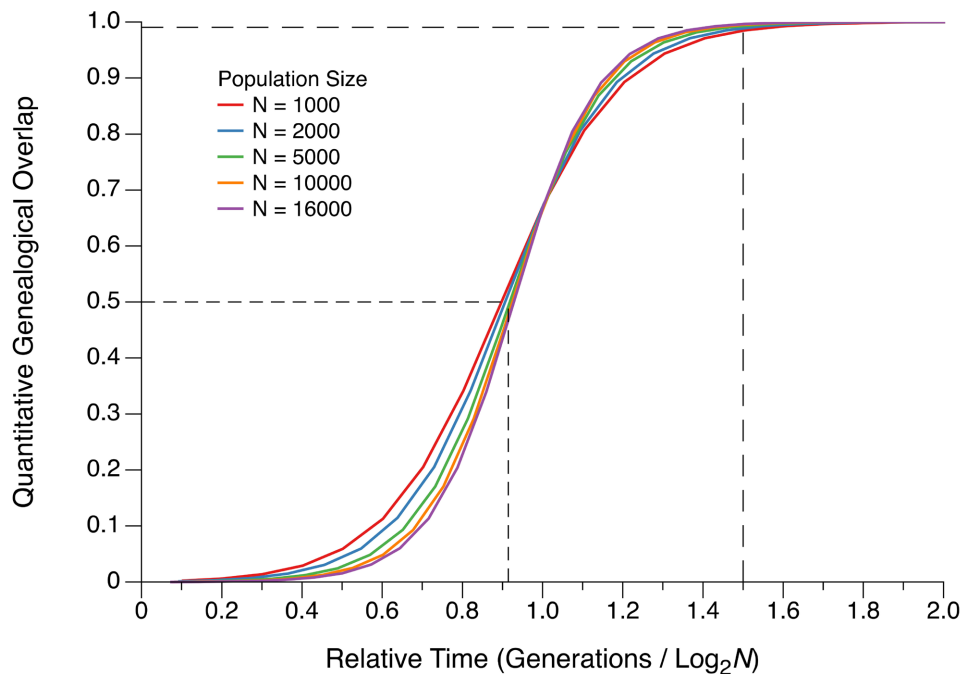
## 2.6. Quantitative Genealogical Overlap

Quantitative overlap is the similarity in the frequencies of shared ancestors in the genealogies of a pair of present-day individuals. I use the metric  $q^{(\alpha, \beta)}(G)$ , introduced by Derrida *et al.* [4], which is the overlap between the trees of individuals  $\alpha$  and  $\beta$  at generation  $G$  in the past. The range of values is 0 - 1. The largest population size simulated was 16,000, due to computer limitations.

Quantitative overlap (Figure 4) is almost indistinguishable from qualitative overlap (Figure 3). Overlap is about 0.5 for all population sizes for ancestors who lived slightly more than  $0.9 \log_2 N$  generations in the past. Considering ancestors who lived about  $1.4 - 1.6 \log_2 N$  generations previously, quantitative pairwise overlap is  $>0.99$ : larger populations require less relative time. Unlike the case of qualitative overlap, there is no requirement that quantitative overlap equal 1.0 once Generation 0 becomes an identical ancestry generation. But, in fact, quantitative overlap is  $>0.9999$  by the time that identical ancestry occurs.



**Figure 3.** Pairwise qualitative genealogical overlap as a function of relative time to ancestor generation.



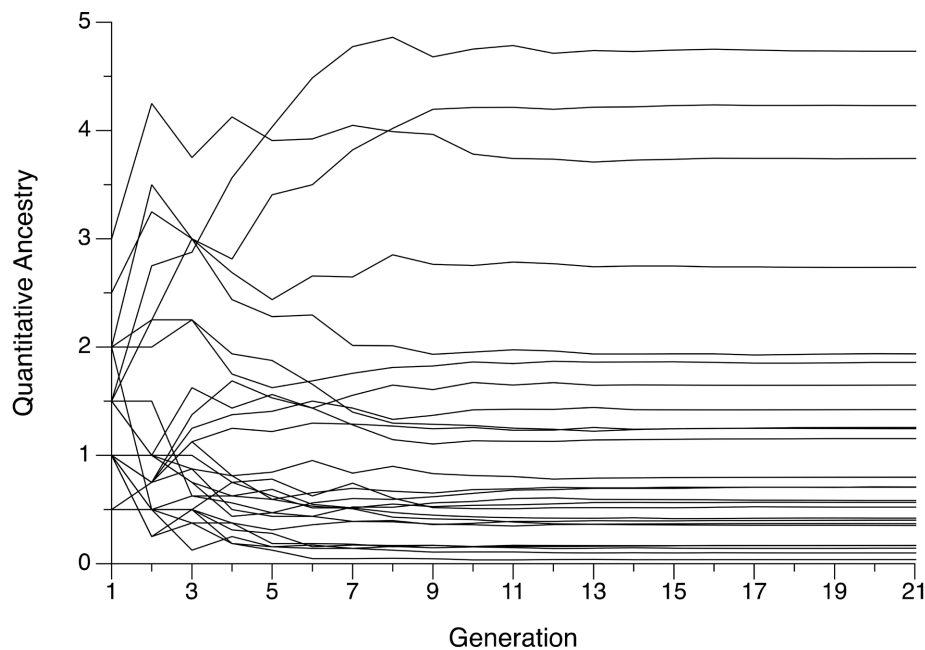
**Figure 4.** Pairwise quantitative genealogical overlap as a function of relative time to ancestor generation.

## 2.7. The Distribution of Ancestors in the Genealogies of Later Generations

In these simulations, reproductive success (number of offspring) is Poisson distributed with mean 2.0. Extinction of Generation 0 lineages is rapid. In fact, the mean time to extinction is about 1.55 generations (independent of  $N$ ), and the last extinction will occur by about  $0.67 \log_2 N$  generations [3]. Consequently, about 80% of the of the original Generation 0 cohort will become persistent ancestors of the population in future generations [1, 3, 4]. A correlate of indefinite persistence is that each Generation 0 member eventually comprises a nearly fixed proportion of the ancestry of future generations. Different Generation 0 members will have different representations in future genealogies, but the distribution of those representations becomes approximately stationary [4, 5].

The stationary distribution of the representation of Generation 0 members in the ancestry of future generations will be illustrated with an example from a single simulation with  $N = 16,000$ , and run for 30 generations. In this example, the last extinction of Generation 0 lineages occurred by Generation 8; 12,792 (79.95%) of the Generation 0 lineages persisted; the MRCA (12 Generation 0 individuals) appeared after 15 generations ( $=1.07 \log_2 N$  generations); and identical ancestry occurred after 27 generations ( $1.93 \log_2 N$  generations). A sample of 30 Generation 0 members was selected from this replicate for purposes of illustration. The lineages of 26 of those persisted, and all 26 became common ancestors of the population by Generation 22. Let  $C_{Gij}$  be the number of times Generation 0 member  $i$  occurs in the genealogy of individual  $j$  in Generation  $G$ .  $C_{Gij}$  is then summed over all individuals  $j$  in the population in Generation  $G$ , and scaled by dividing by  $2^G$ . The resulting metric,  $C_{Gi}$  is plotted for the sample of 26 persistent Generation 0 lineages from this replicate (Figure 5). Scaling by  $1/2^G$  means that the expected value of  $C_{Gi}$  is 1.0 for each Generation 0 member in each subsequent generation, and the sum of  $C_{Gi}$  over all  $N$  Generation 0 members equals  $N$  in every generation.

The representation of each Generation 0 member in the ancestries of subsequent generations is initially highly variable, but then stabilizes fairly quickly. By about  $\log_2 N$  (approximately 14) generations, the distribution is very nearly static (Figure 5), and remains essentially unchanged for all future generations. The eventual shape of this distribution can be roughly inferred from Figure 5. A more detailed description is provided in Derrida *et al.* [4].



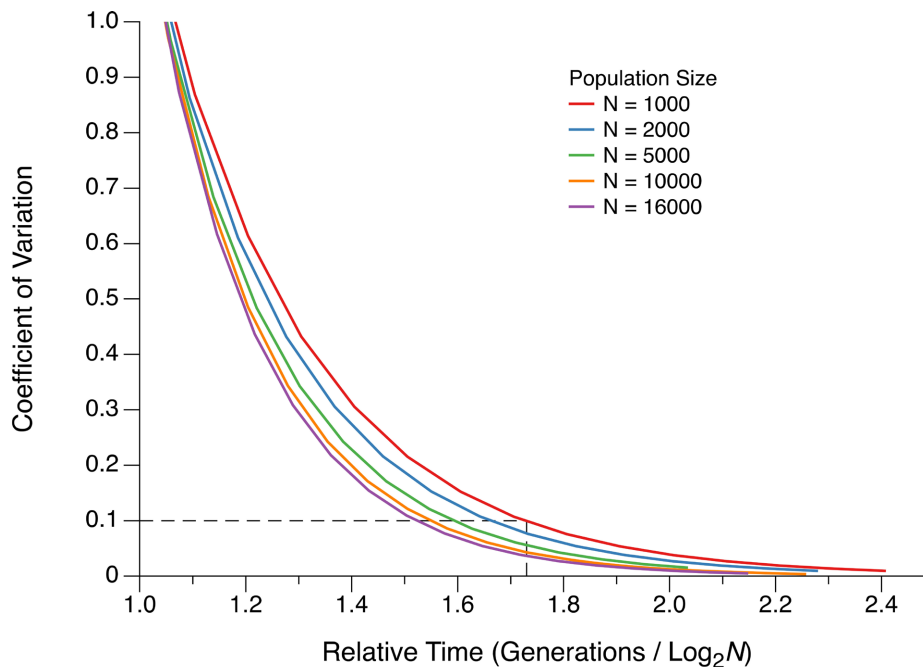
**Figure 5.** Quantitative contribution of a sample of 26 Generation 0 ancestors to genealogies of all members of the population.  $N = 16,000$ . Each line represents one Generation 0 ancestor. The vertical axis is  $C_{Gt}$ , as explained in the text.

## 2.8. The Coefficient of Variation of Quantitative Ancestry across All Individuals

The preceding section illustrates the fact that the quantitative representation of each “successful” Generation 0 ancestor becomes approximately fixed, across generations, when summed over all individuals in the population. High values of pairwise quantitative overlap (Figure 4) would also seem to indicate that a given Generation 0 ancestor has very nearly equal representation in the genealogies of every individual within a generation. In other words, we might expect the scaled variance in the occurrence of a given Generation 0 member in the ancestries of individuals in later generations to become smaller with time. An appropriate statistic is the coefficient of variation (CV), defined as standard deviation/mean.

Clearly, any “successful” Generation 0 member must become a common ancestor of the entire population before it can have equal representation in the genealogies of all individuals in a future generation. In the replicate simulation described in the preceding section, about 72% of the eventual common ancestors had become so by Generation 18, and almost 98% by Generation 21. But even before that, most individuals in the population will be descendants of most Generation 0 members whose lineages have not gone extinct. Consider that with stable population size and Poisson-distributed reproduction, successful reproducers will have about 2.3 offspring, on average. After 15 generations, an “average” Generation 0 member will have  $2.3^{15}$  ( $= 266,635$ ) “descendants”. If  $N = 16,000$ , we might expect that most of the population will be included among those descendants.

A sample of 10% or 200, whichever was greater, of Generation 0 ancestors was used for calculation of the coefficient of variation. The CV was obtained for each Generation 0 lineage from the variance of the scaled  $C_{Gij}$ , the mean of which was  $C_{Gt}$ , as described in the previous section; and the CV then averaged over all persistent lineages in the sample of Generation 0 ancestors. Within about  $1.7 \log_2 N$  generations or less, the CV declined to  $\leq 0.10$  (Figure 6). In fact, for  $N = 10,000$  or  $20,000$ , the CV was less than 0.05. In other words, the quantitative Generation 0 ancestry was very similar for all members of the population. That is consistent with pairwise quantitative overlap  $> 0.999$  by this time for the three larger population sizes (Figure 4); and is to be expected given that each persistent Generation 0 lineage eventually represents a temporally stable portion of the ancestry of the whole population (Figure 5).



**Figure 6.** Coefficient of variation of frequency of Generation 0 ancestors in genealogies of subsequent generations.

### 3. DISCUSSION

The principal finding of these simulations is that pairwise shared ancestry proceeds much more quickly than population-wide common ancestry. The MRCA of a population will have lived very nearly  $\log_2 N$  generations in the past. However, random pairs of individuals have about a 50% chance of sharing one or more ancestors who lived only half as long ago or more recently (Figure 1). Indeed, random pairs have a 100% chance of sharing ancestors who lived no longer than about  $0.7 \log_2 N$  generations previously. In other words, there is universal pairwise shared ancestry: every individual in the population is related to every other individual by shared ancestors who lived at most  $0.7 \log_2 N$  generations in the past. Put another way, there is universal “cousin-ness” of degree  $(0.7 \log_2 N) - 1$ . Similar conclusions apply to metrics of pairwise genealogical overlap (Figure 3, Figure 4). The most recent generation of population-wide identical ancestors will have lived about  $2 \log_2 N$  generations in the past [1, 3]. However genealogical overlap  $> 0.99$  is due to ancestors who lived only about  $1.4 - 1.5 \log_2 N$  generations previously.

To understand why shared ancestry increases much faster than population-wide common ancestry, an idealized example may help. In a constant-size population with biparental reproduction, and in which every pair has exactly two offspring, each individual will have  $4^k$  cousins of degree  $k$  [6]. Cousins of degree  $k$  share common ancestors who lived  $k + 1$  generations in the past. If  $k = 4$ , for example, each individual will have  $4^4 = 256$  fourth cousins due shared ancestors who lived five generations previously. Summing over third, second, and first cousins (and sibs), the expected number of relatives due to ancestors who lived five generations ago, or more recently, is 341 (Table 2). By comparison, each individual will have only  $2^5 = 32$  ancestors five generations in the past. This calculation ignores the effects of finite population size, which must slow down, and eventually stop, the growth in number of cousins as  $k$  increases. On the other hand, variation in reproductive success, such that some pairs have fewer than two and other pairs more than two offspring, will have the effect of increasing the number of cousins. That is because the ancestors of cousins will, on average, have had more than two offspring to make up for pairs who had none or one. In other words, an ancestor who lived  $G$  generations in the past, and who has any present-day descendants, can be expected to have more than  $2^G$  descendants, and those descendants will have more than



**Table 2.** Observed and expected number of cousins\*.

Shared Ancestor Generation	Cousin degree, $k$	$4^k$	Expected (degree $\leq k$ )	Observed	Observed/Expected
1	0 (sibs)	1	1	2	2.00
2	1	4	5	10	2.00
3	2	16	21	42	2.00
4	3	64	85	169	1.99
5	4	256	341	668	1.96
6	5	1024	1365	2537	1.86
7	6	4096	5461	8363	1.53
8	7	16,384	21,845	17,677	—
9	8	65,536	87,381	19,999	—

\* $N = 20,000$ . Data in this table is from the same simulations used for [Figure 2](#). Each entry for “Observed” is the mean of 100 replicates.

$4^{G-1}$  cousins of degree  $(G - 1)$ . This latter effect is illustrated by these simulations, as is the effect of finite population size ([Table 2](#)). In early generations, the observed number of  $k^{\text{th}}$ -degree or closer cousins was twice that expected by the above formula. By six generations, the excess began to diminish, reflecting the effect of finite population size. By eight generations, the expected number of cousins was greater than the population size.

Metrics of shared ancestry scale consistently with  $\log_2 N$  for the different population sizes simulated here. That is a good indication that the present results apply to larger populations. For example, given a population size of one billion, the MRCA will have lived about  $\log_2(1 \text{ billion}) \approx 30$  generations in the past, but there will be universal pairwise shared ancestry due to ancestors who lived only  $0.7 \log_2(1 \text{ billion}) \approx 21$  generations previously.

The results for quantitative pairwise overlap in genealogies ([Figure 4](#)) are very similar to those obtained by Derrida *et al.* [4]. The time required for Generation 0 population-wide ancestry to reach a stationary distribution ([Figure 5](#)), also appears to be consistent with Derrida *et al.* [4]. To my knowledge, results for the coefficient of variation in quantitative ancestry among individuals ([Figure 6](#)) have not been published before.

It should be noted that the present results apply to monogamous reproduction. If mating is promiscuous, progress toward pairwise shared ancestry ([Table 1](#) and [Figure 1](#) and [Figure 2](#)) is faster. However, in almost all cases, related individuals will be half-relatives (half-sibs or half-cousins). Mode of reproduction does not influence measures of genealogical overlap ([Figure 3](#), [Figure 4](#)).

Shared ancestry between pairs of individuals has received less attention than population-wide common ancestry [1-3], although a significant exception is the analysis of quantitative pairwise genealogical overlap [4, 7]. Clearly, overlap  $> 0$  implies shared ancestry. However, those analyses do not directly answer questions such as: what is the probability that a random pair of individuals share one or more ancestors who lived  $G$  generations in the past, or more recently ([Figure 1](#))? Shchur and Nielsen [8] derived the expectation for the number of individuals that would have no relatives of specified degree, say  $2^{\text{nd}}$  cousins, in a sample. Such information is important, for example, for genome-wide association studies. As such, Shchur and Nielsen address a different set of questions than this study. The present simulations could, however, be modified to estimate the same quantities. For example, after three generations of reproduction, draw a sample and determine how many members of the sample share no Generation 0 ancestors with any other member of the sample (*i.e.*, have no  $2^{\text{nd}}$  cousins in the sample).

The present simulations consider only random-mating, *i.e.*, unstructured, populations. Simulations to estimate MRCA and MRIA times in structured populations with migration have been carried out [3]. Under a wide range of assumptions about the number of subpopulations and migration rates, the time required to have an MRCA or MRIA generation is often less than twice, and seldom more than three times, that required for a panmictic population of the same total size. It remains to be seen whether similar scaling applies to measures of pairwise shared ancestry in structured populations.

## ACKNOWLEDGEMENTS

I thank Kiisa Nishikawa for helpful comments on earlier drafts of this paper. I also thank an anonymous reviewer for several very thoughtful suggestions which have been incorporated in the final version.

## 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. 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>
5. Service, P. (2017) Common Genealogical Ancestry. 2. Quantitative Ancestry. Unpublished Manuscript. [https://philservice.typepad.com/Genealogy/Common\\_Genealogical\\_Ancestry\\_Proportional\\_Ancestry.pdf](https://philservice.typepad.com/Genealogy/Common_Genealogical_Ancestry_Proportional_Ancestry.pdf)
6. Butler, R.E. (2011) Number of Distant Cousins. <https://members.storm.ca/~rebutler/pdffiles/NumberofDistantCousins.pdf>
7. Derrida, B., Manrubia, S.C. and Zanette, D.H. (1999) Statistical Properties of Genealogical Trees. *Physical Review Letters*, **82**, 1987-1990. <https://doi.org/10.1103/PhysRevLett.82.1987>
8. Shchur, V. and Nielsen, R. (2018) On the Number of Siblings and p-th Cousins in a Large Population Sample. *Journal of Mathematical Biology*, **77**, 1279-1298. <https://doi.org/10.1007/s00285-018-1252-8>