The Mathematics of Pedigree Collapse: Why Two People Could Not Have Created the Human Race

Abstract

This document explores the mathematical paradox that arises when tracing human ancestry backward in time under the naive assumption of a perfect binary tree. While a simple exponential model suggests an impossibly large number of ancestors in the recent past, population genetics resolves this via *pedigree collapse*—the overlapping of family trees due to intermarriage among distant relatives. We derive mathematical models for the expected number of distinct ancestors and demonstrate why the human population never originated from a single pair of individuals.

Contents

1 Introduction

If each person has two parents, four grandparents, and so on, then going back n generations yields 2^n ancestors in the naive model. For $n = 30$ (roughly 900 years), this gives over one billion ancestors—far exceeding the total world population at that time

($\approx$ 300–500 million). This contradiction shows that the simple binary tree model must be wrong. The correction lies in **pedigree collapse**: ancestors are not all distinct because distant cousins marry, causing the same individual to appear multiple times in a family tree.

2 The Naive Exponential Model

Let $A(n)$ be the number of **distinct** ancestors n generations before a given individual. If all ancestors are unique, then:

$$A(n) = 2^n \tag{1}$$

For example:

- $n = 1$ (parents): $A(1) = 2$
- $n = 2$ (grandparents): $A(2) = 4$
- $n = 3$ (great-grandparents): $A(3) = 8$
- $n = 30$: $A(30) = 2^{30} = 1{,}073{,}741{,}824$

Let P_n be the total human population n generations ago. For $n = 30$, $P_{30} \approx 3 \times 10^8$ to 5×10^8. Clearly,

$$2^n \gg P_n \quad \text{for large } n, \tag{2}$$

which is impossible because $A(n)$ cannot exceed P_n. Therefore, the naive model fails.

3 Pedigree Collapse: The Basic Bound

Pedigree collapse means that the same person occupies multiple positions in the family tree. Hence:

$$A(n) \leq \min\left(2^n,\ P_n\right) \tag{3}$$

For large n, the bound is controlled by P_n, not 2^n. As n increases further, $A(n)$ asymptotically approaches a fraction of the ancestral population size.

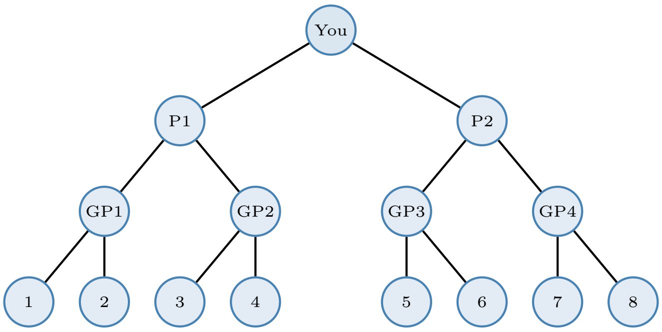

Figure 1: **Figure 1: Naive Binary Tree.** All ancestors are distinct. At generation 3, there are $2^3 = 8$ distinct ancestors (numbered 1–8). Extrapolating backward gives $A(n) = 2^n$ ancestors, which quickly exceeds the total human population.

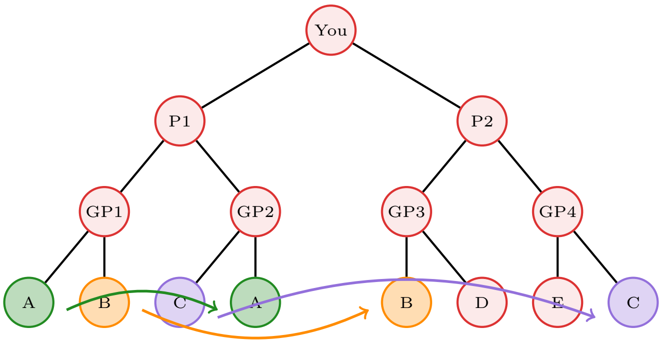

Figure 2: **Figure 2: Collapsed Pedigree.** Due to intermarriage, ancestors A, B, and C appear twice (connected by colored arrows). Total distinct ancestors = 5 (A, B, C, D, E), not 8. Legend: Green A (repeat), Orange B (repeat), Purple C (repeat), D and E (unique).

4 A Random-Mating Model (Constant Population Size)

We now introduce a mathematical model from population genetics. Assume:

- A constant population of N breeding adults each generation.
- Random mating (no selection, no population structure).
- Discrete, non-overlapping generations.

4.1 Expected Number of Distinct Ancestors

From coalescent theory, the expected number of distinct haploid ancestors n generations ago is:

$$\mathbb{E}[A_{\text{haploid}}(n)] = 2N\left(1 - e^{-n/(2N)}\right). \tag{4}$$

4.2 Asymptotic Behavior

From Eq. (4):

- When $n \ll 2N$: $\mathbb{E}[A_{\text{haploid}}(n)] \approx n$ (matches naive model initially).
- When $n \gg 2N$: $\mathbb{E}[A_{\text{haploid}}(n)] \to 2N$ (saturates at population size).

Thus, exponential growth stops once n exceeds $2N$.

5 Numerical Example

Let $N = 10{,}000$ (breeding adults), so $2N = 20{,}000$ haploid lineages.

n (generations)	Approx. years	$\mathbb{E}[A_{\text{haploid}}(n)]$
5	150 (30 yrs/gen)	4.99
10	300 (30 yrs/gen)	9.97
50	1500 (30 yrs/gen)	49.4
100	3000 (30 yrs/gen)	97.6
500	15,000 (30 yrs/gen)	393.5
1000	30,000 (30 yrs/gen)	787
5000	150,000 (30 yrs/gen)	1836

Table 1: Expected number of distinct haploid ancestors for $N = 10,000$ adults.

Observe that even after 500 generations (15,000 years), the number of distinct ancestors is only about 394, not $2^{500} \approx 10^{150}$. The pedigree collapse is extreme.

6 Why Two People Could Not Have Created Humanity

If the entire human population descended from a single pair, then $N = 2$ and $2N = 4$. Equation (4) would give:

$$\mathbb{E}[A_{\text{haploid}}(n)] = 4\left(1 - e^{-n/4}\right). \tag{5}$$

For large n, this approaches 4. That would mean all genetic variation in modern humans comes from just 4 haploid lineages (i.e., two diploid individuals).

However, observed human genetic diversity requires an ancestral population size of at least several thousand. The expected heterozygosity $\mathbb{E}[H]$ in a population of size N is:

$$\mathbb{E}[H] = \frac{4N\mu}{1 + 4N\mu}, \tag{6}$$

where $\mu \approx 2.5 \times 10^{-8}$ is the mutation rate. Solving for observed heterozygosity ≈ 0.001 gives $N \approx 10,000$. Hence, a single ancestral couple is mathematically impossible.

7 Population Bottlenecks and Neanderthal Admixture

Population bottlenecks intensify pedigree collapse. The **Toba supereruption** ($\approx$ 74,000 years ago) has been hypothesized to have caused a human bottleneck, though evidence remains debated [1]. For a bottleneck to N_{small} individuals, the number of distinct ancestors saturates at $2N_{\text{small}}$.

Neanderthal admixture: Non-African populations carry $\approx$ 1.5%–2.1% Neanderthal-derived DNA, indicating interbreeding $\approx$ 50,000–60,000 years ago [2]. This admixture required a Neanderthal population of thousands, not a single couple.

8 Effective vs. Census Population Size

The effective population size N_e is typically much smaller than census size N_c. For humans, $N_e \approx$ 10,000–30,000 while $N_c \approx$ 7 billion today. All equations in this document use N_e.

9 Conclusion

The naive model $A(n) = 2^n$ fails due to pedigree collapse. The correct model is:

$$\mathbb{E}[A(n)] = 2N\left(1 - e^{-n/(2N)}\right), \tag{7}$$

which saturates at $2N$. Genetic data require $N \gg 2$, disproving a single-couple origin.

Glossary of Terms

- **MRCA (Most Recent Common Ancestor)**: The most recent individual from whom all living humans are descended along some lineage. For all humans, the MRCA lived $\tilde{2}$,000–5,000 years ago.
- **IAP (Identical Ancestors Point)**: The time in the past when every person in the ancestral population who has any living descendants is an ancestor of *every* present-day individual. For humans, the IAP is estimated around 10,000–20,000 years ago.
- **Coalescent Theory**: A retrospective mathematical model in population genetics that traces lineages backward in time to their common ancestors. The expected time for two lineages to coalesce is $2N$ generations.
- **Effective Population Size (N_e)**: The size of an idealized Wright-Fisher population that would lose genetic diversity at the same rate as the actual population. For humans, $N_e \approx$ 10,000–30,000, much smaller than the census size.
- **Pedigree Collapse**: The phenomenon where the same ancestor appears multiple times in a family tree due to intermarriage among relatives. This causes the number of distinct ancestors to grow much more slowly than 2^n.
- **Heterozygosity**: The proportion of gene loci that are heterozygous (having two different alleles). Expected heterozygosity $\mathbb{E}[H] = \frac{4N\mu}{1+4N\mu}$, where μ is the mutation rate.
- **Population Bottleneck**: A sharp reduction in population size, which intensifies pedigree collapse and reduces genetic diversity.

References

[1] Ambrose, S. H. (1998). *Late Pleistocene human population bottlenecks, volcanic winter, and differentiation of modern humans.* Journal of Human Evolution, 34(6), 623-651. DOI: 10.1006/jhev.1998.0219

[2] Prüfer, K., et al. (2014). *The complete genome sequence of a Neanderthal from the Altai Mountains.* Nature, 505(7481), 43-49. DOI: 10.1038/nature12886

[3] Chang, J. T. (1999). *Recent common ancestors of all present-day individuals.* Advances in Applied Probability, 31(4), 1002-1026.

[4] Rohde, D. L., Olson, S., & Chang, J. T. (2004). *Modelling the recent common ancestry of all living humans.* Nature, 431(7008), 562-566.

[5] Wakeley, J. (2008). *Coalescent Theory: An Introduction.* Roberts & Company Publishers.