

Coalescence time distributions for hypothesis testing

-Kapil Rajaraman (rajaramn@uiuc.edu)
498BIN, HW# 2

This essay will be an overview of Maryellen Ruvolo's work on studying modern human origins using multiple datasets. My interest in the subject was aroused when reading Roger Lewin's book "Patterns of Evolution", which referred me to the Feb 1996 issue of Molecular Phylogenetics and Evolution, a special issue to celebrate Morris Goodman's 70th birthday. (The issue is free from biological technicalities for the most part, and is a good place to see what the issues in primate-hominoid evolution are).

Evolutionary hypotheses

There seems to be no doubt that the earliest hominids arose in Africa. The first hominid to leave Africa was *H. erectus*, and it is universally accepted that this happened around 1.8 my (million years) ago. However, it was still unanswered (at the time of the article) if this time was the last time all living humans were connected by a common ancestor. One theory known as the "multiregional hypothesis" says that this is true, implying that human population differences are very old. Also, it is held that the transition from *H. erectus* to *H. sapiens* was made separately in different parts of the world, and cohesiveness was maintained by gene flow. A second "candelabra" theory does not hypothesize gene flow.

Another theory – known as the "Out of Africa" or "Noah's Ark" model – states that some time after *H. erectus* spread through the world, another speciation occurred, and *H. sapiens* emerged in Africa. These humans then replaced the existing *H. erectus* as they moved outwards. This theory posits younger population differences. Molecular evidence from mtDNA has put an approximate date of 250,000 years for the second exodus.

Molecular data studies

As we follow a history of genetic lineages back in time, these lineages join up into a common ancestral type called the coalescent; the age of this ancestor is called the coalescent time. However, to reconstruct phylogenies from genetic data, there should be a way of relating the gene tree to the population tree.

Daughter populations take some samples of alleles from the original population on division, and so the coalescence times of each population's alleles are usually greater than or equal to the time of population divergence. However, in some cases, allelic loss can lead to reduced coalescence time observations, as seen in Fig. 1. Moreover, the correct alleles have to be used to make observations. For example, the HLA (human leukocyte antigen) tree gives a coalescence time of 35 my. This is because the alleles under study have been preserved from the times of *H. habilis* and the australopithecines. The problem, however, is that it cannot be said whether any allele under study is an anciently polymorphic (i.e. preserved) one or not, making results questionable.

Coalescence time distributions

The steps in this experiment are to (1) collect datasets from many genes; (2) to estimate coalescence times for each dataset; (3) to plot the frequency distribution of times; and (4) to compare the observed distribution with model predictions. These predictions are shown in Fig. 2. The difference in shapes of the multi-regional and candelabra models is due to the hypothesized absence of gene flow in the candelabra model. This absence of gene flow reduces the effective population size, which in turn leads to increased allelic loss. This reduces the observed coalescence time. Even without allelic loss, gene flow maintains older coalescent times, as shown in Fig. 3.

The distributions shown in the figure are very general, and a more rigorous information set would need to incorporate frequency distribution of mutation rates of all the genes in the human genome and the proportion of genes that are anciently polymorphic. But, even at this qualitative level, the experimental predictions should be able to distinguish the three models.

Prevailing tests utilize the branch length methods, which use the amount of genetic distance accumulated along the branches of a gene tree to infer divergence dates (after calibrating the molecular clock). The advantage of the coalescence time methods is that error bars associated with the evolutionary process can be estimated. Also, as the size of the datasets increases, the coalescent times approach the dates set by the relative branch lengths.

Applications to existing data

In the article, Ruvolo describes all the experiments that have been performed to collect molecular data. Genetic evidence from protein polymorphisms, mitochondrial DNA, Y-chromosomal sequences, and other tests give recent dates of approximately 200,000 years for human genetic ancestors. Some tests give older dates of 1.3 my, 3 my, and 500,000 years. With this evidence, the out-of-Africa model seems to be the likeliest for modern human origins. (Further refinements to the experiments have been suggested to make sure that the loci which are studied are unlinked. With these refinements, it is expected that the out-of-Africa model would be made even likelier).

Discussion

The out-of-Africa model has been upheld by almost all molecular tests, but some researchers – mainly Wolpoff – claim that the multiregional model is correct. This group feels that the molecular clock used to calibrate mutation accumulation time is faulty. When it was pointed out that the calibration would have to be off by a factor by 5 or so, the proponents of the multi-regional model claim that this does not matter, because the natural loss of genetic material prevents accurate reconstruction of phylogenetic histories. This view, however, does not seem to be supported by the larger part of the anthropological community. Population wave data studies have shown that at the time of

population divergences and expansions, there were too few humans in existence to be compatible with the multiregional theory!

Current research

A number of researchers have been using coalescent time distributions for phylogenetic studies since, and there has been considerable refining of the technique to account for variable population sizes, colonization, etc. I did not find any work specifically relating to the out-of-Africa versus multiregional debate, but a recent article has claimed that the Homo separated from the chimpanzees 10 my ago, a factor of 2 over the more universally accepted date of 5 my ago. This goes to show that the issue of primate phylogeny is still not conclusively resolved, and we may expect to see more than a few debates in the near future.

References

1. Roger Lewin , *Patterns in Evolution – The new molecular view*, 1996.
2. Maryellen Ruvolo, *A New Approach to Studying Modern Human Origins: Hypotheses testing with Coalescence Time Distributions* and references therein, *Molecular Phylogenetics and Evolution*, **5** (Feb 1996), pp. 202-219
3. *Molecular Phylogenetics and Evolution*, Feb 1996 issue

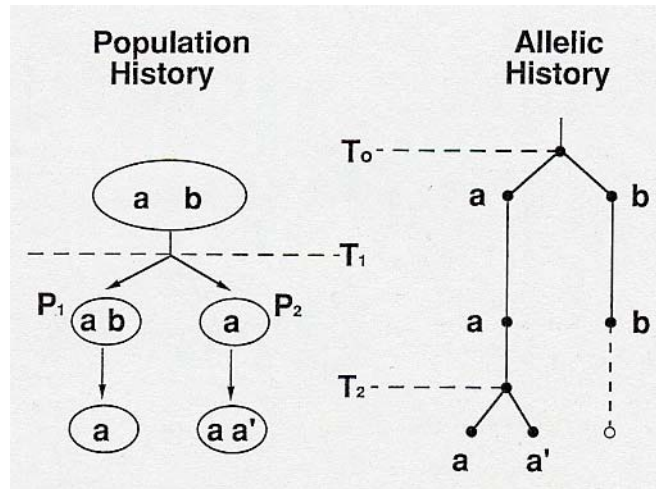


Fig. 1

Scenario showing how a molecular coalescent can occur after a population divergence.

(Left) Population history shows two alleles (a,b) in an ancestrally polymorphic population. Population divergence occurs at time T_1 , and only one population P_1 receives both alleles during population formation. Subsequently, allele b is lost from P_1 . A new allele a' arises in P_2 after divergence.

(Right) Allelic history. If no alleles had been lost, the coalescence times of all alleles (a,a',b) would have been T_0 , which is greater than T_1 . Instead, observed alleles (a,a') have coalescence time T_2 , which is less than T_1 , the time of population divergence.

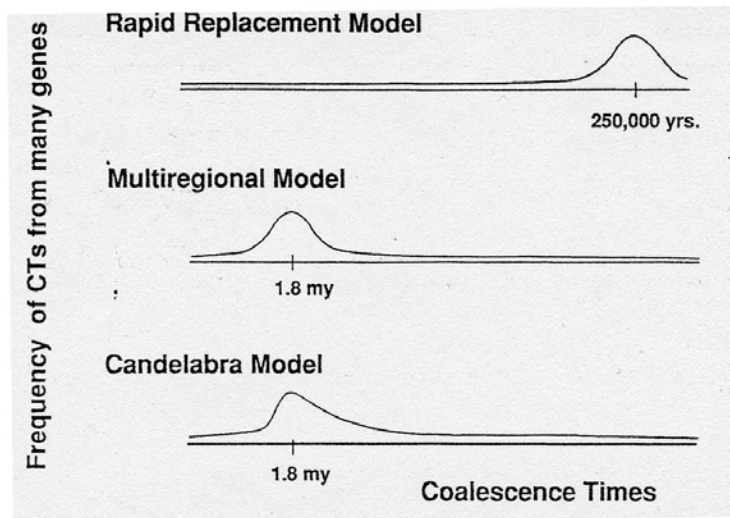


Fig. 2

Qualitatively predicted coalescence time frequency distributions

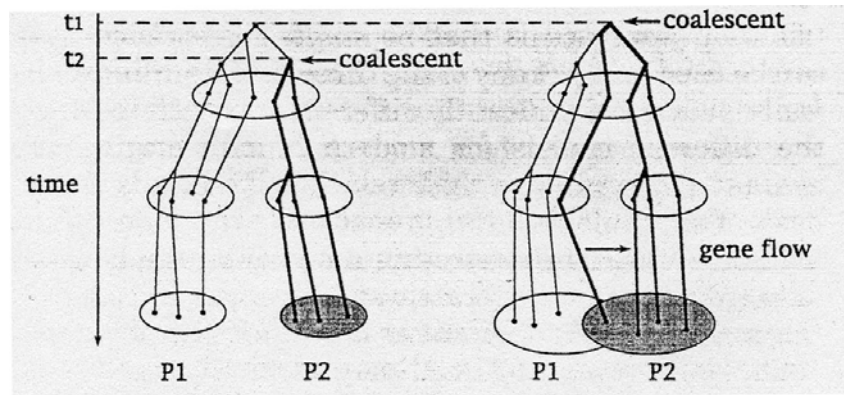


Fig. 3
Gene flow acts to maintain older coalescence times.

In the absence of gene flow (left), the alleles in Population P2 have a coalescent at t_2 .
With gene flow (right), the coalescent is older, at t_1 .