

The Two Phases of the Coalescent and Fixation Processes

Introduction

The coalescent process which traces back the current population to a common ancestor and the fixation process which follows an individual until the population is fixed for its descendants are heuristically inverse processes, yet the time reversal of one is seldom the other. This is because several generations will share the same most recent common ancestor, and several generations will first achieve fixation for one of their genes in the same generation. If the original individual is the most recent common ancestor of the present generation, and the present generation is the population in which the original individual becomes fixed, then the processes are inverses of each other. In general, however, if a gene is followed to fixation, the most recent common ancestor of the generation in which it becomes fixed will be more recent than the original gene. Similarly, if the present generation is traced back to its most recent common ancestor, that gene will have become fixed prior to the present generation. But a fixation/coalescence inverse process from most recent common ancestor to first generation of fixation (or its reverse) will be a subset of any fixation or coalescent process.

The present work considers this aspect of the structure of the coalescent and fixation processes. The inverse process shall be referred to as the transition phase, since it manifests the actual increase from a single copy to the entire population (or contraction from the entire population to a single individual). The several generations of a coalescent process which share the same most recent common ancestor, and the several generations of a fixation process which attain fixation in the same generation shall be called the stasis phase. Because the expected fixation and coalescent times are equal, and those processes share the transition phase, the expected lengths of the stasis phase are the same for the coalescent and fixation processes.

Notation

The previous concepts can be elucidated by introducing appropriate notation. Start at some generation t , and let T_i be the first generation that the population is fixed for some gene in generation t , then the expected fixation time is the expected value of $T_i - t$. Next let t_i be the generation of the most recent common ancestor of the population in generation T_i , then the expected length of the transition phase will be the expected value of $T_i - t_i$. T_{i+1} (T_{i-1}) can be defined as the next (previous) generation when the population first became fixed for a different most recent ancestor, and t_{i+1} (t_{i-1}) the generations of the respective most recent common ancestors. Then all the generations from T_{i-1} to $T_i - 1$ share the same most recent common ancestor (in generation t_{i-1}), and all the generations from $t_i + 1$ to t_{i+1} first attain fixation for one of their genes in the same generation (T_{i+1}). The same notation could have been defined starting at an arbitrary generation, and going back to the generation of its most recent common ancestor rather than forward to its fixation.

The intervals T_{i-1} to $T_i - 1$ and $t_i + 1$ to t_{i+1} , which I shall denote as ΔT and Δt , contain the stasis phases for coalescence and fixation, respectively. Hence I shall call them stasis intervals. Note the usage of “phase” and “interval”: the stasis phase is the realized stasis period, it is a stasis interval truncated at the initial (or present) generation. Because the initial generation can be anywhere in the stasis interval, the average fixation (coalescent) time should be half the expected value of Δt (ΔT) (weighted by the lengths of the intervals) added to the expected value of the transition phase ($E[T_i - t_i]$). Hence the expected fixation time (which is equal to the expected coalescent time) is $\frac{1}{2}E[(\Delta t)^2]/E[\Delta t] + E[T_i - t_i]$.

The adjacent figure illustrates these definitions for a simulation of a haploid population of six individuals. The gene first becomes fixed in generation T_1 , for which generation the most recent common ancestor is in generation t_1 . Generations t_1 to T_1 and t_2 to T_2 are transition phases; generations $t_1 + 1$ to t_2 are a stasis interval for fixation; and generations T_1 to $T_2 - 1$ are a stasis interval for coalescence. The original generation t would have occurred somewhere in a stasis interval for fixation.

Characterization of the Stasis Phases

The transition phase is the actual increase of a gene from a single copy to the entire population for fixation, and the reverse for coalescence. The length of the transition phase is the difference between the generation in which the ancestral gene becomes fixed, and the generation of the most recent common ancestor of that population.

The stasis phase of fixation heuristically has the ancestral gene as a single copy before it branches to spread to the population; actually the gene may branch and have several copies during that phase, and the branches may persist during part of the transition phase, but those branched lineages will die out before fixation occurs. From the coalescent perspective of going back in time, the fixation stasis phase precedes the most recent common ancestor. The length of the stasis phase is the difference between the initial generation and the most recent common ancestor of the population in which the original gene became fixed.

The stasis phase of coalescence is generation(s) when the entire population shares the same most recent common ancestor; the transition phase (which precedes the stasis phase in real time) begins (going backward in time) when the population contains an individual not descended from that ancestor (i.e., there is a more ancient branch in the pedigree). From the fixation perspective of going forward in time, the coalescence stasis phase is the generations after fixation for the most recent common ancestor of the population until the present generation. The length of the coalescence stasis phase is the difference between the present and the first generation in which the population has the specified most recent common ancestor.

Note that the stasis phase for a given coalescent (or fixation) process will be a subset of a stasis interval, which includes all generations sharing the same most recent common ancestor (or generation of first fixation), including generations after the present generation (or before the initial generation).

Extreme Examples

If every member of the population replaces itself for $n - 1$ generations, and then one individual parents the entire next generation, the length of the transition phase will be 1, and the length of the stasis interval (ΔT or Δt) will be $n - 1$. The average fixation/coalescence time will be $(n + 1)/2$ generations. If the length of the stasis phase were a random variable X , the weighted expected value would need to be calculated as noted above.

A stasis phase of length 0 is obtained if the member of the ancestral lineage (individuals whose descendants will not go extinct) produces two progeny every generation, every other individual produces one progeny, except that the individual whose ancestors left the ancestral lineage furthest in the past does not reproduce. This follows since every generation will contain a most recent common ancestor for some future generation, and every generation will be the first generation of fixation for some previous generation. If the population has N individuals, then an ancestral gene will become fixed in $N - 1$ generations; that will be the transition time, fixation time, and coalescent time.

Poisson Progeny Distribution

The binomial or Poisson progeny distribution is employed with the assumption that the future depends only the present, and not previous generations. In particular, at the time of a fixation event (T_i), the time until the next fixation event ($T_{i+1} - T_i$) will be less than or equal to the expected fixation time, because at time T_i there may be multiple copies of the next gene destined for fixation. Therefore, the average length of the stasis interval will be less than the expected fixation time; however, this refers to the unweighted average of the length of the stasis interval.

Numerical simulations were performed for 1000 fixations each in haploid populations of 100 and 200 individuals. The average times until fixation were 199 and 390 generations, respectively, which are approximately equal to $2N$. At fixation, the average times since the most recent common ancestor were 97 and 195 generations, respectively. Hence the average length of the transition phase was half of the fixation time, and the weighted average of the stasis intervals was equal to the average fixation time.

Discussion

This study was motivated by the need to clarify the relation between coalescent and fixation events. Indeed, the expected coalescent and fixation times are equal, but the expected time since a common ancestor at the generation when fixation occurs is not the same as the expected coalescent time in general, nor is the expected time until fixation of a gene which is a most recent common ancestor equal to the expected fixation time in general. When studying the fixation of a gene or the coalescence of a population, the actual transition from a single copy to the entire population or from an entire population to the single copy will be less than the fixation or coalescent time.

One implication of these results is that hitchhiking occurs in half of the fixation time (for random mating with Poisson progeny distribution) because it is only during the transition phase that crossing over could affect monomorphism at a linked locus. Of course, this does not address the role of mutation in polymorphism.

In fact, these results are really not important to the general questions of polymorphism and evolution. Dead end lineages contribute to the variation of a population. The breadth of the coalescent process as well as the coalescent time impacts how much mutation (which provides variation) occurs during fixation. The minimal genetic history of a population is the lineage of single genes which eventually become fixed, for such a lineage there is no concept of variation or coalescence.

The concise genetic history of a population is the lineage of the single genes in each generation which are destined for fixation. The genetic diversity which we study is the embellishment of that lineage. The present work provides another perspective on the nature of this embellishment.

