

# The Coalescent of a Selected Allele

R. B. Campbell

Department of Mathematics  
University of Northern Iowa  
Cedar Falls IA 50614-0506

[campbell@math.uni.edu](mailto:campbell@math.uni.edu)  
[www.math.uni.edu/~campbell](http://www.math.uni.edu/~campbell)

(319) 273-2447

June 2003      Chico, CA

## The Coalescent

The coalescent in its most general sense refers to all aspects of the ancestral pedigree of the population. Two aspects of this are the time since a common ancestor (which is the same as the time until fixation at equilibrium) and the shape (reduced family tree). We produce the approximation  $(.5 + \ln(Ns))/s$  for the expected fixation time of a selected mutation and show that the coalescent for a selected mutation is less concave than the coalescent for a neutral mutation.

The coalescent was originally studied for neutral mutations (Kingman 1982). Kaplan, Dardeen, and Hudson (1988) used selection to construct a virtually subdivided population where the coalescent could be studied. Neuhauser and Krone (1997) demonstrated that selection affected the time scale, but not the structure, of the coalescent process under certain assumptions. But in general, the coalescent in the presence of selection is not yet understood.

It is important to recognize that as a selected mutation becomes fixed, the lineages of many copies of the mutation will die off, hence the coalescent is a subset of all the copies of the mutation which occur. In fact, the coalescent grows much slower than the allele frequency during fixation; one model for neutral mutations has the size of the coalescent increase as  $1/(1 + (1/N) - (t/2N))$  while the allele frequency increases linearly as  $t/2$  where  $N$  is the population size and  $t$  is the number of generations since the common ancestor. This provides that there are only about two individuals in the coalescent at  $t = N$  ( $2N$  is the time until fixation), although there are approximately  $N/2$  copies of the mutant allele in the population at  $t = N$ .

## Analysis

Our analysis is predicated on the observation that drift governs the dynamics of a selected allele when it is rare, but selection governs its dynamics when it is frequent.

The demarcation point between the drift and selection growth phases is rather arbitrary, we shall use the time when the number of copies of the mutant allele is equal to  $1/s$ , which will be approximately  $1/(2s)$  generations after the mutation first occurs (if drift governs the dynamics of allele frequencies to that time). If the mutation were neutral (i.e.,  $s$  were just a parameter), there would be approximately  $4Ns/(4Ns - 1)$  members of the coalescent at this time (which is close to 1 if  $4Ns \gg 1$ ). However, under selection, the probability of fixation of an allele is approximately  $2s$ , hence  $2s$  of the alleles should be immortal (have nonterminating lineages), hence be in the coalescent. Since  $2s \times 1/s = 2$ , this will increase the size of the coalescent to 2 or 3, which is still rather small.

## Coalescent Time

The coalescent time can be calculated as the time to reach frequency  $1/s$  plus the time to go from frequency  $1/s$  to  $N$ . For a neutral allele, the time to reach  $1/s$  is  $1/(2s)$  by the branching process approximation (the probability of surviving  $t$  generations is  $1/(2t)$ , hence the expected population size conditioned on not being extinct is  $2t$ ). The diffusion approximation will extend the time because the variance is less under the diffusion approximation than in a branching process, but if  $1/s \ll N$  this should not be significant. Furthermore, selection will shorten the time it takes an allele to reach a specified frequency. Using the expected value of the population size also overestimates the time until the population size is reached, populations with a higher allele frequency are more likely to fix that allele (and populations with a lower frequency more likely to have it go extinct), so conditioned on fixation the time to reach  $1/s$  will be less (when the high frequency populations first attained the frequency  $1/s$ ). A deterministic calculation for the time to reach frequency  $1/s$  (assuming selective advantage  $s$ ) yields  $t = -\ln s/s$  which is greater than  $1/(2s)$  if  $s < .06$ , hence drift overwhelms the deterministic change. The deterministic time to increase from  $1/s$  to  $N$  is  $t = \ln(Ns)/s$ , hence the total time until fixation

is approximately  $(.5 + \ln(Ns))/s$ . If  $Ns < 1$ ,  $1/s > N$  and the second summand is not used because fixation occurs during the drift phase (this fixation time is off by a factor of 4 if  $Ns = 1$  (i.e.,  $.5(1/s) = .5N \neq 2N$ ), which is the discrepancy between the diffusion approximation and the branching model). If  $Ns \gg 1$ , the time is approximately  $\ln(Ns)/s$ , which is less than  $2N$  which neutrality would provide, and less than  $\ln N/s$  which selection would provide. The fixation time calculated from the diffusion approximation by van Herwaarden and van der Wal (2002) is essentially  $\ln(Ns)/s$ .

### Coalescent Shape

We have demonstrated the intuitive result that the coalescent time in the presence of selection is less than the coalescent time under neutrality. We now present the less intuitive result that this is not just a rescaling of time for the coalescent process, rather the shape of the coalescent changes: the coalescent is less concave. For the neutral case, it takes  $2N$  generations for a mutation to become fixed, halfway through that process is  $N$  generations. From the approximation to the coalescent that the pedigree size  $t$  generations ago is  $N(2/t)$ , it follows that the pedigree size is 2 when the fixation process is half over. As previously shown, in the presence of selection, the coalescent time is approximately  $(.5 + \ln(Ns))/s$ . If  $Ns = 100$ ,  $((.5 + \ln(Ns))/s)/2N = .026$ , hence the fixation time will be approximately  $.052N$ . The mutation will increase more rapidly under selection than under neutrality, hence the frequency of the mutant allele at time  $.026N$  will be at least what it was under neutrality, i.e.,  $.013N$ . But by the branching process approximation,  $2s$  of the alleles should be immortal (have lineages which do not become extinct), hence all those alleles should be in the coalescent. But  $2s \times .013N = 2.6$  (assuming  $Ns = 100$ ), which is greater than 2, so the coalescent will need to be larger halfway through the fixation process. This entails earlier growth in the coalescent, so the initial tail of slow growth is not as long, hence the flourish of growth at the end is not as pronounced (note that the reduction in the coalescent time, in this case by a factor of 40, greatly increases the rate of growth for the entire coalescent, so the final flourish of growth will be greater than under neutrality, but the relative increase of its growth will not be as great as for other portions of the trajectory).

Kaplan, Dardeen, and Hudson. 1988. GENETICS 120:819-829.

Kingman. 1982. STOCH. PROC. APPLICAT. 13:235-248.

Neuhauser and Krone. 1997. GENETICS 145:519-534.

van Herwaarden and van der Wal. 2002. THEOR. POPUL. BIOL. 61:311-318.