

R. B. Campbell  
Department of Mathematics  
University of Northern Iowa  
Cedar Falls, IA 50614-0506  
http://www.math.uni.edu/~campbell  
campbell@math.uni.edu  
(319)273-2447

June 2004  
Fort Collins  
Colorado

## The Coalescent with Selection

### Introduction

In the case of selective neutrality, the expected time until fixation of a new mutation conditioned on fixation (the fixation time) is equal to the expected time since a common ancestor (the coalescent time) (Campbell, 1999). The fixation time can be parsed as the time from first occurrence of the mutation to the most recent ancestor of the fixed population, plus the time from the most recent ancestor of the fixed population until fixation; both these quantities are meaningful with selection. The coalescent time can be parsed into the time from the present to fixation, plus the time from fixation to the most recent common ancestor (going backward in time); but selection does not occur after a mutation is fixed (after it is fixed, its fitness is one), hence the time between the present and fixation is not meaningful with selection. The times between the most recent common ancestor and fixation, and between occurrence of a mutation and the most recent common ancestor (summing which yields fixation time) for a selected mutation are calculated below.

The coalescent/branching structure of the coalescent of Kingman (1982) is robust, the effect of selection is manifested in the time between the coalescent/branching events (Neuhauser and Krone, 1997) which depends on the population size. For a selected mutation, this is the number of that allele in the population. It is necessary to distinguish between three numbers: the total number of alleles in the population (the population size  $N$ ), the number of copies of the selected allele in the population, and the number of copies of the selected allele in the coalescent (*i.e.*, the size of the coalescent  $k$ ). It is the number of copies of the selected allele in the population which determines the rate of coalescence.

### The coalescent time of a selected allele

Although the initial rate of increase of a selected allele will manifest great stochasticity, once there are several copies of the allele, the rate of increase should be close to the deterministic exponential increase specified by the fitness of the selected allele. Because the most recent common ancestor will generally be more recent than the first occurrence of a favorable mutation, there will be several copies of the mutation at the time of the most recent common ancestor, and deterministic exponential growth will be a good approximation to the number of mutant alleles in the population. The specific question which we are addressing is the time between fixation of a selected allele and the most recent common ancestor. Exponential growth gives the number of copies of the selected allele in the population  $t$  generations before fixation as  $N(t) = Ne^{-st}$  where  $N$  is the population size and  $s$  is the Malthusian parameter. This allows us to rewrite  $\frac{dk}{dt} = \frac{-k(k-1)}{2N}$  as  $\frac{dk}{dt} = \frac{-k(k-1)}{2Ne^{-st}}$  where  $k$  is the number of copies of the selected mutant allele in the coalescent. (The negative signs occur because  $t$  is measured backward in time.) Solving this differential equation with the initial condition  $k(0) = N + .59$  (which is chosen so that  $\int_{1.59}^{2.59} \frac{1}{k(k-1)} dk = .5$ ) provides that the most recent common ancestor ( $k = 1.59$ ) occurred approximately  $\frac{\ln 2Ns}{s}$  generations before fixation. (initial and common ancestor  $k$  values close to  $N + .59$  and 1.59 give the same approximate result, but  $N$  and 1 cannot be used.) Essentially the same result is obtained by assuming drift for the early increase of the mutation and deterministic increase for the latter increase. A similar result has been obtained by van Herwaarden and van der Wal (2002). Note that this is less than the deterministic time to increase from 1 to  $N$  copies,  $\frac{\ln N}{s}$ .

(over)

## The neutral case

The argument that an allele can only come from its kind can be applied in the case of a neutral mutation. In this case the frequency of a mutation, conditioned on fixation, increases linearly for  $2N$  generations. The governing differential equation is  $\frac{dk}{dt} = \frac{-k(k-1)}{2(N-t/2)}$ , which for the same initial condition  $k(0) = N + .59$  yields for the most recent common ancestor ( $k = 1.59$ )  $t = \frac{4}{3}N$ . This is consistent with the well known result that in the neutral case the expected time since a common ancestor is  $2N$  generations, because restricting the ancestry entails considering only the generations between fixation and the most recent common ancestor, but  $2N$  generations allows generations between the present and fixation. Simulations of populations with 100 and 200 individuals showed that the number of generations between the present and fixation, and the number of generations between fixation and the most recent common ancestor are about equal. The  $\frac{4}{3}N$  generations before fixation is approximately  $N$  generations which would be expected. The approximation in choosing appropriate initial conditions and final frequency for the differential equation further suggests that this agreement is reasonable.

## The fixation time

The fixation time can be calculated by adding the time from the first occurrence to the most recent common ancestor to  $\frac{\ln 2Ns}{s}$ , the time from the most recent common ancestor to fixation. The question is how much time elapses from the first occurrence of a selected mutation destined for fixation until two of its descendants have lineages which persist to fixation.

In all populations, there is a single lineage which contains all the common ancestors. Those individuals average approximately one sibling per generation (under the Poisson progeny distribution). The branching process approximation gives the probability that one of those siblings persists as  $2s$ , hence the expected time until a sibling that persists is  $\frac{1}{2s}$  generations. This provides the fixation time for a selected allele  $\frac{1}{2s} + \frac{\ln 2Ns}{s}$  generations. (The approximation which obtained the coalescent time  $\frac{\ln 2Ns}{s}$  ignored terms with  $s$  in the denominator, but they were smaller than  $\frac{1}{2s}$  and  $\frac{\ln 2Ns}{s}$  for  $Ns \gg 1$ .)

## Discussion

The results of this work are that the time from most recent common ancestor until fixation of a selected mutation is approximately  $\frac{\ln 2Ns}{s}$  generations but the time from occurrence of a selected mutation until fixation is approximately  $\frac{1}{2s} + \frac{\ln 2Ns}{s}$  generations.

Selection has been interpreted as  $4Ns \gg 1$ , and approximations based on this have been employed; selection cannot be contrasted with neutrality by taking the limit as  $s \rightarrow 0$ .

Only the total coalescent time has been calculated, but because the coalescent rate is inversely proportional to the number of potential ancestors, the coalescent/branching is accelerated most near the common ancestor and least near fixation. The acceleration also occurs in the neutral case, where it can be interpreted as the inclusion of branches which contributed to fixation but died off before later populations sharing the same common ancestor. Because the coalescent rate depends on the number of mutant alleles rather than the coalescent size, an exact description is difficult.

We are only looking at one locus, with selection at that locus. We are not considering the effect of selection at other loci on the coalescent (Barton and Etheridge, 2004). Rather we are essentially looking at the coalescent in an exponentially growing population (Slatkin and Hudson, 1991).

## References

- Campbell. 1999. THEOR. POPUL. BIOL. 55:260-269.
- Barton and Etheridge. 2004. GENETICS 166:1115-1131.
- 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.
- Slatkin and Hudson. 1991. GENETICS 129:555-562.
- van Herwaarden and van der Wal. 2002. THEOR. POPUL. BIOL. 61:311-318.