

Coalescent size versus coalescent time with strong selection

Fairbanks AK/June 2005

R. B. Campbell/Department of Mathematics/University of Northern Iowa/
Cedar Falls IA 50614-0506/campbell@math.uni.edu

Introduction

As noted by Kingman (1982a,b), the coalescent structure is robust. The structure (topology) remains if selection is in some manner equivalent for all individuals, the effect of selection is rather to change the rate at which the coalescent is traversed. This results in a significant alteration to the shape of the coalescent. In particular, with strong selection the coalescent near the root is traversed much more rapidly than the coalescent near fixation, hence the reduction in the coalescent time is much greater than the reduction in the total number of individuals in the coalescent. We illustrate this with calculations for the coalescent time for the entire population, the coalescent time for two individuals, and the total size (number of individuals) of the coalescent.

Coalescent of a selected allele

Our approach to studying the coalescent of a selected allele is noting that the parent of an individual must be of the same type as the individual. Hence one does not use the entire population size, but the number of selected alleles in the previous generation, to calculate the coalescent (Slatkin 1996). This is equivalent to the coalescent of an allele in a growing population. Because the most recent common ancestor will generally occur many generations after the mutation first appeared, there will be several copies of the selected allele at the time of the most recent common ancestor, and deterministic exponential increase of the selected allele can be assumed (Slatkin and Hudson, 1991). For comparison with the neutral case, linear growth is assumed for a neutral mutation based on the growth of a branching process conditioned on non-extinction.

Coalescent time for the population

We do not use the estimate $\frac{2\ln(2N)}{s}$ (Campbell 1999, Durrett 2002) for the coalescent time because $\frac{2\ln(2N)}{s} > 2N$ whenever $\ln(2N) > Ns$, which includes reasonable values of N and s . Rather we use $\frac{2\ln(2Ns)}{s}$ (van Herwaarden and van der Wal 2002); this estimate can also be obtained by combining initial drift with deterministic latter increase as suggested by Ewens (1979) or by calculating the coalescent time within the increasing population of the selected mutation. With this selection coalescent time the ratio of coalescent time with selection to neutral coalescent time is $\frac{\ln(2Ns)/s}{2N} = \frac{\ln(2Ns)}{2Ns}$, which depends only on Ns , and is equal to .15 for $Ns = 10$ and .026 for $Ns = 100$. Hence strong selection entails a large reduction in the coalescent time.

Coalescent time for two individuals

The coalescent time under selection for two individuals was calculated numerically as $1 + \sum_{i=1}^{(\ln N)/s-1} \prod_{t=1}^i (1 - \frac{1}{Ne^{-ts}})$. ($\ln N/s$ is the number of generations to increase exponentially from 1 to N copies, and Ne^{-ts} is the number of selected alleles at time t .) For s ranging from .0001 to .1 the results were approximately $.2N$ for $Ns = 10$ and $.04N$ for $Ns = 100$. The neutral expected coalescent time for two individuals is N , hence the reduction in the coalescent time is less than for the entire population, but of the same order of magnitude. Because two individuals may coalesce long after the common ancestor of the population where growth of the coalescent is accelerated most, such lesser reduction in the coalescent time is expected.

Table 1: Relative coalescent times and sizes

	$\frac{\ln 2N_s}{2N}$	(selection coalescent size)/(neutral coalescent size)					
		$N = 10^2$	10^3	10^4	10^5	10^6	10^7
$N_s = 10$	0.15	0.47	0.60	0.69	0.75	0.79	
$N_s = 100$	0.026		0.32	0.45	0.55	0.62	0.68

Cumulative coalescent size

The cumulative coalescent size was obtained by finding the size of the coalescent (number of individuals in a generation) as a function of t , and summing that from the present back to the most recent common ancestor (this was done as deterministic functions). [It merits mention that the values obtained for the neutral case are all within 4% of the estimate $2N(\ln(2N) - .05)$ (Campbell 2003).] The results as presented in Table 1 show that selection only reduces the cumulative size of the coalescent to between 32% and 79% of the neutral cumulative coalescent size.

Discussion

This study confirms the well known result that it is very hard to detect selection. Although selection greatly accelerates the rate of fixation of a beneficial mutation, the effect of the selective sweep on genetic variation is far less significant. Genetic variation in the population (measured as number of alleles segregating) due to neutral mutation at other loci is proportional to the number of individuals in the coalescent, because segregating mutations will have originated in those individuals. Hence the genetic variation after fixation of a strongly selected mutation will be about the same as in the absence of a selective sweep (between 32% and 79% of the value for a neutral coalescent for the parameters used above).

Literature cited

- Campbell, R. B.** 1999. The coalescent time in the presence of background fertility selection. *Theor. Popul. Biol.* 55,260-269.
- Campbell, R. B.** 2003. A logistic branching process for population genetics. *J. Theor. Biol.* 225,195-203.
- Durrett, R.** 2002. *Probability Models for DNA Sequence Evolution*. Springer-Verlag. New York.
- Ewens, W. J.** 1979. *Mathematical Population Genetics*. Biomathematics vol. 9. Springer-Verlag. New York.
- Kingman, J. F. C.** 1982a. The coalescent. *Stoch. Proc. Applicat.* 13,235-248.
- Kingman, J. F. C.** 1982b. On the genealogy of large populations. *J. Appl. Probability* 19A,27-43.
- Slatkin, M.** 1996. Gene genealogies within mutant allelic classes. *Genetics* 143, 579-587.
- Slatkin, M. and R. R. Hudson.** 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555-562.
- van Herwaarden, O. A. and N. J. van der Wal.** 2002. Extinction time and age of an allele in a large finite population. *Theor. Popul. Biol.* 61, 311-318.