

## Fixation of a Gene by Drift is Unlikely in Large Populations, neither Identity by Descent nor Identity by Type

R. B. Campbell/University of Northern Iowa/*campbell@math.uni.edu*

June 2011

### Introduction

The paradigm for evolution is that a single mutation occurs which, through natural selection or drift, replaces the wild-type and becomes fixed in the population. This is incorrect in two respects: there is not a single ancestor of all the genes at a locus (common ancestors exist for nucleotide base pairs, not entire genes), and mutations do not become fixed in the population (there are always multiple forms present). The alleles at a locus in a large population are not identical, neither by descent nor by type. At the level of the gene, evolution is not a sequence of steps, but an ongoing process with inherent variation.

### Model/Assumptions

The standard model of random mating with a Poisson progeny distribution and selective neutrality is employed. Both the frequency of mutation at a base pair site and the frequency of crossing over between adjacent base pair sites are  $10^{-8}$ . A gene is 1000 contiguous base pairs. The population size is  $10^6$ , hence this is our definition of large.

### Identity by Descent

Although the coalescence process provides that all the base pairs at any site are identical by descent, crossing over allows that different base pair sites may have different ancestors. We can bound the balance between crossing over and coalesce with the inequality  $r \leq E[k(k-1)/(4N)]$ , where  $N$  is the population size,  $k$  is the number of individuals ancestral for some base pairs in the gene, and  $r$  is the frequency of recombination within a gene. The result (<http://www.math.uni.edu/~campbell/genecoalfly.pdf>) is that for the parameter values in the assumptions the probability that a single individual contained ancestral material for all the base pair sites in a gene is less than .0001. (Indeed, in some generations in the past a single individual will carry genetic material ancestral to all base pair sites, hence this probability is the relative frequency of such generations.) In general, a gene does not have a common ancestor, but an ancestral pool.

### Identity by Type

To assess the probability of identity by type in the entire population, we ignore recombination (so that there is a common ancestor of the entire gene) and compute the probability that no mutation occurred since the common ancestor. It is necessary to know how many descendants of the most recent common ancestor are ancestors of the present population. This is the cumulative size of the coalescent which is approximately  $2N(\ln(2N) - .05)$  (<http://www.math.uni.edu/~campbell/evol05.pdf>). Only if no mutations occurred in any of these ancestors will there be identity by type. Therefore, assuming that mutation is a Poisson process, for the assumed parameter values the probability of identity by type (i.e., no mutation occurred since the MRCA) is approximately  $\exp(-2000000(\ln(2000000) - .05) \times 10^{-5}) = 10^{-126}$ .

### Different Parameters

If the population size is 1, the population will usually manifest both identity by descent and identity by type. The calculations for identity by descent employ the product of the population size, gene size, and crossover rate. Decreasing this product by 10 raises the upper bound for probability of a common ancestor to 0.2; increasing this product by a factor of 10 reduces the upper bound to  $10^{-15}$ . Because  $\ln N$  does not change as rapidly as  $N$ , the probability of identity by type is essentially a function of the product of the population size, gene size, and mutation rate. Decreasing this product by a factor of 10 increases the probability of identity by type to  $10^{-11}$  or  $10^{-13}$ , increasing this product by a factor of 10 decreases the probability of identity by type to essentially 0. If there are introns within a gene, that will not effect the mutation rate, but crossing over will be more likely, hence identity by descent less likely. Selection can maintain a polymorphism, or increase the probability of identity by type.

### Discussion

The parameter values employed ( $N = 10^6$ , 1000 base pairs in a gene, frequency of mutation or recombination =  $10^{-8}$ ) demarcate whether or not genetic identity (by descent or type) occurs. In large populations, evolution is a continuous process, always in transition; populations do not move from one gene to another. Genes do become fixed in small populations, hence an isolated subpopulation can establish a distinct genome and speciation can occur.