**Lab Assignment of Computational biology**

**Genetic Processes in Finite Populations**

**Introduction**

Genetic drift is the random movement of allele frequencies between generations. More specifically, this movement is random with respect to the direction in which the allele frequencies change. Of course, the probabilities of movement in one direction or the other need not be equal, but even lopsided probabilities do not guarantee that in any particular population or in any particular generation the movement will be in one perfectly predictable direction. This randomness is in contrast to the effects of a deterministic force like natural selection in a very large population; when one allele is favored over another because of the increased fitness of phenotypes carrying that allele, the favored allele will always increase in frequency. This distinction is the reason we call genetic drift a stochastic phenomenon.

Two factors govern what happens in genetic drift. First, the effective population size determines the magnitude of this random change in allele frequencies in a single generation. Random sampling effects are more important in smaller populations because the probability that a sample of alleles will deviate in frequencies from the true frequencies is greater in small populations than in large ones. Second, the true allele frequencies determine the probability of movement in any particular direction. If one allele's true frequency is 0.90 and the other's is 0.10, a small sample is more likely to show an increase in the frequency of the common allele than a decrease.

Unless other forces are operating, drift leads inevitably to fixation of one allele. Let's investigate the influence of population size and initial allele frequencies on the cumulative changes in allele frequencies in the following simulations.

**Requirement**

Write a computer program to simulation the genetic drift.

Parameters:

1, run time: 2-30 generation

2, population size: 10-2000

3, Initial loci: 5

4, set frequency individually: F1=?, F2=?, F3=?, F4=?,F5=?

The output of your program should be the frequency of each loci at each generation, plot it in excel.

Here is a reference code by perl:

<http://sayer.lab.nig.ac.jp/~saitou/evolutionary_genomics/random_drift_win.pl>

Here is a reference software:

http://www.cbs.umn.edu/research/resources/populus

Bonus:

Add the natural selection to your program by choosing the fitness of each loci.