**INSTALLATION/COMPILING/RUNNING**

To compile this program, keep all the files contents that end in .cpp and .h within the same directory, and type:

g++ -O3 -I /Path/To/Boost/Library \*.cpp

where the filepath to the boost C++ library (www.boost.org) is included after the –I flag. This should create a program named “a.out”, which may be renamed. For the moment, ignore files in the “AsexVersion” subdirectory.

**“Parameters” File**

This file must be in the same directory as the compiled program (above) as it contains 10 input parameters for the simulator. These parameter values must be on the second line of the file and, from left to right, include:

1. Number of loci
2. Number of individuals (population size)
3. Number of generations to run the simulator
4. Total recombination rate, i.e. the physical per base pair rate multiplied by the genomic segment simulated (see 8.))
5. Geometric distribution parameter to determine tract lengths transferred between individuals, must be between 0 and 1, and the mean tract length is the reciprocal of the value specified here
6. Additive genetic variance (each locus has additive effect , where *L* is the number of loci, and *VA* is the value specified here)
7. Epistatic genetic variance (determines the variance specified in the normal distribution  from which pairwise epistatic effects are drawn, where *VI* is specified here
8. Circular genome size (in bp)
9. The length of the segment (in bp) in which loci are randomly distributed, must be less than or equal to the genome size specified above
10. Number of simulation replicates

Note that there is no mutation rate; each locus is represented as a +/- 1 which are each at 50% frequency in the population at the beginning of the simulation. These alleles are also randomly distributed among individuals such that loci start out in linkage equilibrium (but slight linkage disequilibrium may arise from small population sizes). Allele frequencies change and linkage disequilibrium increases with the sampling of individuals across generations.

**OUTPUT FILES**

There are two types of output files from this program labeled "Results" and "EpistasisTable", and enumerated by simulation replicate.

**“EpistasisTable” file**

The “EpistasisTable” file has a list of all the pairwise epistasis selection coefficients, drawn from a normal distribution with mean zero and variance specified in the “Parameters” file that needs to be in the same directory. This variance corresponds to the *VI* in the variance of the normal distribution specified as (2 *VI/L(L-1)*), where *L* is the number of loci. These selection coefficients are ordered such that all pairwise interactions involving the leftmost locus are printed first, going from left to right through all other loci until the first *L*-1 pairs are printed. Here, the interaction in position *L*-1 in the list corresponds to the interaction effect between the leftmost and rightmost loci. After this, all pairwise interactions involving the next leftmost locus (to the right of the first locus considered) are printed, again going from left to right through all loci until *L*-2 pairs are printed. Remaining pairwise interactions are printed in a similar fashion.

**“Results” file**

This file begins with some general information from the simulation. First printed are the positions (POSITIONS) of the loci, which were randomly drawn integers between 1 and the size of the genome specified in the “Parameters” file. Second, the additive fitness effect of each locus is printed (AddFitConst). This is a single number:. Third are the epistatic fitness effects for each pairwise interaction (EpistasisTable\_Ns) as specified in the “EpistasisTable” file but multiplied by the population size of the simulation. There should be *L*-1 lines, where the first line represents the *L*-1 interactions with the first (leftmost) locus, the second line represents the *L-2* unique interactions the second locus, and so forth.

Next in this file are some summary statistics from the beginning of the simulation at time t=0. The first of these summary statistics is Dprime\_BY\_DIST\_gen0, which represents all the D’ values for each locus pair, where D’ here is a popular measure of linkage disequilibrium (Lewontin 1964). D’ values are printed for each locus pair in the same way as the EpistasisTable\_Ns section. You also might notice “ThetaDecay\_1.0”, which represents the remaining diversity (as measured as the average number of pairwise differences). Here it is 1.0 because 100% of the diversity remains because the simulation has not started.

More summary statistics are then printed, including ThetaPi\_gen0 (average number of pairwise differences at generation 0), Wmean\_gen0 (mean fitness), Wvar\_gen0 (the variance in fitness), Wskew\_gen0 (skew, or third moment of the fitness distribution), and WabsDist\_gen0, which is followed by the absolute fitness value for each individual in the population. Only relative fitness values are used to sample individuals with replacement across generations to mimic a Wright-Fisher population.

These summary statistics are then printed at specific generations, and labels are followed by “\_gen#” where # is the number of generations that have elapsed since t=0, and included in these same lines are a “ThetaDecay” label which represents the proportion of diversity that remains. Diversity depletes over time due to genetic drift and natural selection.

**COMPARING RESULTS TO CLONAL/ASEXUAL EVOLUTION**

I found it useful to conduct these simulations with a control that has no recombination. These simulations start out with all loci in linkage equilibrium, and linkage disequilibrium builds up over time from genetic drift and selection. Selection on epistatic interactions can spread certain haplotypes through a population and generate more linkage disequilibrium, beyond what we would expect with just genetic drift. However, the amount of linkage disequilibrium that can be generated by selection on epistatic interactions depends on recombination, so it is natural to compare observed results with recombination to simulations with no recombination, to see if linkage increases with time at similar or different rates.

To do this, move into the “AsexVersion” directory, and compile the program using the exact same g++ command above. This compiles the same program but with one exception: instead of randomly drawing pairwise epistatic fitness effects from the specified normal distribution, it loads the same values that were used in the previous simulation in order to make results directly comparable. Consequently, the “EpistasisTable” files need to be in the same directory as the compiled program, as well as the same “Parameters” file that was used for simulations with recombination. It also loads the position information of the loci used in previous simulations with recombination, so the “Results” files also need to be in the same directory. To make this easy, you can simply rename this compiled program generated in this directory and move it to the previous directory in which you simulated epistasis with recombination.

This program prints results files named “AsexResults” that have the same structure as the results file printed from the program that includes recombination.