

## Instructions for the Popsim program

### To start

This program allows you to investigate the effects of natural selection, genetic drift, mutation and gene flow (migration) on the frequencies of a single gene in a population.

The command which starts the simulation is **Go** from the **Run** menu. However when you first access the program you will find that this item is inactivated. Before you can activate it, you must tell the program:

- (a) what selection there will be
- (b) what population size
- (c) where the gene frequencies should start.

These values are grouped together in the **Parameters** menu.

### Other parameters

There are four other values which can optionally be set in the **Parameters** menu. You can nominate:

- (1) Mutation, which can be from m to + or back mutation from + to m
- (2) Gene flow (migration)
- (3) The number of generations (initially set to 50)
- (4) Speed of the simulation, which can either be increased or decreased.

You can tell if any population parameter has been set. If so, there will be a check mark next to the item in the **Parameters** menu.

### Choose model

In some, but not all, implementations of the Popsim program, the top item of the **Parameters** menu is **Choose model**. This allows you to choose between a number of models which have been pre-set. This includes models of selection, drift, mutation, and various combinations. Each model allows a choice between finite and infinite population (see below).

**It is important to note that the Choose model command simply sets the 'defaults' for each model. You still need to confirm the various parameters in the usual way.**

However you will be prompted with a choice of parameters which is appropriate for the model you have nominated. You can accept or change these. There may also be a series of exercises and questions for each model, accessible from the **Display** menu.

## Control of the simulation

To stop the simulation at any time, hold down the mouse button. You can then re-start from the **Run** menu. At any time it is also possible to access the **Repeat** command from the **File** menu. This allows a re-start of the same simulation, or alternatively you may change one or more parameters. Alternatively you can use the **New** command to re-start the simulation with all new parameters.

Other items which can be accessed from the **File** menu include a Print command (assuming that a printer is available), and a **Save As...** command for saving the output to disk as a text file. There is also a **Get Info** command which shows the current parameters set for the run.

## Displaying results

The results of simulations appear in a text window. You can ask the program to display means and variances from the **Display** window.

You can also display the results graphically at any stage during the simulation. The command **Update graph** will result in a graph of the current simulation, with generation number plotted on the X-axis and gene frequency on the Y-axis.

The **Show window** item enables you to go between text and graph. However the text window will automatically be chosen by the program except when graphs are being drawn.

## The model

The simulation uses the 'wild type' notation, and assumes the following single locus genotypes:

$+ / +$                        $m / +$                        $m / m$

The  $m$  (mutant) gene can be any mutation. Despite the notation, it can be recessive or dominant. It can also be deleterious, favourable or neutral.

Although the simulation takes into account the three genotypes, for simplicity the program prints out only gene frequencies or gene numbers. Since the frequency of the  $+$  allele is just the complement of the  $m$  allele, it is in fact only necessary to print out the  $m$  gene frequency.

## Population size

This program has two basic modes. You can specify

- (1) Infinite population size
- (2) A particular finite population size.

In the first mode, the program runs a deterministic simulation. It prints out the expected gene frequency at any generation for the given selection and/or mutation values specified. A single value is printed out, which is the frequency of the  $m$  gene.

In the second mode, each stochastic population simulation is a separate event. The program simulates as many populations as possible, the number being chosen to fit across the screen, and prints out the number of  $m$  genes for each. Note that the total number of genes is 2x the specified population size.

## Selection

The program allows you to specify any selection model. The defaults for the selection values may be the following:

Genotypes	+ / +	$m$ / +	$m$ / $m$
Selective values	1.00	1.00	1.00

which implies no selection.

You can nominate any intensity of selection for any genotype, but the selective value cannot be negative. Also the program will ask you to confirm the selective values if one of the three selective values is not equal to unity, since selective models are conventionally given in this way. You can either divide through by a suitable factor to make one of the selective values equal to one, or confirm the selective values you have entered.

## Mutation

The specification of mutation is optional in the program. The stochastic simulation runs in a rather different way depending on whether or not mutation is specified.

A population can become 'fixed' when either the + or the  $m$  gene is lost. These two states are denoted by + or - in the absence of mutation. If all populations are fixed, then the simulation stops.

In the presence of mutation, either forward mutation from + to  $m$  or back mutation from + to  $m$ , fixation or loss of the  $m$  gene are denoted by . or a blank respectively. The simulation will not stop when all populations are fixed, since new mutations may be produced at any time.

## **Gene flow**

Specification of gene flow, like mutation, is optional. There are two types of gene flow:

- (1) Between populations. This assumes that each of the simulated finite populations is an 'island' population which obtains immigrants at the same rate from all other populations. If you specify a rate such as 0.01, then 99% of the genes will be non-immigrant in each generation, and 1% will be immigrant chosen randomly from each of the other populations.
- (2) Long range. This assumes that there is a source of immigrant genes from a population whose gene frequency remains unchanged and can be specified. This type of gene flow can also be assumed for deterministic calculations with an infinite population.