Instructions for waiting time experiments:

A special version (2.5.1) of Mendel's Accountant (Mendel) allows us to determine the waiting time for establishing a new string of nucleotides within a given biological population, at a particular location within the organism's genome. At present Mendel's primary limitation in this regard is it has a lower limit for the designated mutation rate per string (mutation rate per string must be 10^{-5} or higher). This limitation has to do with limits of current random number generators. When lower mutations are of interest, runs can be done with higher mutation rates, and then fixation times corresponding to a lower mutation rate can be approximated using proportional correction upward of the waiting times (or by extrapolation).

When doing waiting time experiments, Mendel starts with a specified string of nucleotides. That string undergoes mutation, and the resulting mutant strings segregate and drift within the population from generation to generation. Whenever the target string arises, the individuals carrying it are given a specified reproductive advantage. Any target string that arises is protected from any further mutation. Mendel allows the mutation/selection process to continue until the target string reaches fixation.

After the target string reaches fixation, Mendel summarizes and plots the results of the experiment, and provides the following data points: a) number of generations to first instance; b) generations to the effective instance; c) generations from effective instance to fixation; d) total number of generations to final fixation; and e) total number of instances arising during the experiment.

Details on how to do waiting time experiments –

If you use a MacIntosh computer running OS X Yosemite, download Mendel 2.5.1 from http://sourceforge.net/projects/mendelsaccount/?source=typ_redirect

If your Mac has a different operating system, when you try to do a Mendel run it might immediately indicate it finished running (but it actually never ran). In this case you may need to update your operating system.

If you do not have a Mac with the appropriate operating system, we will soon have a Windows version of Mendel available for download. Until then, Dr. Franzine Smith may be able to help you (fsmith@fmsfound.org).

If you have downloaded Mendel, unzip it, double click on the Mendel fold, and then double click on "start server". A box will open saying the server has been started (now close this box). A **Mendel tab (labeled SciPaaS or SPC)** will automatically appear on your default browser (Google Chrome, Safari, and Firefox all work), and you will see the Mendel control panel.

1. To begin an experiment you can click "start", or you can directly start to enter your desired parameter settings. Each new run is automatically assigned its own unique identifying code. In the box at the upper right, enter a short description of the experiment or else use a case name of your own choosing.

2. To do a waiting time experiment - on the far right choose the tab **special applications**. Click on the check box labeled *waiting time experiments*. A default starting sequence is shown – change it to the

starting sequence desired. Likewise, specify the desired target string (it must be the same length as the starting sequence). You can choose to have (or not have) some matching letters between the starting and target sequence. Caution - if the string is 6 or more nucleotides long – you may run out of patience or memory before the target string arises and goes to fixation. Lastly, choose the fitness benefit that the target string will confer on carrier individuals.

3. When you choose the waiting time option – You will notice (under *neutral mutations* - immediately above) that all mutations and all strings are automatically made neutral (all strings are non-beneficial - except the target string). Other settings that are automatically altered are: a) under the **population tab** - *recombination model* is auto-set to "recombination suppressed"; b) under the **mutation tab** - *distribution type* is auto-set to "all mutations equal". Under the **computation tab** – if your run is substantial in size you should uncheck the box for *automatically allocate memory*, and then manually increase "maximum number of neutral mutations per individual" to 100,000 or perhaps more (as needed). If an experiment runs out of allocated memory, the run will simply stop, and then provide a diagnostic warning at the end of the output file.

4. Lastly, go to the **basic tab**. Adjust *Total non-neutral mutation rate* as follows. Note - this label is misleading – for this application it should read *Total mutation rate per string* (we need to fix this). To specify this number you need to know the mutation rate (per nucleotide) of the organism under study. If it is a diploid organism this rate should be multiplied by two. Then multiply times the number of nucleotides in the string. This will give you the mutation rate (just within the string), per individual. If the organism's mutation rate is too low (I think less than 10^{-5}), you will need to adjust upward the mutation rate - as necessary for the experiment to initiate. If the mutation rate is adjusted upward ,by a certain factor, then when the run is complete the observed waiting times must all be corrected upward by the same factor (see below). Eventually we will solve this technical problem.

5. Under the **basic tab**, modify the population size as desired. Larger populations will shorten waiting times, but will require extra memory. The largest population we can do (depending on string length), is about 1 million.

6. Under the **basic tab** modify the generation number. Depending on the string length and population size, you may need to go up to 500,000 or more generations to observe target string fixation (but using too high a number can require extra memory allocation). As long as you use a high enough generation number, the target string will eventually be fixed (i.e., reaching an allele frequency of more than 99%), and when this happens the run will automatically stop and the waiting time statistics will be generated. If you do not use a high enough generation number, the target string will not have enough time to be fixed and the experiment will simply go to completion with incomplete results.

7. Other possible settings of potential interest –

a) Under basic tab, you may wish to change reproductive rate (defines intensity of selection – all excess population is always selected away every generation).).

b) Under the mutations tab you can alter the dominance settings (first box must be zero or one).

c) Under the selection tab you can change the settings that affect the efficiency of selection.

d) Under the computation tab, you can change the random number seed, for doing replicates.

8. Once you have decided on all the settings, click on the green button that says "continue", and when you are satisfied with the displayed review of settings, click on the green button that says "execute". The run should begin, and the monitor will start to update regularly. For certain experiments there may be a delay before any progress is seen. During the run you can get updates on the run at any time by click on "inputs", or "plots", or output", or "monitor, etc.

9. Within the output you will see that the data is summarized periodically, showing progress. For waiting time experiments these updates include: a) display of the strings of two randomly selected individuals, b) the number of times the target string has arisen, and c) the current allele frequency of the string in the population. Also, allele frequencies of neutral strings (which are not the target) can be seen when clicking first on "plot", and then clicking on allele frequency.

At the end of a run, near the bottom of the output file, there is a summary of the waiting time statistics (see example below). This includes generations to first instance, generations to effective instance, generations to final fixation, and total number of instances. Just below that is displayed a list of all the instances of the string, when they arose and how long each one persisted. The effective instance is also shown.

```
-----------------------------------------------------
POLYGENIC BENEFICIALS SUMMARY:
First_instance_gen, Last_instance_gen, Fixation_gen, Total_instances
        4                   493              520           190
```

Use of elevated mutation rates and subsequent correction of waiting times. When modeling populations of higher organisms, we typically need to enhance mutation rate. This is for two reasons: 1) to get runs to finish is a reasonable time; and 2) to overcome the technical limitation of mutation rates less than roughly .00001. In such cases, after the experiment we need to correct all waiting times proportional to the amount we enhanced the mutation rate. For a human population we routinely assume a mutation rate per nucleotide of 0.001. This is 10^4 higher than the known human mutation rate. We then multiply this times two (for diploidy), and also multiplied by the string length, and we put this number in mutation rate box (this is the mutation rate per diploid string). Because we enhanced mutation rate, when the run is done we must correct for this. We must increase all the waiting times by the same factor (10,000). *This is less than perfect, but the resulting bias is conservative (corrected waiting times end up being significantly shorter than would be seen if we had used the natural mutation rate). To overcome this we will eventually have to write our own random number generator, but even then, each run will take days, weeks, or months.*

Other helpful hints:

1. To do a replicate or slight modification of the same run, just click "restart", and make needed changes.
2. For an entirely new run, starting with the default settings, just click "start".
3. To review or restart any previous run, just click on "jobs".
4. Plots and data files can be cut and pasted as desired for use in publications and Excel spread sheets.