Application of Mendel Accountant in Population Genetic Studies

Esmaeilzadeh Mahdi [1], Kazemzadeh Fariba [2]

[1.] Department of Basic Science, Shirvan Branch, Islamic Azad University, Shirvan, Iran
[2.] Young Researchers Club, Shirvan Branch, Islamic Azad University, Shirvan, Iran
mehdi_dna@yahoo.com

Abstract: During the last years gene interaction networks are increasingly being used for the assessment and interpretation of biological measurements. Knowledge of the interaction partners of an unknown protein allows scientists to understand the complex relationships between genetic products, helps to reveal unknown biological functions and pathways, and get a more detailed picture of an organism's complexity. Being able to measure all protein interactions under all relevant conditions is virtually impossible. Hence, computational methods integrating different datasets for predicting gene interactions are needed. Mendel's Accountant is an advanced numerical simulation program for modeling genetic change over time and was developed collaboratively by Sanford, Baumgardner, Brewer, Gibson and ReMine.Mendel's Accountant (hereafter referred to as "Mendel") is a user-friendly biologically realistic simulation program for investigating the processes of mutation and selection in sexually reproducing diploid populations. Using a standard personal computer, the MENDEL program can be used to generate and track millions of mutations within a single population.MENDEL provides biologists with a new tool for research and teaching, and allows for the modeling of complex biological scenarios that would have previously been impossible.
[Esmaeilzadeh M, Kazmzadeh F. **Application of Mendel Accountant in Population Genetic Studies**. *Life Sci J* 2012;9(4):5317-5320] (ISSN:1097-8135). http://www.lifesciencesite.com. 791

Keywords: Mendel's Accountant, Mutation, Population Genetics

1. Introduction

Population geneticists have used mathematical modeling for over 75 years to understand better how mutation and selection affect population dynamics. Recent advances in numerical simulation and the wide availability of low cost computational resources now make possible an alternative way to understand how populations change over time. Numerical simulation offers the ability to treat complex biological situations where an analytical solution would be cumbersome, if not impossible. Numerical simulation allows the study of the complex interactions of many biological factors simultaneously. This is generally not practical using traditional methods. The numerical approach provides great flexibility and allows a researcher or student to explore parameter space quite rapidly, without detailed knowledge of the mathematical techniques that underlie the classical theoretical approach.

At its most basic level, the task of modeling mutation and selection in a population over many generations can be viewed as a bookkeeping problem in which random events play a major role. Mutations are continuously entering and leaving any population.

When a new mutation arises, it may or may not be transmitted to an individual's progeny, depending on whether or not the chromosome segment carrying the mutation segregates into the gamete from which the progeny is derived. Generally speaking, mutations that occur near one another on the same chromosome are likely to be inherited together. Therefore, tracking mutation location in the genome is important if one desires to account for mutational linkage. In addition, in higher organisms during meiosis there are about two crossovers per chromosome pair (Santiago & Cabellero, 2000). This random phenomenon of crossover also must be part of the simulation in order to treat linkage in a realistic manner.

Random mutations tend to differ greatly from one another in their effects on genotypic fitness. The fitness effect of a given mutation can be positive or negative, can range from lethal to beneficial, and can vary from fully dominant to fully recessive.

How the effects of multiple mutations (at different loci within the same individual) combine with one another (additively or multiplicatively) also influences the overall genetic fitness of an individual. The effectiveness of selection (that is, its power to alter individual mutation frequencies) is limited by the surplus population available, which in turn depends on the population's average fertility level. Selection efficiency is further limited by factors such as random fluctuations in environmental conditions. Generally speaking, reproduction in nature has a significant random component and is only partially correlated with the fitness of the genotype. All these variables influence actual genetic change over time

and must be modeled accurately if a simulation is to be biologically relevant.

2. Approach

Although there are many programs for genetic data analysis, comparatively little effort has been devoted to software development for detailed simulation of the processes of mutation and selection (Balloux, 2001). Numerical strategies for population genetics modeling have been under discussion for several decades (Crosby, 1973; Fraser & Burnell, 1970), yet it is only recently that computing resources have become widely available to allow large realistic forward-time simulations. The forward-time approach offers the distinct advantage of being able to treat random mutations and natural selection under complex mating/recombination scenarios.

Mendel represents an advance in forward-time simulations by incorporating several improvements over previous simulation tools:

(1) Mendel adds the ability to model mutations as having a continuous, natural distribution of mutation effects.

(2) Mendel allows a user-specified ratio of dominant to recessive mutations.

(3) Mendel uses an infinite sites model, where segregating mutations are distinct and their number is unlimited (or limited only slightly by computer capacities).

(4) Mendel incorporates the concept of heritability and accounts for environmental variance.

(5) Mendel uses realistic chromosome structure with realistic stochastic crossover and recombination, and a high number of linkage blocks (up to order 105).Users can specify the number of chromosome pairs.

(6) Mendel is tuned for speed-efficiency and memory usage to handle large populations and high mutation rates.

(7) Mendel allows control of genetic parameters via a graphical user interface (Figure 1), thereby allowing non-programmers to construct sophisticated simulations.

(8) Mendel provides several forms of graphical output, allowing the user to see the results as the simulation proceeds (Figure 2 shows one of the plots).

In addition, Mendel provides a variety of options for mating, bottleneck events, and population substructure. It is computationally efficient, allowing many problems of interest to be run on ordinary personal computers. In addition, because Mendel is parallelized with MPI (Message Passing Interface), it can exploit multiple processors to run:

(a) Multiple interacting heterogeneous tribes (b) multiple replications of a single case, or (c) a very large population comprised of sub-populations but with sufficient migration to maintain a high degree of genetic homogeneity.

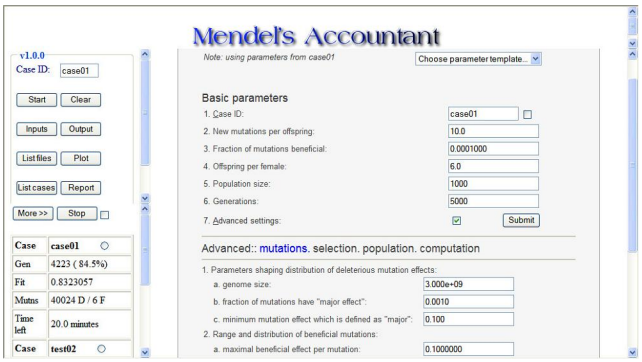

Figure 1. Web user interface of Mendel ' s Accountant showing a portion of the input window

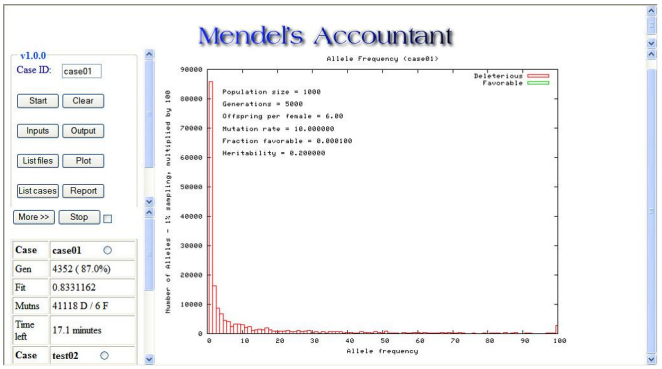

Figure 2. Web user interface of Mendel ' s Accountant showing one of the several output plots the program generates. This plot displays the distribution of deleterious mutations with respect to fitness effect. Red bars represent mutation distribution in the absence of selection. Blue and green bars represent actual accumulated recessive and dominant mutations, respectively, in the presence of selection. The two bars representing mutation classes with effects nearest zero extend beyond the vertical scale of the plot.

3. Analysis

Mendel's input parameters include: number of offspring per female, mutation rate, fraction of mutations which are beneficial, fraction of mutations that are recessive, high-impact mutation threshold, fraction of mutations with effect greater than threshold (two parameters that specify the distribution of mutation effects), number of linkage blocks, number of chromosomes, genome size, mutation effect combining method, heritability of genotypic fitness, type of selection, number of generations, and population size. Mendel's output report is provided at regular generation intervals and includes summary statistics on number and types of

mutations, mean population fitness, fitness standard deviation, and related information. In addition, data for each generation is stored in various files and is also plotted in output figures.

In the example we present below, we employ the following input parameters: number of offspring per female = 6 (4 surplus offspring selected away), mutation rate = 10 per offspring, fraction of mutations which are beneficial = 0.01, fraction of mutations that are recessive = 0.8, high-impact mutation threshold = 0.1, fraction of mutations with effect greater than threshold = 0.001, number of linkage blocks =1000, number of chromosomes = 23, genome size = 3 billion, mutation effect combination method = multiplicative, heritability of genotypic fitness = 0.2, type of selection = probability, number of generations = 5,000, and population size = 1000. Although the current human population size is more than six billion, we have found that population sizes above 1,000 result in only marginal increases in selection efficiency. It is reasonable to expect that, beyond a certain level, larger population size will not result in more efficient selection, because of increased environmental variance.

Some of the output from this example is displayed in the following figures. Fig. 3a shows the mean mutation count per individual plotted with respect to time. A noteworthy aspect of this figure is a nearly exact linear accumulation of mutations, a feature we observe consistently across a broad region of parameter space. The slope of this line is governed primarily by the mutation rate. Selection intensity modifies the slope of this line only to a limited degree. This is because of the preponderance of unselectable "nearly-neutral" deleterious mutations (as further described below).

Fig. 3b shows an initial non-linear genotypic fitness decline, which soon becomes essentially a linear decline. We observe this pattern across most of the parameter space we have explored. Mendel defines an individual's genotypic fitness as 1.0 plus the combined positive and negative effects of all the individual's mutations. In this case mutation effects are being combined multiplicatively. We have found that the slope of this curve (fitness change over time) is determined primarily by three things – the mutation rate, the average mutational effect, and the selection intensity.

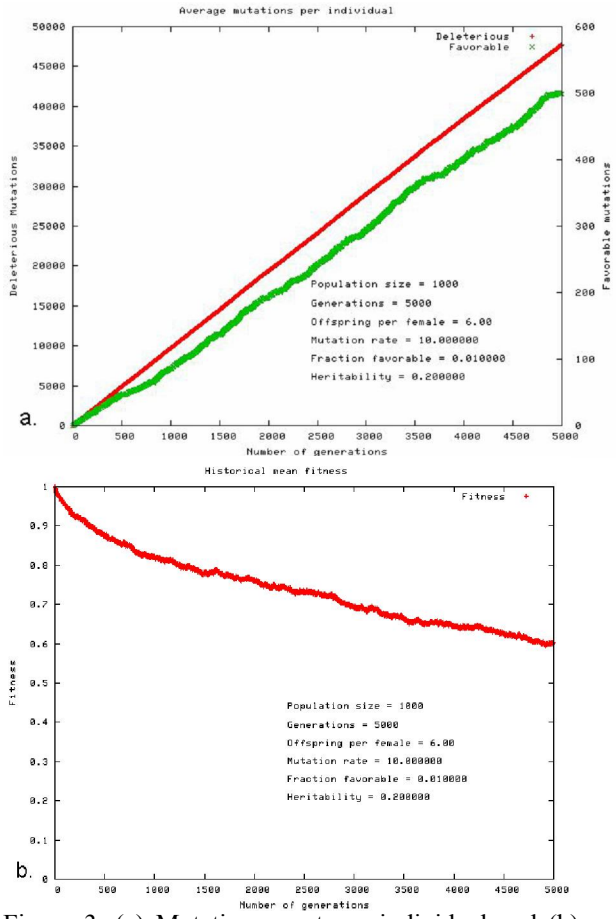

Figure 3 .(a) Mutation count per individual and (b) mean population fitness, plotted for 5,000 generations. (a) shows that deleterious mutations accumulate in close to a strict linear fashion (reaching 47,730–scale on left). Beneficial mutations also accumulate in a linear manner, but their lower number results in sampling error fluctuations (reaching 498– scale on right). (b) shows a progressive decline in population fitness. oak forest (at HB, 30) as compared to pine forest (at HB, 23). Species richness was higher (7.4) at HB and lower at HT (5.0) in oak forest. Similar pattern was found in pine forest, i.e., maximum species richness was at HB (10.5) and minimum at HT (4.7).

Fig. 4 shows the distribution of mutation effects of accumulating deleterious mutations. Mendel employs a distribution of mutation effects (prior to selection), which reflects what is found in nature – a continuous distribution essentially exponential in character. Input parameters such as genome size and the fraction of high-impact mutations define the exact shape of the mutation-effect distribution curve. Because of the shape of the mutation-effect curve, lethal mutations will always be very rare, and a large fraction of deleterious mutations will have near-zero impact.

When strong selection is applied, regardless of the other input parameters, high impact mutations are consistently eliminated quite effectively – especially the dominant ones. However, across a wide range of parameter space the bins nearest to zero fill at essentially the same rate, regardless of whether or not selection is being applied. Experimentally, these "nearly-neutral" mutations are consistently found to be un-selectable – in accordance with mathematical theory (Kimura, 1979; 1983). Mutations with intermediate levels of impact accumulate at intermediate rates. The transition zone between selectable and un-selectable mutations is very wide, especially for recessive mutations. The actual point at which mutations become un-selectable depends on numerous input parameters, but is readily apparent in Mendel's output for any given scenario.

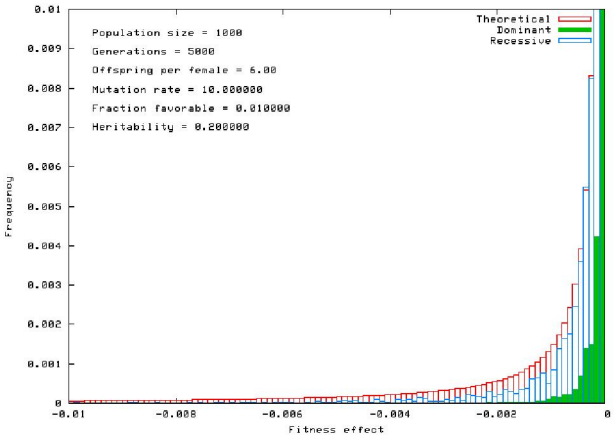

Figure 4. Distributions of accumulating mutations are shown above. Red bins represent the expected mutation accumulation when no selection is applied. Blue bins represent actual accumulation of recessive mutations. Green bins represent actual accumulation of dominant mutations. The magnitude of each mutation's effect is shown on the x-axis, which is a linear scale. The bin nearest zero represents mutations which change fitness by a factor between .0001 and .00001. Mutations with a magnitude of less than .00001 were not tracked or plotted.

4. Conclusions

The program Mendel's Accountant provides a biologically realistic platform for analyzing the problem of mutation accumulation. This program demonstrates that the problem of deleterious mutation accumulation is very serious under a wide range of scenarios and across a vast portion of parameter space. The relentless accumulation of deleterious mutations is primarily due to the existence of un-selectable "nearly- neutral" mutations, but the genetic load problem is greatly amplified when mutation rates are high. Intensified natural selection only marginally slows the accumulation of deleterious mutations. Preliminary Mendel experiments indicate that the most effective means of slowing mutation accumulation and reducing a population's genetic load is by reduction of the mutation rate. This study clearly indicates that more research is needed. Mendel's Accountant is freely available to users and can be downloaded at either http://mendelsaccountant.info or http://sourceforge.net/ projects/ mendelsaccount.

References

Santiago, E., & Cabellero, A. (2000). Application of reproductive technologies to the conservation of genetic resources. *Conservation Biology, 14*, 1831–1836.

Balloux, F. (2001). Computer Note- EASYPOP (Version 1.7): A computer program for population genetics simulations. *Journal of Genetics, 92*(3).

Crosby, J. L. (1973). *Computer simulation in genetics*. NewYork, New York: John Wiley.

Fraser, A., & Burnell, D. (1970). *Computer models in genetics*.New York, New York: McGraw-Hill.

Kimura, M(1979). *Model of effectively neutral mutations in which selective constraint is incorporated*. PNAS **76** : 3440-3444.

Kimura, M(1983). *Neutral Theory of Molecular Evolution*. Cambridge University Press, New York :30-31.

Sanford, J., Baumgardner, J., Brewer, W et al (2007). Mendel's accountant: a biologically realistic forwards time population genetics program**.** *Scalable computing practice and experience*;8(2): 147–165

9/16/2012