Using Computer Simulation to Understand Mutation Accumulation Dynamics and Genetic Load

John Sanford[1], John Baumgardner[2], Wes Brewer[3], Paul Gibson[4], and Walter ReMine[5]

[1] Dept. Hort. Sci., Cornell University, Geneva, NY, 14456, USA
`jcs21@cornell.edu`
[2] Los Alamos National Laboratory, Los Alamos, NM, USA, retired
[3] Computational Engineering, Mississippi State University, MS, USA
[4] Dept. Plant, Soil and Agric. Syst., Southern Illinois University, Carbondale, IL, USA
[5] Science and Math Dept., Northwestern College, St. Paul, MN, USA

Abstract. Long-standing theoretical concerns about mutation accumulation within the human population can now be addressed with numerical simulation. We apply a biologically realistic forward-time population genetics program to study human mutation accumulation under a wide-range of circumstances. Using realistic estimates for the relevant biological parameters, we investigate the rate of mutation accumulation, the distribution of the fitness effects of the accumulating mutations, and the overall effect on mean genotypic fitness. Our numerical simulations consistently show that deleterious mutations accumulate linearly across a large portion of the relevant parameter space. This appears to be primarily due to the predominance of nearly-neutral mutations. The problem of mutation accumulation becomes severe when mutation rates are high. Numerical simulations strongly support earlier theoretical and mathematical studies indicating that human mutation accumulation is a serious concern. Our simulations indicate that reduction of mutation rate is the most effective means for addressing this problem.

Keywords: genetic load, Mendel's Accountant, mutation accumulation, population genetics, simulation.

1 Introduction

The problem of genetic load has concerned geneticists for over 50 years [1][2]. Theoretically, high mutation rates and the natural inefficiencies of selection both appear to ensure the accumulation of deleterious mutations within the genomes of higher organisms [3]. These concerns have been accentuated by the apparent reduction of selection pressures within human populations within historical time frames [4]. All these earlier concerns were based upon purely theoretical considerations.

Advances in computer science and the increasing power of simulation programs provide us with a new way of understanding the problem of mutation accumulation. The use of numerical simulation allows us to test empirically previous mathematical analyses, which are otherwise inherently abstract and difficult to test.

Such simulations allow us to examine in precise detail complex biological scenarios which otherwise would require extreme simplification and generalization before any type of mathematical analysis would be possible.

2   The Program

The computer program "Mendel's Accountant" (hereafter referred to simply as Mendel) has been developed to provide a biologically realistic forward-time numerical simulation of mutation accumulation [5]. This is a highly flexible program which for the first time effectively models natural mutation distributions, environmental variance, and improved modeling of linkage/recombination. Mendel is designed to model sexually reproducing diploid organisms. Mendel tracks individual mutations in a detailed manner from parents to progeny through many generations. Mutations are modeled so as to have a continuous range of effects from lethal to beneficial and to vary in expression from fully dominant to fully recessive. Each mutation's unique identifier encodes its genotypic fitness effect, whether it is recessive or dominant, and its location within the genome (the specific linkage block where it resides within a specific chromosome). This allows realistic treatment of linkage of mutations along segments of chromosomes. Mutational effects may be combined either in a multiplicative or additive manner to yield an overall genotypic fitness for each new offspring.

Mendel is designed to track large numbers of distinct mutations by using a single four-byte integer word to store the mutation's unique identifier. This allows up to about four billion different unique mutations to be tracked in a given population. The number of mutations per individual is limited by the available processor memory and the population size before selection. As an example, 1.6 GB of memory available for storing mutation identifiers translates into a maximum of 40,000 mutations in each of 10,000 individuals. Thus, Mendel is effectively an infinite sites model, in contrast with k-allele and stepwise models that both impose highly restrictive limits on the number and variety of mutations. Mendel offers the option of tracking only those mutations whose fitness effect exceeds a user-specified threshold. This threshold usually is chosen to lie in that region of extremely small mutation effect which is beyond the reach of selection. Typically, we find that half to two-thirds of all mutations lie in this un-selectable region. The fitness effects of the untracked mutations under this option are nevertheless accounted for in the composite fitness effect of the linkage block where the mutation occurs. This option allows the user to investigate scenarios that involve two to three times more total mutations that would be possible otherwise.

Mendel offers the important option of including the effects of environmental variation. Environmental variance, specified via a heritability parameter and a non-scaling noise standard deviation, combines with genotypic fitness to yield the phenotypic fitness. Selection then acts on phenotypic fitness to eliminate that fraction of the offspring (the population surplus) required to maintain the user-specified population size. The surplus population is a consequence of the specified fertility, as implied by the average number of offspring per female. Mendel provides the user the

choice of either natural selection (probability selection) or artificial selection (truncation selection). Because Mendel is optimized for memory efficiency and speed, many non-trivial scenarios can be run on a desktop or laptop computer. Moreover, because Mendel is parallelized with MPI, it readily handles large population sizes and complex population substructure on cluster computers. Mendel's graphical user interface is designed to make the specification of a scenario intuitive and simple, while also providing a variety of visual representations of the program output. Mendel is therefore a versatile research tool. It is also useful as an interactive teaching resource for a variety of settings ranging from introductory courses in biology to more advanced courses in population genetics.

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. 1a 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 un-selectable "nearly-neutral" deleterious mutations (as further described below).
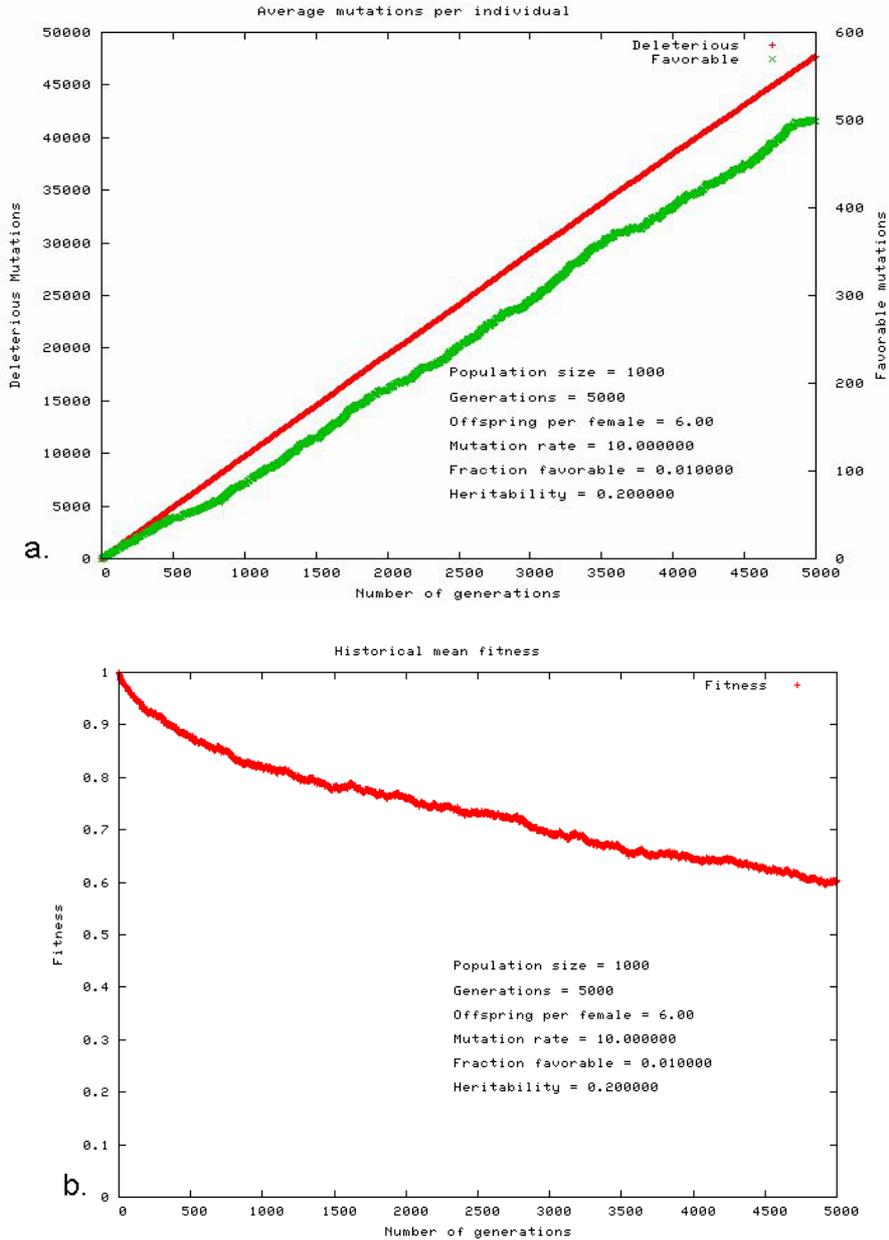
Fig. 1. (a) Mutation count per individual and (b) mean population fitness, plotted for 5,000 generations. (a) shows that deleterious mutations accumulate in close to a strict linear fashion (reaching 47,730–scale on left). Beneficial mutations also accumulate in a linear manner, but their lower number results in sampling error fluctuations (reaching 498– scale on right). (b) shows a progressive decline in population fitness.

Fig. 1b 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.

Fig. 2 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
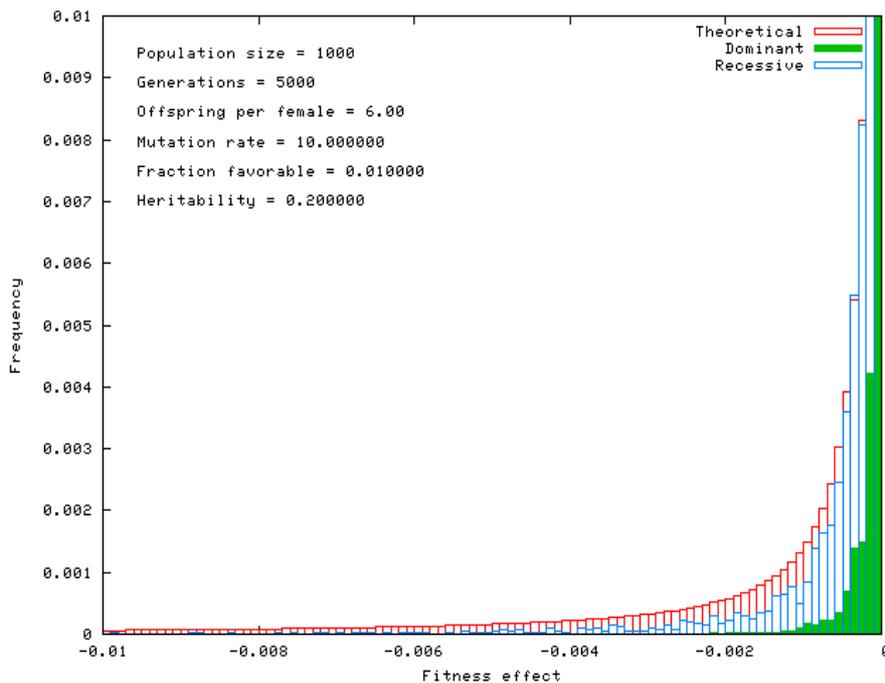

Fig. 2. Distributions of accumulating mutations are shown above. Red bins represent the expected mutation accumulation when no selection is applied. Blue bins represent actual accumulation of recessive mutations. Green bins represent actual accumulation of dominant mutations. The magnitude of each mutation's 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.

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 [6][7]. 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.

Fig. 3 shows that over time many alleles move toward fixation. The movement toward fixation is extremely slow for both deleterious and beneficial mutations – consistent with the mathematical predictions of Haldane [8]. However, over long periods of time, even with intense selection, a significant number of deleterious mutations consistently become fixed.

All these findings strongly support previous theoretical and mathematical analyses [1], [3], [9], [10] which have predicted that deleterious mutation accumulation in the human population is a very real biological concern.
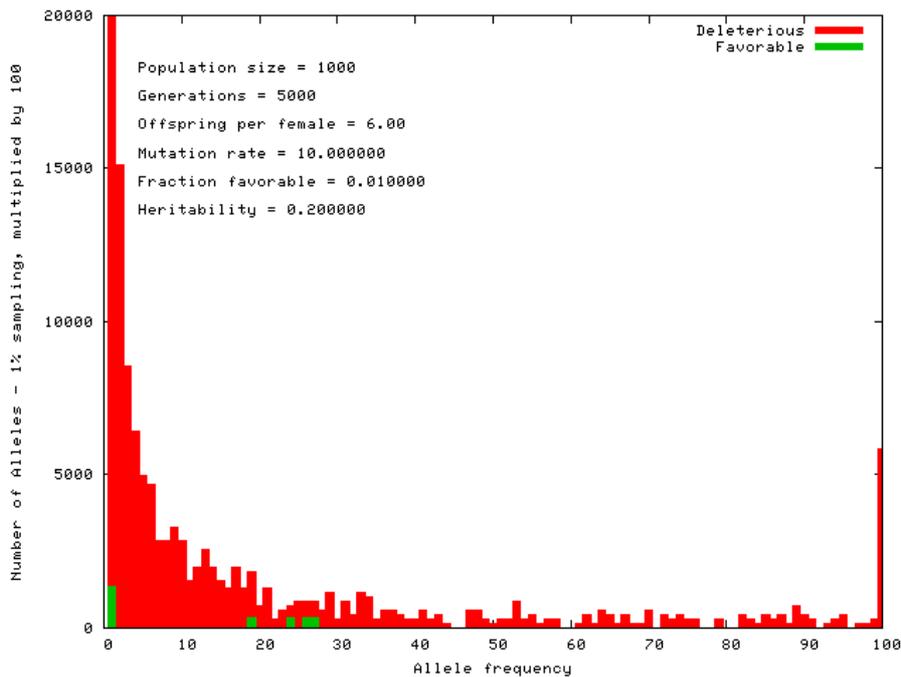
Fig. 3. Mutant allele frequencies are shown above, with rare alleles (<1%) on the far left, and fixed or nearly fixed alleles (>99%) on the far right. Deleterious mutations are shown in red, beneficial mutations are shown in green. In this instance 5,845 deleterious mutations have been fixed after 5,000 generations. No beneficial mutations were fixed in this example.

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

1. Muller, H.J.: Our load of mutations. Amer. J. Human Genetics **2** (1950) 111-176.
2. Wallace, B.: Fifty years of genetic load. J. Hered. **78** (1987) 134-142.
3. Kondrashov, A.S.: Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? J. Theor. Biol. **175** (1995) 583-594.
4. Crow, J.F.: The high spontaneous mutation rate: a health risk? PNAS **94** (1997) 8380-8386.
5. Sanford, J., Baumgardner, J., Gibson, P., Brewer, W., Remine, W.: Mendel's Accountant: a biologically realistic forward-time population genetics program. SCPE, **8(2)** (submitted).
6. Kimura, M.: Model of effectively neutral mutations in which selective constraint is incorporated. PNAS **76** (1979) 3440-3444.
7. Kimura, M.: Neutral Theory of Molecular Evolution. Cambridge University Press, New York (1983) 30-31.
8. Haldane, J.B.S.: The cost of natural selection. J. Genetics **55** (1957) 511-524.
9. Muller, H. J.: The relation of recombination to mutational advance. Mutation Research **1** (1964) 2-9.
10. Loewe, L.: Quantifying the genomic decay paradox due to Muller's ratchet in human mitochondrial DNA. Genetical Research **87** (2006) 133-159.