A Statistical Model of Genetic Imprinting

Alice Ohlson*
MIT Department of Physics
(Dated: May 14, 2009)

Parental imprinting is one epigenetic factor that can influence the expression of a gene in a diploid organism. The spread of a gene subject to imprinting through a population is compared to that of a gene unaffected by imprinting in a simulation. It is discovered that parental imprinting greatly affects the probability of fixation and the time to fixation of a new genetic mutation.

1. INTRODUCTION

In many statistical models of population genetics, it is assumed for simplicity that the phenotype of an organism is a function of its genotype alone. However, there are many additional natural factors which also play a role in phenotype determination and evolution, such as interactions with other species (coevolution), the environment, non-random mating, and epigenetic inheritance.[1] The following discussion will address parental imprinting, in which the phenotype of an organism is not solely based on its genotype, but rather on the genotypes of the organism's parents.

2. BIOLOGICAL BACKGROUND

Parental imprinting is an epigenetic factor that can influence gene expression in diploid organisms. In paternal imprinting, a gene is expressed only if it was inherited from the mother, and the copy inherited from the father is inactive. Maternal imprinting is the opposite. Imprinting is not due to changes in the DNA of an organism, rather it is due to the methylation of the inactivated gene.[2] This can affect the phenotype of a single organism in one generation dramatically. For example, an inherited recessive allele can be expressed as dominant because it is the only active copy of the gene in the organism. The question that will be addressed below is: Can imprinting dramatically affect inheritance patterns over the span of many generations?

3. MODEL WITH NO IMPRINTING

A model of gene inheritance was built in MATLAB[3], based on the structure of the model proposed by Fisher, Haldane, and Wright[1]. The simulation starts with two pools of alleles, one pool representing males and one females. (The 'male'/'female' designation is not important, it is just a convenient way of marking the two gene pools.) This is a two-allele model, so each pool is filled with a certain number of allele A_1 and of allele A_2. Each allele is marked with the phenotype of the organism to which it belongs.

Mating is simulated by picking one allele randomly from each of the male/female pools, with statistical weights determined by the relative fitness of the phenotype from which the alleles originated. The phenotype produced by the two alleles is determined and then half of the allele pairs are placed in the next-generation female pool and half in the next-generation male pool. The population size is kept the same from generation to generation for simplicity, and the gene pools are updated in parallel after each generation.

The fitness of the homozygote A_1A_1 is 1, the fitness of the homozygote A_2A_2 is $1+s$, and the fitness of the heterozygote is $1+hs$.

If allele A_2 is a random mutation, then the initial conditions of the model are: 1 male heterozygote, 99 male homozygotes, and 100 female homozygotes. The probability that the mutation will become fixed, or be present in all N individuals after some generations (with the loss of allele A_1), is given by Equation 1 (assuming that $h = 0.5$).[4]

$$\Pi = \frac{1 - e^{-s}}{1 - e^{-2Ns}} \tag{1}$$

The graph of the function in Equation 1 is shown in Figure 6.

4. MODEL WITH IMPRINTING

The model from Section 3 was modified to include imprinting. The only difference was the statistical weighting of genes in the gene pools. The weights were based on the phenotype of the organism, which was based on which allele was inherited from the mother (paternal imprinting). The fitness of the organisms expressing allele A_1 was 1, and the fitness of the individuals expressing A_2 was $1+s$.

If A_2 is a new mutation, then when a male has the first mutation, the initial conditions of the model are: 1 male heterozygote, 99 male homozygotes, and 100 female homozygotes (all expressing allele A_1, with fitness 1). When a female has the first mutation, the initial conditions are: 1 female heterozygote, 99 female homozygotes, and 100 male homozygotes (all expressing allele A_1, with fitness 1).

*Electronic address: aohlson@mit.edu

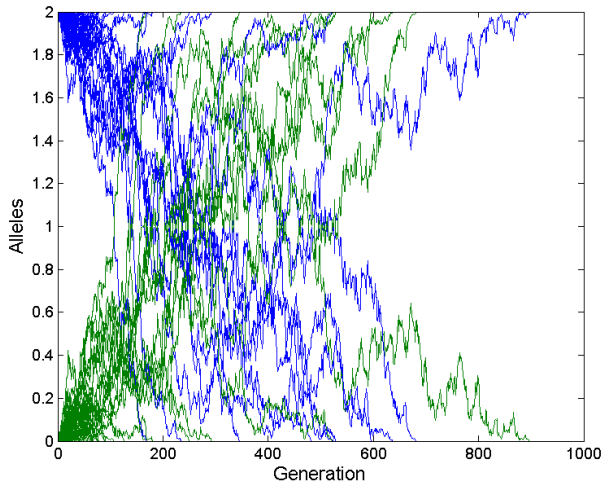

FIG. 1: No imprinting. In Figures 1 through 5 the blue lines represent the number of A_1 alleles and the green lines are the number of A_2 alleles, divided by the number of individuals in the population.

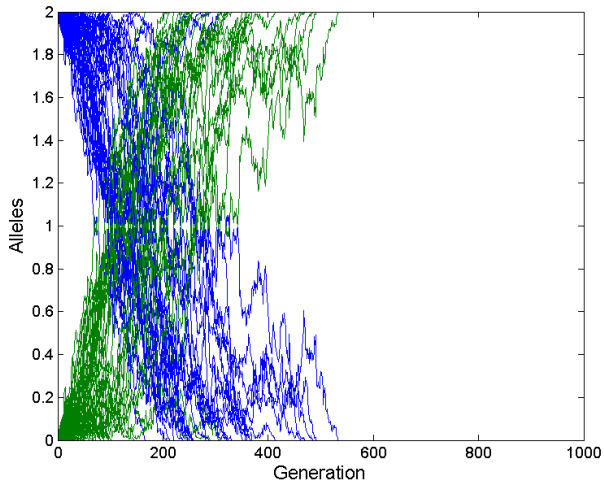

FIG. 2: Paternal imprinting, male has initial mutation.

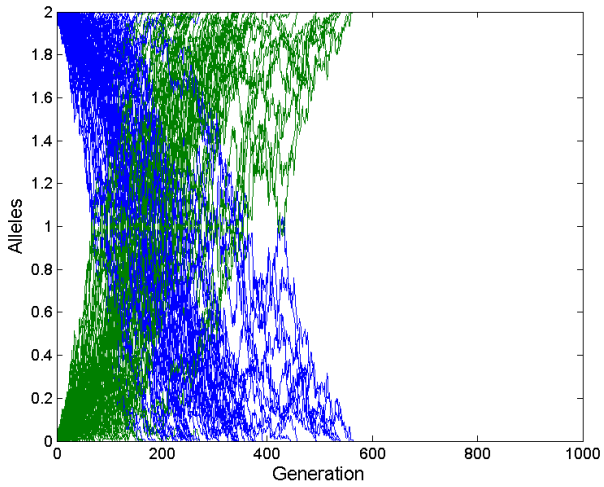

FIG. 3: Paternal imprinting, female has initial mutation.

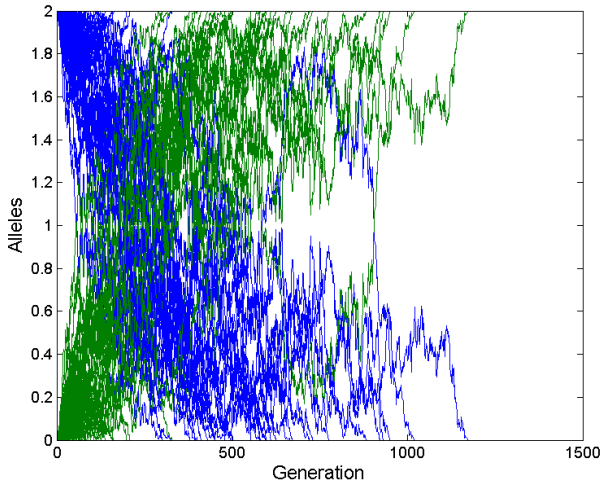

FIG. 4: No imprinting, dominant mutation.

5. MODEL COMPARISON

The simulations were run for the following parameters: $N = 200$, $s = 0.02$, $h = 0.5$. The results for the simulation with no imprinting are shown in Figure 1. Out of 1,000 trials, the mutation was fixed 9 times. The results for the simulation with paternal imprinting are shown in Figure 2, where the initial mutation is in a male. In 1,000 trials the mutation was fixed 28 times. Figure 3 shows the results of the simulation with imprinting, where the initial mutation is in a female. In 1,000 trials the mutation was fixed 52 times.

A value of $h = 0.5$ in the non-imprinting model indicates incomplete dominance (as in Figure 1). The behavior of allele frequencies in the case of a dominant new mutation ($h = 1$) is shown in Figure 4, in which the mutation is fixed 31 times out of 1,000. Figure 5 shows the non-imprinting model for a recessive mutation ($h = 0$), where the mutation is fixed 13 times out of 1,000.

6. RESULTS

The probability of fixation as a function of selection ($\Pi(s)$) was calculated for each of the scenarios discussed above: the non-imprinting model with $h = 0$ (recessive mutation), $h = 0.5$ (incomplete dominance), and $h = 1$ (dominant mutation), and the paternal imprinting model with the original mutation being in a female and in a male. Figure 6 shows that the results of each scenario are quite different. For negative values of s (indicating a

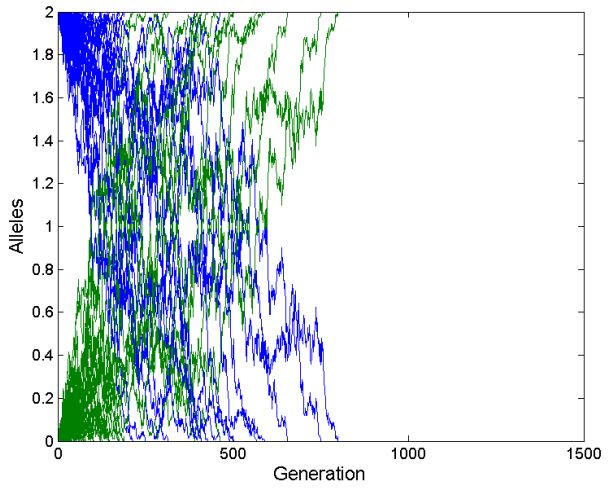

FIG. 5: No imprinting, recessive mutation.

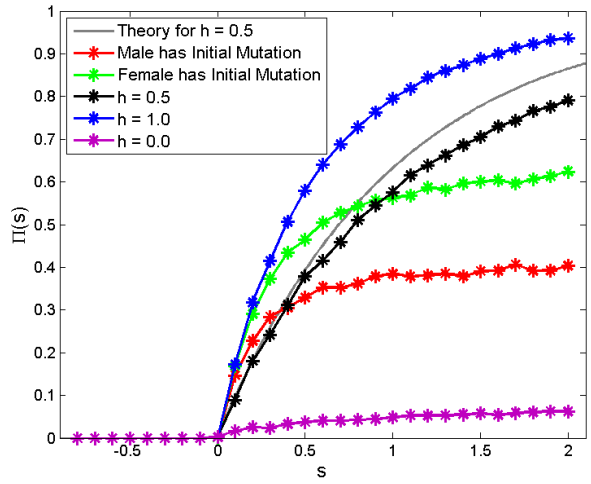

FIG. 6: Probability of fixation as a function of selection ($\Pi(s)$) for the no-imprinting model (with $h = 0, 0.5, 1$) and the imprinting model (for the initial mutation being in the non-imprinting and the imprinting gender).

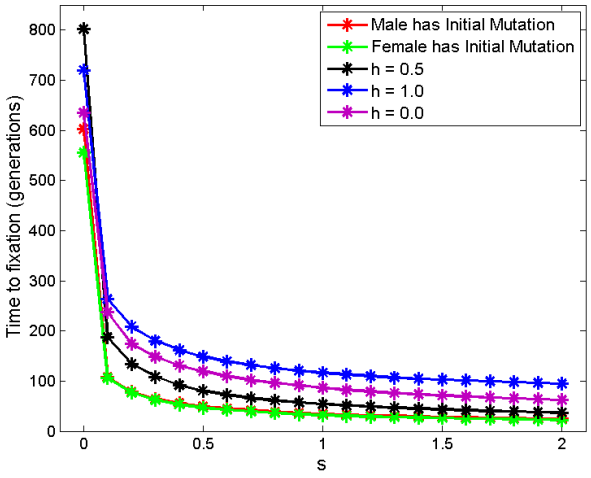

FIG. 7: Time to fixation of a random mutation as a function of selection.

deleterious mutation), all the models indicate near zero probability for the mutation to become fixed in the population. For small positive values of s (weak positive selection), the effect of imprinting is similar to that of a dominant mutation. For higher values of s (strong positive selection), the probability of fixation becomes roughly constant. This indicates that even for a very advantageous mutation, selection cannot overcome the power of imprinting. There is always a finite probability that a new mutation will be inherited mainly through the males, and will be lost before it has a chance to be expressed when inherited through the females (in the case of paternal imprinting), regardless of the value of s. However, there is less of a chance of this happening when the original mutation is in a female, and thus the probability levels off at a higher value when the original mutation is in a female.

The time to fixation (i.e. the number of generations for the mutation to become fixed, given that it does become fixed) was measured as a function of s and is shown in Figure 7. Mutations in genes with parental imprinting become fixed faster than mutations in genes without parental imprinting. Although in the non-imprinting model, the time to fixation varies with h, in the imprinting model the time does not depend on the gender in which the initial mutation occurred.

7. IMPRINTING IN HUMAN BIOLOGY

Parental imprinting was only discovered twenty years ago[2], and thus there is not yet any data on the behavior of mutations in imprinted genes over many generations. However, parental imprinting has already been implicated in a few human diseases, such as Prader-Willi syndrome and Beckwith-Wiedemann syndrome, and when the process fails it can cause several types of cancer.[5] Therefore the study of imprinting is an important one in modern genetics, and hopefully in the next few years the combination of more accumulated data and further theoretical work will lead to interesting results.

8. CONCLUSIONS

In order to study parental imprinting, two simulations were built according to the model proposed by Fisher, Haldane, and Wright. The first was a model of diploid genetic evolution without any epigenetic effects, and the second incorporated imprinting. The models were run for

several choices of parameters: the gender of the person in which the original mutation occurred, the dominance parameter h, and the selection coefficient s. The simulations indicate that the presence of parental imprinting can dramatically change the probability of fixation and time to fixation of a new mutation in a population.

[1] B. Drossel, Advances in Physics **50**, 209 (2001).

[2] A. J. Griffiths, S. R. Wessler, R. C. Lewontin, W. M. Gelbart, D. T. Suzuki, and J. H. Miller, *Introduction to Genetic Analysis* (W.H. Freeman and Company, New York, 2005), pp. 324–325, 8th ed.

[3] MATLAB 7.4.0, The Mathworks (R2007a).

[4] *8.592 Class Notes, MIT* (2009).

[5] M. Lalande, Annual Review of Genetics **30**, 173 (1996).