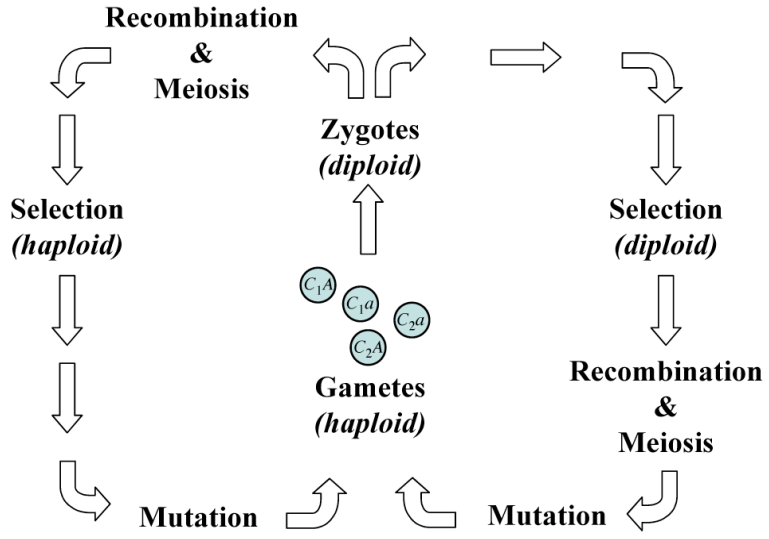
From Chapter 8 of: ***A Biologist's Guide to Mathematical Modeling*****, by Otto & Day.**

An important genetic distinction between some organisms is whether they are haploid (i.e., they carry only one copy of each gene) or diploid (i.e., they carry two copies of each gene). For example, humans are primarily diploid because the majority of their life cycle is carried out in the diploid state, even though they have a haploid stage (sperm and eggs). In contrast, many fungi, algae, and unicellular organisms are primarily haploid. Why have some organisms evolved to become diploid while others have evolved to become haploid? Let's construct a model to gain some insight into this question.

Consider an organism that reproduces sexually and is capable of growth and development in either the haploid or diploid stage. Because the organism is sexual, it will necessarily pass through a haploid stage after meiosis and a diploid stage after the union of gametes (Figure 8.11). To allow the proportion of time spent in each state to evolve, we suppose that there is a gene that alters the life cycle. At this gene, allele C_1 causes meiosis to occur early in life (before natural selection has acted), resulting in a predominantly haploid life cycle (left pathway, Figure 8.11). In contrast, allele C_2 causes a delay in meiosis until after selection, resulting in a predominantly diploid life cycle (right pathway, Figure 8.11). To simplify matters, we assume that C_2 is dominant, so that C_1C_2 individuals are also diploid.

Figure 8.11: Evolution of haploidy and diploidy. A general life cycle is shown in which an organism either undergoes meiosis early and experiences selection as a haploid (left pathway) or undergoes meiosis late and experiences selection as a diploid (right pathway).



Let's also suppose that there are no intrinsic costs or benefits to being haploid or diploid, and therefore the frequency of allele C_2 would remain constant over time if it were the only gene considered in the model. There are certainly other genes in the genome under selection, however, and we thus include a second gene under selection with two alleles, where allele *A* is the most fit but mutates regularly to a less fit allele, *a*, at rate μ. Specifically, we assume that the fitnesses of individuals are:

Haploid Fitness		Diploid Fitness		
A	*a*	*AA*	*Aa*	*aa*
1	$1 - s$	1	$1 - h\,s$	$1 - s$

where *h* is the coefficient of dominance and *s* is the selective disadvantage. Both *h* and *s* are assumed positive so that the *a* allele is deleterious. We also assume that, at the end of the life cycle, all individuals produce haploid gametes that unite at random to begin the next generation (Figure 8.11).

We census the population at the gamete stage immediately after mutation, letting the frequency of C_1A, C_1a, C_2A, and C_2a gametes equal x_1, x_2, x_3, and x_4, respectively (Figure 8.11). Using a mating table like Table 8.1, it is possible to track the frequency of these genotypes as they unite, undergo early meiosis (in C_1C_1 individuals), undergo selection, and then undergo late meiosis (in C_1C_2 and C_2C_2 individuals). The only difference from Table 8.1 is that C_1C_1 individuals experience selection only after meiosis. Working through such a table, the frequencies of the four chromosome types after meiosis but before mutation are:

$$\begin{aligned}
x_1' &= \left(x_1^2 + x_1 x_2 + x_1 x_3 + (1-r)\, x_1 x_4 (1 - h\,s) + r\, x_2 x_3 (1 - h\,s)\right)/\overline{W} \\
x_2' &= \left(x_1 x_2 (1-s) + r\, x_1 x_4 (1 - h\,s) + x_2^2 (1-s) + (1-r)\, x_2 x_3 (1 - h\,s) + x_2 x_4 (1-s)\right)/\overline{W} \\
x_3' &= \left(x_1 x_3 + x_1 x_4 r\,(1 - h\,s) + x_2 x_3 (1-r)\,(1 - h\,s) + x_3^2 + x_3 x_4 (1 - h\,s)\right)/\overline{W} \\
x_4' &= \left(x_1 x_4 (1-r)\,(1 - h\,s) + x_2 x_3 (1 - h\,s)\, r + x_2 x_4 (1-s) + x_3 x_4 (1 - h\,s) + x_4^2 (1-s)\right)/\overline{W}
\end{aligned} \tag{8.46a}$$

where *r* is the rate of recombination between the two genes, and $\overline{W}$ is the mean fitness of the population (the sum of the numerators on the right-hand side). Finally, we allow mutations to occur from *A* to *a* (mutations occurring in the reverse direction are assumed to be very rare and are ignored), which gives us the frequency of the four types of gametes in the next generation:

$$\begin{aligned}
x_1'' &= (1-\mu) x_1' \\
x_2'' &= \mu\, x_1' + x_2' \\
x_3'' &= (1-\mu) x_3' \\
x_4'' &= \mu\, x_3' + x_4'
\end{aligned} \tag{8.46b}$$

If we want to understand how life cycles evolve, one approach is to assume that the population has a certain life cycle and then determine when a mutation altering the life cycle can invade. To this end, let's consider a haploid population, with allele C_1 fixed, into which we will introduce the C_2 allele, thereby generating diploid individuals.

First, we must determine the equilibrium reached by a haploid population when only allele C_1 is present ($x_3 = x_4 = 0$). In Problem 8.3, you are asked to find the equilibrium where selection is balanced by mutation. This equilibrium occurs at $\hat{x}_1 = 1 - \frac{\mu}{s}$ and $\hat{x}_2 = \frac{\mu}{s}$, as we found in the one-gene model (Problem 5.4).

Next, let's find out what happens after C_2 arises at some small frequency, causing its carriers to remain diploid throughout selection. The allele will ultimately decrease in frequency if the equilibrium $\hat{x}_1 = 1 - \mu/s,\ \hat{x}_2 = \mu/s,\ \hat{x}_3 = 0,\ \hat{x}_4 = 0$ is locally stable, and the population will remain haploid. Therefore, we can determine when we expect the allele to die out by performing a local stability analysis of this equilibrium. The stability matrix for this model is

$$\mathbf{J} = \left(\begin{array}{cccc} \frac{\partial x_1''}{\partial x_1} & \frac{\partial x_1''}{\partial x_2} & \frac{\partial x_1''}{\partial x_3} & \frac{\partial x_1''}{\partial x_4} \\ \frac{\partial x_2''}{\partial x_1} & \frac{\partial x_2''}{\partial x_2} & \frac{\partial x_2''}{\partial x_3} & \frac{\partial x_2''}{\partial x_4} \\ \frac{\partial x_3''}{\partial x_1} & \frac{\partial x_3''}{\partial x_2} & \frac{\partial x_3''}{\partial x_3} & \frac{\partial x_3''}{\partial x_4} \\ \frac{\partial x_4''}{\partial x_1} & \frac{\partial x_4''}{\partial x_2} & \frac{\partial x_4''}{\partial x_3} & \frac{\partial x_4''}{\partial x_4} \end{array} \right) \Bigg|_{\substack{x_1 = 1 - \frac{\mu}{s}, x_2 = \frac{\mu}{s}, \\ x_3 = 0, x_4 = 0}} \tag{8.47}$$

where the x_i'' are given by equations (8.46). Analyzing this matrix is aided by the fact that the 2×2 sub-matrix on the bottom left is full of zeros: $\left(\begin{array}{cc} \frac{\partial x_3''}{\partial x_1} & \frac{\partial x_3''}{\partial x_2} \\ \frac{\partial x_4''}{\partial x_1} & \frac{\partial x_4''}{\partial x_2} \end{array} \right) \Bigg|_{\substack{x_1 = 1 - \frac{\mu}{s}, x_2 = \frac{\mu}{s}, \\ x_3 = 0, x_4 = 0}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Thus, matrix (8.47) has a "block triangular form" (see Primer 2), and its four eigenvalues are given by the eigenvalues of the two sub-matrices along the diagonal, which we label **A** and **B**:

$$\mathbf{A} = \left(\begin{array}{cc} \frac{\partial x_1''}{\partial x_1} & \frac{\partial x_1''}{\partial x_2} \\ \frac{\partial x_2''}{\partial x_1} & \frac{\partial x_2''}{\partial x_2} \end{array} \right) \Bigg|_{\substack{x_1 = 1 - \frac{\mu}{s}, x_2 = \frac{\mu}{s}, \\ x_3 = 0, x_4 = 0}} = \begin{pmatrix} \frac{\mu(1-s)}{(1-\mu)s} & -\frac{(s-\mu)(1-s)}{(1-\mu)s} \\ -\frac{\mu(1-s)}{(1-\mu)s} & \frac{(s-\mu)(1-s)}{(1-\mu)s} \end{pmatrix} \tag{8.48a}$$

$$\mathbf{B} = \left.\begin{pmatrix} \frac{\partial x_3''}{\partial x_3} & \frac{\partial x_3''}{\partial x_4} \\ \frac{\partial x_4''}{\partial x_3} & \frac{\partial x_4''}{\partial x_4} \end{pmatrix}\right|_{\substack{x_1 = 1-\frac{\mu}{s}, x_2 = \frac{\mu}{s}, \\ x_3 = 0, x_4 = 0}} = \begin{pmatrix} 1 - \frac{\mu(r + hs - rhs)}{s} & \frac{(s-\mu)(1-hs)r}{s} \\ \frac{\mu(r+s-rhs) - \mu^2(r+hs-rhs)}{(1-\mu)s} & 1 - \frac{(s-\mu)(hs + (1-\mu)(1-hs)r)}{(1-\mu)s} \end{pmatrix} \quad (8.48b)$$

(Rule P2.27).

From the partial derivatives contained in the sub-matrix **A**, we can tell that this sub-matrix describes the sensitivity of the recursions for x_1 and x_2 to displacements in x_1 and x_2. The eigenvalues of this sub-matrix thus describe the stability of the equilibrium in the absence of the new mutant C_2 allele. Using Rule P2.20 and factoring, these eigenvalues are given by $\lambda = 0$ and $\lambda = \frac{1-s}{1-\mu}$. The zero eigenvalue indicates that the model has only three effective dimensions, and again this occurs because all four gamete frequencies must sum to one. The eigenvalue of $\frac{1-s}{1-\mu}$ is positive and less than one as long as the mutation rate is small relative to selection ($\mu < s$), which is both a reasonable assumption and necessary for $\hat{x}_1 = 1 - \frac{\mu}{s}$ to be a valid equilibrium frequency. Thus, when the C_2 allele is absent, the mutation-selection balance equilibrium is stable when it exists.

The real question of interest is whether this equilibrium is stable if we perturb it by introducing the C_2 allele (i.e., if we have x_3 and/or x_4 not equal to zero). As seen by the partial derivatives contained in sub-matrix **B**, the eigenvalues of **B** address this question. These eigenvalues can be calculated, but they're ugly and difficult to interpret. To obtain interpretable results, we use the perturbation method under the assumption that the mutation rate is very small, $\mu = \zeta$.

First, let's find λ_0, the leading term in the eigenvalue (8.45), by setting the mutation rate to zero in sub-matrix **B**. The sub-matrix is then triangular and has eigenvalues equal to the diagonal elements; $\lambda_0 = 1$ and $\lambda_0 = (1-hs)(1-r)$. As we have assumed that all of the parameters are positive, $\lambda_0 = 1$ is the leading eigenvalue (i.e., the one with the largest absolute value).

Because the leading term, $\lambda_0 = 1$, falls on the border between stability and instability, we must seek out the next order term, λ_1, in (8.45). Plugging (8.45) in for λ in the characteristic polynomial of **B**, taking the Taylor series in ζ and keeping only terms to first order in ζ, we get an equation that λ_1 must satisfy:

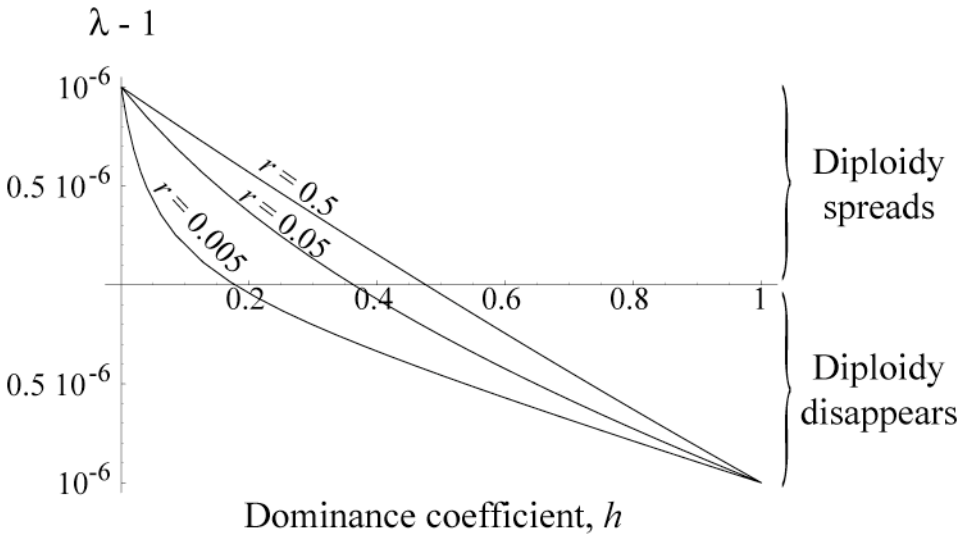
$$-\mu\left(r(1-2h)(1-hs) - h^2 s - \lambda_1(r+s-rhs)\right) = 0, \quad (8.49)$$

which we have written in terms of the original parameter, μ (see Problem 8.9). Solving (8.49) for λ_1 and plugging λ_0 and λ_1 into (8.45), we get an approximation for the leading eigenvalue:

$$\lambda = 1 + \frac{r(1-2h)(1-hs) - h^2 s}{r + s - rhs}\mu + O(\mu^2), \tag{8.50}$$

which is extremely close to the exact numerical value (Figure 8.12). The denominator of (8.50) is always positive, but the numerator is more difficult to interpret. If $h > 1/2$ then all terms in the numerator are negative and λ is less than one, indicating that diploids cannot invade a haploid population. If $h < 1/2$, then the two terms in the numerator are of opposite sign, and diploidy can invade ($\lambda > 1$) as long as the first term is larger. This requires that $r > \frac{h^2 s}{(1-2h)(1-hs)}$.

Figure 8.12: Evolution of diploidy. The leading eigenvalue minus one ($\lambda - 1$) is shown as a function of the dominance coefficient. A mutation causing diploid life cycles spreads within a haploid population only when the curve lies above the horizontal axis. This requires that h is small enough ($\lambda - 1 > 0$); the larger the recombination rate between the genes, the broader the range of dominance conditions under which diploidy evolves. For the parameters chosen (s = 0.1, $\mu = 10^{-6}$), the curves for the leading eigenvalue based on the approximation (8.50) or a numerical evaluation of the eigenvalues of the matrix (8.48a) cannot be distinguished.



What do these results mean biologically? Intuitively, one might expect diploidy to be favored because mutant alleles can be "masked" by the good copy of the allele in heterozygotes. This does provide diploidy with an advantage, but this advantage is counter-balanced by the fact that a diploid individual has two chances of carrying a deleterious mutation. Thus, only when

masking is strong enough ($h < 1/2$) would the average fitness of diploids be better than haploids, assuming that they have the same frequency of deleterious alleles.

The final twist to our result, however, is that diploidy spreads only if recombination rates are high enough, even when mutations are better masked in diploid individuals ($h < 1/2$). Because of masking among diploid ancestors, mutant alleles are more likely to survive and persist among the descendants carrying the diploid allele, C_2. Consequently, chromosomes with the diploid allele, C_2, are more loaded with deleterious mutations, while chromosomes with the haploid allele, C_1, are more effectively purged of deleterious mutations. The tighter the recombination rate between the ploidy and selected genes, the greater the difference in mutant allele frequency expected between C_1–bearing and C_2–bearing chromosomes, and the less likely diploids are to invade a population. While diploidy protects the individual from its burden of deleterious mutations, it does so at the expense of future generations, which are more likely to inherit deleterious mutations. Thus, diploidy is favored only when there is enough genetic mixing among the chromosomes in a population (see also Otto 1994; Otto and Goldstein 1992; Otto and Marks 1996; Perrot et al. 1991).

REFERENCES:

Otto, S. P. 1994. The role of deleterious and beneficial mutations in the evolution of ploidy levels. Lectures on Mathematics in the Life Sciences 25:69-96.

Otto, S. P., and D. B. Goldstein. 1992. Recombination and the evolution of diploidy. Genetics 131:745-751.

Otto, S. P., and J. Marks. 1996. Mating systems and the evolutionary transition between haploidy and diploidy. Biological Journal of the Linnean Society 57:197-218.

Perrot, V., S. Richerd, and M. Valero. 1991. Transition from haploidy to diploidy. Nature 351:315-7.