6.877 Lecture Notes, 11/30/05: the cost and benefits of sex

1. The cost of sex: why sex?

There is a two-fold cost to sexual reproduction in stable populations (diploid). Each female, whether sexual or not, will leave behind, on average, two offspring. On average, in sexual species, one of the two offspring will be male, and the other female. Each of these offspring carries one haploid complement of the female's genetic material; the other haploid genome come from the male parent. A parthenogenetic female will also leave behind two offspring, but both will be female because they are clones of their mother. Moreoever, all of the genetic material, four haploid complements, come from the mother. Thus, the parthenogenetic female leaves behind <u>twice as much</u> of her genetic material as does the sexual material.

Imagine a sexual species in which a dominant mutation arises that causes such parthenogenesis. The genome carrying the mutation will appear in two individuals in the next generation, four after that, and so on, doubling in number each generation. Meanwhile, all genomes without the mutation will be replacing themselves once every generation, on average. Eventually, the species will be made up entirely of parthenogenetic individuals. This is the two-fold cost of sexual reproduction.

There must be some problem with this scenario or else all species would be parthenogenetic. One possibility is that mutations to parthenogenesis either do not occur or else are so deleterious that they swamp their two-fold advantage. However, in many groups of species parthenogenesis does recur, so we must look elsewhere for an explanation of its ultimate failure.

Sex in diploids is associated with both segregation and meotic crossing over; parthenogenetic species have neither. A natural place to look for the advantage of sex is in the examination of segregation and recombination. Both can confer advantages to a species. Let's see how.

2. Two possible explanations for overcoming the two-fold cost of sex

2.1 Kirkpatrick's Rocket: segregation can overcome the two-fold cost

A 1989 paper by Kirkpatrick and Jenkins builds a model showing that the substitution of an advantageous mutation for a less favorable form requires two mutations in a lineage, whereas substitution in a sexual species requires only one mutation. If the evolutionary success of a species depends on its ability to evolve rapidly, and if advantageous mutations are relatively rare, then it is easy to imagine that the parthenogens will be at a distadvantage.

This is plausible, but the argument isn't convincing unless it can be shown to enhance the fitness of sexual species sufficiently to overcome their two-fold cost of sex. The Kirkpatrick model did this by first finding the number of loci in a parthenogenetic species that are waiting for a second mutation that allows the fixation of the advantageous

mutation and then calculating the cost of being heterozygous at these loci. Here's how their model works.

We assume that the fixation of an advantageous mutation in a parthenogen (asexual species) occurs in two steps, as illustrated below. In the beginning, a typical locus is imagined to be fixed for the A_2 allele, with the fitness of the genotypes as follows:

Genotype:	A_2A_2	A_1A_2	A_1A_1
Fitness:	1	$1+hs$	$1+s$

The A_1 allele is favored, so $s>0$. The symbol h denotes the usual heterozygote factor that diminishes the fitness of the heterozygote with respect to the homozygote A_1A_1 form, so $h<1$. The rate of mutation to the A_1 allele is u. The population size is N. So, each generation, $2Nu$ A_1 mutations enter the population, on average. However, as we have earlier shown from the diffusion approximation, only some of these will fix, due to stochastic effects; you may recall that the fixation probability is approximately $2s$, or $2hs$, if there is a heterozygous effect. Thus the rate of entry of A_1 alleles into the population is:
   $4Nush$
If there are L homozygous loci experiencing this sort of selection, then the rate of conversion from homozygous A_2A_2 forms to heterozygous A_1A_2 forms is:
   $4NuLsh$
The environment is assumed to be changing in such a way that L is constant rather than decreasing with the substitution of each advantageous mutation.

Now we imagine the second stage in this process, the conversion of A_1A_2 heterozygotes to A_1A_1 homozgyotes (which are the most fit forms). We will again do this by introducing a diffusion approximation, and then find the steady-state point at which the rate of flow producing A_1A_2 heterozygotes is equal to the conversion rate to A_1A_1 homozygotes. By calculating the fitness of this mixed heterozygote/homozygote population and comparing to the most fit situation, we will see how much more fit a sexual species (which does not have to use this two-stage approach) can be than an a-sexual species.

For the second stage then, we require the fixation of an A_1 allele on the chromosomes that have an A_2 allele, in the heterozygotes. The mean number of A_1 mutations entering each generation at <u>one</u> locus is $2Nu$ (rather than $4N$ this time, because each individual we are talking about is an A_1A_2 heterozygote). The survival probability of these mutations is, by the diffusion approximation as before, twice the selective advantage of A_1A_1 over A_1A_2 . This selection coefficient of an A_1A_1 homozygote <u>relative</u> to the A_1A_2 heterozygote, call it $s\,'$, can be found by solving:

$$1+s'=\frac{1+s}{1+hs}, \text{ so}$$

$$s'=\frac{s(1-h)}{1+hs}$$

So, the <u>rate</u> of conversion of a particular locus from a population of heterozygotes to one of homozygotes is:

$$\frac{2Nus(1-h)}{1+hs}$$

If n loci are currently heterozygous (at some point in time), then the total rate of conversion from A_1A_2 to A_1A_1 for the whole genome is:

$$\frac{2Nuns(1-h)}{1+hs}$$

The equilibrium or steady-state number of heterozygous loci is found by setting the rate of conversion to heterozygotes (the number flowing in to the heterozygous A_1A_2 pool from the A_2A_2 pool) equal to the rate of conversion from A_1A_2 to A_1A_1, and then solving for n:

$$4NuLs = \frac{2Nu\hat{n}s(1-h)}{1+hs}, \text{ or}$$

$$\hat{n} = \frac{2Lh(1+hs)}{1-h}$$

We now show that this induces a 'fitness' load on a heterozygous/homozygous mixed population that does not arise in the sexual case. In a sexual species, the A_1 allele sweeps through the population without waiting for a second mutation: it can get two A_1 alleles by segregation, from another individual. The asexual species is at a relative disadvantage while waiting for the second mutation to sweep through. How much of a disadvantage is this? We can calculate it as follows. The fitness contribution of each of the $\hat{n}$ heterozygous loci to the particular individual's overall fitness is $1+hs$. If the overall fitness is the product of the fitnesses of individual loci (multiplicative gene interaction or epistasis – remember, that multiplicative epistasis says that the effects of genes on an individual's survival probability are independent.), then the total fitness contribution of $\hat{n}$ heterozygous loci is $(1+hs)^{\hat{n}}$. The relative advantage of the sexual species is thus:

$$W_s = \left(\frac{1+s}{1+hs}\right)^{\hat{n}}$$

This is the benefit of sex. When does this beat the two-fold cost of sex? Suppose, e.g., that $s=0.01$, $L=100$, $h=1/2$, then the steady-state value $\hat{n}$ can be computed as 201, and the W_s value works out to 2.7, which is sufficient. (This is a pretty high selection coefficient of 1%; if $s=0.001$, .1%, then we need more than 100 loci involved – in fact, you should be able to work out that 693 different loci need to be selected in this fashion. You might want to reflect on the biological plausibility of this number.)

2.2 Kondrashov's Hatchet: benefits of recombination can over-come the two-fold cost of sex

We saw in an earlier lecture that sex might help by recombination: it allows for the removal of deleterious mutations that otherwise accumulate in the "Muller's ratchet" fashion. One problem with that model is that it assumed multiplicative epistasis, which is not clearly seen in actual data. It is contradicted by a number of experimental studies. In one example, T. Mukai (1968) accumulated mutations on the second chromosome of *Drosophila melanogaster* for 60 generations, under conditions that minimized natural selection. As the mutations accumulated, the relative viability of the second chromosome when homozygous decreased (viability as measured by survival of the adult flies). Mukai estimated that the mutation rate to deleterious mutations with measurable effects was

0.1411 per second chromosome per generation. In 60 generations, a typical second chromosome would have 60 x 0.1411 = 8.46 mutations. We can graph # of mutations against viability and get a scatter plot, then see whether multiplicative epistasis fits this data. (The rightmost point would correspond to 8.46 mutations, on the x-axis.)

Image removed due to copyright reasons.

When we look at this figure, we can see that there are three basic types of curves we can try to fit, corresponding to three models of epistasis. A convex curve is characteristic of multiplicative epistasis, of the form:

$$w_n = (1-s)^n$$

where n is the number of deleterious mutations, s is a measure of the effect of each mutation, and w_n is the relative viability of a fly with n homozygous deleterious mutations.

A straight line is characteristic of additive epistasis, where each gene chips in an incremental amount that accumulates by a constant with each new mutation:

$$w_n = 1-sn$$

A concave curve is characteristic of synergistic epistasis, a quadratic model with:

$$w_s = 1 - sn - an^2$$

This curve fits the data much better than additive or multiplicative epistasis. With synergistic epistasis, fitness drops off faster than expected from the effects of mutation in isolation – things go 'from bad to worse.' This does not invalidate Muller's ratchet, but it does change its rate. What is worse, though, is that it violates the assumptions on which we built the model for the ratchet – a Poisson distribution of the number of mutations per genome.

To remedy this, Kondrashov (1988) proposed a direct model of synergistic epistasis, and showed that this alone could be used to demonstrate that sexual recombination could eliminate deleterious mutations better than asexually reproducing species could. Let's see how this model works. Kondrashov proposed a model of truncation selection where the number of deleterious mutations can accumulate, up until some threshold k; before

the threshold is reached, there is no effect; with *k* or more mutations, the organism dies – the hatchet falls.

Intuitively, the difference between asexual and sexual populations will then be as follows. In asexual populations, as before in the Muller's ratchet model, deleterious mutations can only accumulate more and more, until they reach the threshold *k*. Then, one more mutation will kill individuals, and of course there is still some chance that individuals will get 0 new mutations, and survive. In this situation, all individuals are 'crowded' in a peak at *k* and the variance of the population is quite small. Thus, if there is, say, a 50% chance of a new mutation, ½ of the population will die (since ½ will move from *k* to *k*+1 mutations). In contrast, consider a sexually recombining population. In this case, by recombination, it is possible for a new individual to move from *k* to *k*-1 mutations. Below, the black dots are deleterious mutations along the two chromosomes and the dotted lines are the cross-over points, resulting in one new chromosome with *fewer* deleterious mutations than before (2 instead of 3):
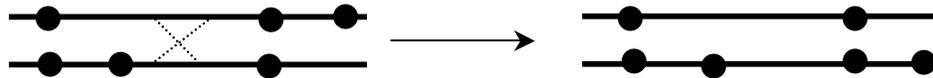
Thus, overall, in a sexual species the total probability mass can be distributed much more broadly, over the full range of 0, 1, 2, …, *k*–1 mutations. If we now apply a chance of one more deleterious mutation, this will fall not on the entire population as before, but only on a much smaller fraction, say, 10%. Thus, the selective death of the sexual population will be much less, compared to the asexual one – its overall fitness is higher. It is this selective advantage that the Kondrashov model can quantify, and we shall see that it can easily be much larger than the 2-fold disadvantage of sex. We can summarize the model via the following picture:
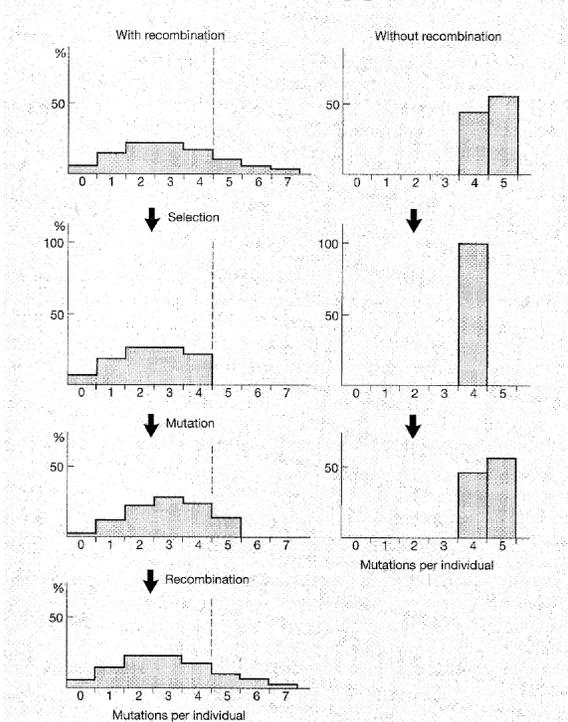
To pursue the model then. We start by considering an asexual population. Mutation is added as with the Muller's ratchet case: each generation, every offspring receives a Poisson-distributed number of new deleterious mutations. The mean of the Poission is u. Through time, the number of deleterious mutations in any asexual lineage will increase until it reaches k. From then on, any offspring with additional mutations will die. After a sufficient period of time, every individual in the population will have exactly k mutations. From this point on, the frequency of offspring in the population with <u>no</u> additional mutations is just the 0th order term in the Poisson distribution with mean u, i.e., e^{-u}. Therefore, the probability that an offspring has one or more mutations (killing it) is $1-e^{-u}$.

At steady state, the mean fitness of the asexual population is thus:
$$\bar{w} = e^{-u} \times 1 + (1 - e^{-u}) \times 0 = e^{-u}$$
On the assumption that 1 denotes 'the maximum fitness, and anything less is worse, we thus have that the 'selectional load' L, or the maximum fitness minus the actual fitness, relative to the maximum fitness, is $1-e^{-u}$. For example, if the (Poisson) mean number of new mutations is 2, then the load is $1-e^{-2}$, or 0.86.

Now consider a sexually reproducing population. (It is perhaps easiest to imagine an asexual population turning sexual.) With meiosis and crossing over, individuals would appear that have fewer than k deleterious mutations. They along with those with k mutations, would live, while those with more than k mutations would die. The result is that the mean number of deleterious alleles would be less than in the asexual population. With sex, selection lowers the mean number of mutations per individual, while mutation increases the number. Eventually, steady state is reached where the selectional load must be lower than in the asexual case – let us see where the balance is.

Kondrashov's hatchet kills off all offspring with more than k mutations, which reduces the average number of deleterious mutations per individual by an amount that is by definition the selectional differential, S, where S is the mean value of the phenotype of individuals *after* selection. That is, S measures the amount by which selection changes the some property of the individuals, in this case, their mean number of mutations – it lowers this, as remarked. (If this were some other quantitative trait, say, number of bristle hairs on a *Drosophila*, or height, then S would measure how much selection 'pushes' that value in one generation.) At steady state, $S=u$; that is, the mean number of mutations being removed by selection balances the mean number being introduced. What is the mutational load at this balance point? We want to know if this is sufficiently smaller than the asexual fitness 'load' to overcome the cost of sex.

To compute this, we need one formula that we have not yet derived, which is called the <u>intensity of selection with respect to the **proportion** of the population selected p, $i(p)$.</u>

This turns out to be: S/std-deviation(total # of mutations), or $\dfrac{S}{\sqrt{V_P}}$, where $\sqrt{V_P}$ is the variance in the trait being measured, in this case, the number of deleterious variations. We'll derive that formula just below, but let's proceed. Note that if $u=S$, then, dividing both sides by $\sqrt{V_P}$, we have:

$$i(p) = \frac{u}{\sqrt{V_P}}$$

which is the ratio of the mean number of mutations divided by the standard deviation of the total number of mutations per individual (the spread of the mutation distribution). This has been measured. In the case of *Drosophila*, Kondrashov argued that $u=2$ and $\sqrt{V_P} = 10$, which means that $i(p) = 0.2$. If we do some numerical computations relating $i(p)$ back to the actual proportion of individuals that must be 'killed off' (see below), we find that this intensity occurs when about 10% of the population is killed off in each generation. This is the 'selectional load' of a sexual population at steady-state. Recall now that the load for the asexual population with the same mutation rate $u=2$ was 0.86. The ration of sexual to asexual mean fitness for each is one minus these loads, so is roughly 0.9/0.14, or 6.4, easily larger than the 2-fold cost of sex. If $u=1$, then the benefit of sex also exceeds the cost. This is another example, then, where we can plausibly show that the two-fold cost of sex might be countered by another force.

Derivation of intensity of selection
To derive the formula for the intensity of selection, we assume that the parental phenotype follows a normal distribution (e.g., the distribution of number of deleterious alleles is normally distributed). So we get a picture like this, where α represents the proportion of parents that are selected. The mean midpoint of selection parents, as shown in the figure is S, the selection differential.
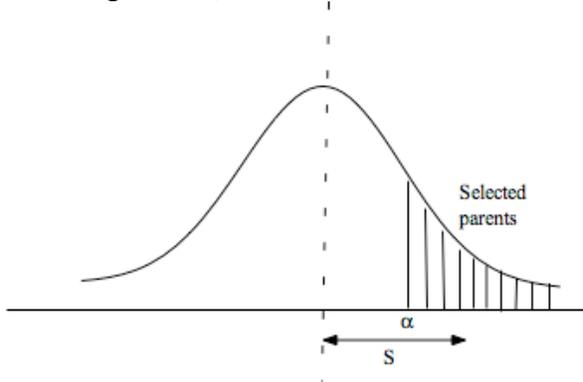
If we denote by V_P the variance of the parental phenotypes, then the probability density of the parents with mean 0 is this:

$$\frac{1}{\sqrt{2\pi V_P}} \exp\left(\frac{-x^2}{2V_P}\right)$$

The proportion p that is selected is the hatched area from α on, or:

$$p = \int_\alpha^\infty \frac{1}{\sqrt{2\pi V_P}} \exp\left(\frac{-x^2}{2V_P}\right) dx$$

where α is the largest number that is less than the values of the selected parents. The proportion may be simplified by changing the variable of integration to $y = \frac{x}{\sqrt{V_P}}$, yielding:

$$p = \int_{\alpha/\sqrt{V_P}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

The selection differential is the mean of the selected parents, the mean of the parents in the shaded area in the figure. The density of the selected parents (parents *can* be dense), is the truncated normal density,

$$\frac{C}{\sqrt{2\pi V_P}} \exp\left(\frac{-x^2}{2V_P}\right), x > \alpha$$

where C is chosen to make the integral of the density equal to one,

$$C = \left(\int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi V_P}} \exp\left(\frac{-x^2}{2V_P}\right) dx\right)^{-1}$$

The mean of the selected parents is thus

$$S = C \int_{\alpha}^{\infty} \frac{1}{\sqrt{2\pi V_P}} \exp\left(\frac{-x^2}{2V_P}\right) dx$$

We simplify this by changing the variable of integration as before to $y = \dfrac{x}{\sqrt{V_P}}$, obtaining:

$$S = \sqrt{V_P} i(\alpha / \sqrt{V_P}), \text{ where}$$

$$i(\alpha / \sqrt{V_P}) = \frac{\int_{\alpha/\sqrt{V_P}}^{\infty} \frac{1}{\sqrt{2\pi}} y e^{-y^2/2} dy}{\int_{\alpha/\sqrt{V_P}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy}$$

The function i is called the intensity of selection. Note that both this and the proportion selected are functions of the quantity $\dfrac{\alpha}{\sqrt{V_P}}$. They are both monotonic functions of this quantity; p is a decreasing function and S is an increasing function. Consequently, for each value of p there is a unique corresponding value of the intensity of selection, and we can view the intensity of selection as function of p rather than as a function of $\dfrac{\alpha}{\sqrt{V_P}}$.

Alas, the functional relation cannot be written as a simple formula because both p and $i(\dfrac{\alpha}{\sqrt{V_P}})$ are functions of integrals. However, it is easy to evaluate the integrals by computer and so obtain the figure below. This figure relates the intensity of selection to the proportion of selected individuals. The selection differential is then obtained as $S = \sqrt{V_P} i(p)$, which is what we were after. (This formula assumes that the selection differential uses a large number of measured individuals.)

Image removed due to copyright reasons.

Intensity of selection (y-axis) as a function of $p,$ proportion of individuals selected