A mutant hunt in a model vertebrate

Paper to read for this section

Mullins, M.C. *et al.* (1994) Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Current Biology* **4**: 189-202

Zebrafish as a genetically usable model system

As discussed in Chapter 1, different model systems are used to study different biological processes. Which model system is appropriate to address a particular biological question will depend on the kind of question and the extent to which one is seeking the universal themes or the details that determine how different kinds of organisms can do different things.

There are two vertebrate organisms for which significant genetic manipulation is practiced, mice and zebrafish, *Danio rerio*. Classical genetic studies on mice have been going on for many years and stocks of mice with various mutations are available. Among mammals, mice are small, have a relatively short generation time, and breed prolifically. Tissue culture systems based on mouse tissues are widely available, and many viruses that infect mice have provided important studies on subjects ranging from replication to gene expression to hormonal regulation to immunology and cancer. Finally, it is now relatively routine to reintroduce engineered genes back into mice at specific chromosomal locations (we will discuss this later in the course).

Given this, why work on zebrafish, which are more distantly related to humans, and which have genomes that are larger than humans (XX check this)? Zebrafish certainly didn't have a historical head start as an experimental system relative to mice. Serious genetic studies on zebrafish were started by George Streisinger around 1980 at the University of Oregon.

Zebrafish have lots of nice characteristics that are enumerated by Mullins. She particularly emphasizes the ability to examine embryonic development in living organisms. As external egg layers, zebrafish embryos develop outside their mothers' bodies, unlike mice. In addition, the embryos are not significantly pigmented, so all of the different organ systems can be seen easily. These are features

that zebrafish share with many of the classic systems used to study developmental biology, such as amphibians. Zebrafish are also small compared to frogs, less likely to escape or bite than mice, and are hypoallergenic compared to rodents. Sex determination in zebrafish can be manipulated by the environment you grow them in; in principle you can get as many males or females as you need for the kind of genetic cross you want to do. In the setup described by Mullins, hundreds of small fish tanks, each containing about 50 fish, can be maintained in parallel. You can even stack little fish bowls on your bench like petri dishes at your bench. When I visited Streisinger's lab in the early 1980's, I was struck by how that would appeal to a lab that had formerly worked on bacteriophage.

Special issues for metazoan genetics

These features were factors in Streisinger's choice of zebrafish as a potentially useful genetic system. However, one of the features of zebrafish that he considered important reflects fundamental difference between microbial and metazoan genetic systems.

Haploids vs. Diploids. Yeast and *E. coli* mutants can be isolated from haploid cells that can be propagated by simple cell division to generate more cells that are genetically the same. The phenotypes of mutations are generally easier to detect in haploid organisms than in diploids because most mutations are **recessive**. When a diploid wild-type cell is mutagenized, the probability of generating mutations in both copies of the same gene is very very low. Thus, mutations will occur that would have detectable phenotypes in a haploid or homozygous cell, but their phenotypes will not be observed because the wild-type copy of the gene supplies sufficient amounts of the gene product to complement the mutant.

In a haploid organisms, this just isn't a problem. You isolate the mutations in the absence of a wild-type homolog. One way to get around this problem in a diploid organism would be to start with a situation where at least part of the genome is effectively haploid. This is the case for the X chromosome in males, and in parts of any chromosome that can be made heterozygous where one homolog is partially deleted.

This could be done in diploid yeast, although there is usually no point, since the haploid forms are

haploid over the whole genome. It could be useful for looking for genes that affect aspects of the life cycle that are specific to diploid cells. For example, to screen for mutations that affect meiosis one has to use diploid cells. Note that in yeast, the needed heterozygous diploid can be propagated from a single cell into a large population of diploid, heterozygous individuals, which can then be subjected to mutagenesis.

In metazoans, phenotypes are only detectable in stages of the life cycle after fertilization. Not only are sperm and eggs generally unable to grow on their own, their genomes are dormant until well into the process of embryogenesis. All of the proteins and RNA molecules in a sperm cell or an egg were synthesized by the diploid parent. Thus, while the haploid, gametic forms in the life cycle can be mutagenized, we can't see the consequences of the mutations until the mutagenized genome is in the same nucleus as a wild-type copy. Unlike yeast, even when or if you can make an animal that has the appropriate genotype, that animal can only reproduce by passing half of its genome onto each of its offspring.

The way in which this is traditionally solved is to use carefully designed controlled breeding experiments to generate stocks of animals with the appropriate genotypes. The second half of the Mullins paper illustrates how this was done to screen for recessive embryonic lethal mutations. The details of this will be discussed below, but note that mutant phenotypes were detected in the third generation after the mutagenesis step. Since the generation time is about three months, this means that you don't find the interesting mutants until 9 months after the original mutagenesis.

Just as it takes a long time to identify mutants, every kind of strain construction is complicated by the fact that diploid metazoans don't reproduce asexually. The offspring of crosses from the same parents are genetically different because even though they all get half of a genome's worth of chromosomes from each parent, they get different subsets of from each parent due to the combination of the independent segregation of the chromosomes and the recombination of different chromosomal segments.

Inbreeding.

To construct genetically homogeneous populations of animals, repeated cycles of inbreeding and controlled crosses between inbred stocks are needed. Inbreeding reduces the number of alleles that are present in a stock. Imagine that we start with a pair of fish that have allelic differences at three loci. Represent this as one parent having the genotype *A/A, B/B* and the other has the genotype *a/a, b/b*. Assume that the lowercase allele in each case is a recessive mutation that can be detected as a phenotype in a homozygous animal. Suppose that we want to make a stock with the genotype *a/a, B/B*. In the first generation of progeny (the F1) the genotypes of all of the progeny will be *A/a, B/b*. Next, inbreed the progeny to obtain the F2 generation. The genotypes among the F2 will be different, and the expected genotypes can be enumerated by using a Punnet square (Figure 3-1).

			Second parent			
			A		*a*	
			B	*b*	*B*	*b*
First parent	*A*	*B*	*A/A, B/B*	*A/A, B/b*	*A/a, B/B*	*A/a, B/b*
		b	*A/A, b/B*	*A/A, b/b*	*A/a, b/B*	*A/a, b/b*
	a	*B*	*a/A, B/B*	*a/A, B/b*	***a/a, B/B***	***a/a, B/b***
		b	*a/A, b/B*	*a/A, b/b*	***a/a, b/B***	*a/a, b/b*

Figure 3-1. Genotypes for the F2 progeny from inbreeding. Progeny where the *a* phenotype are boxed, and the subset that will appear wt for *B* are indicated in boldface.

The fish that have the *a* phenotype could have three genotypes with respect to *B*. Suppose we can observe the *b* phenotype and eliminate the fish that are *b/b*. To continue toward the goal of the desired stock, we next need to inbreed the a/a fish from the F2 generation. However, since we can't tell whether the fish with the *B* phenotype are *B/B* or *B/b*, we have to consider all of the possibilities for how we will choose the parents of the F3 generation (Figure 3-2).

If we can mate each parent to more than one partner, and we had enough crosses so that we could be reasonably certain that all possibilities were sampled, then the F3 phenotypes will tell us which F2 animals were homozygous or heterozygous at the B locus. We can then pick the homozygous animals as founders for a line of inbred animals the desired genotype. For some kinds of animals this is

practical (I think it would work for fish, but I'm not sure XX Ask Bruce Riley); for *Drosophila,* where sperm storage by females requires the use of virgins for strain construction, this is not possible.

		Second parent		
		B/B	*B/b*	*b/B*
First parent	*B/B*	all *B/B*	*B/B* + *B/b* 1:1	
	B/b	*B/B* + *B/b* 1:1	*B/B* + *B/b* + *b/b* 1:2:1	
	b/B			

Figure 3-2. Genotypes for the F3 progeny from inbreeding. Progeny where the *a* phenotype are boxed, and the subset that will appear wt for *B* are indicated in boldface.

If we can't go back and figure out which groups were from homozygous F2 parents, we can at least discard the progeny of those crosses that give *b/b* siblings. If we now pick F3 animals and continue the inbreeding, the possible parental genotypes remain the same as in the F2 X F2 cross, but the relative abundance of B/B and B/b genotypes will be shifted from 1:2 for the F2 cross to 2:1 in the F3 cross. With each succeeding round of inbreeding, two things are happening. First, the probability of picking a B/B parent increases with each round. Second, as soon as you pick two *B/B* parents, there is no longer a source of the *b* allele in the population. The B allele becomes **fixed**.

In real inbreeding experiments, the experimenter has to worry about getting homozygosity at more than one additional locus. In addition, inbreeding generates homozygosity even if the phenotypes of the different alleles at different loci are not detectable. At some point, how the chromosomes segregate to generate gametes becomes irrelevant because both homologs from each parent are the same.

Parthenogenesis

Streisinger hoped to shorten the number of generations needed for genetic manipulations needed to do zebrafish genetics by taking advantage of the ability to stimulate zebrafish eggs to develop into adults in the absence of fertilization by sperm, a process called **parthenogenesis**. Parthenogenic fish will be genetically haploid, in the sense that they will only have the chromosomes from one parent that made it into the egg.

Now consider how we would make an inbred line using parthenogenesis. Instead of inbreeding the F1 fish, use the eggs of F1 females to generate parthenogenic fish. Note that although different parthenogenic fish from the same mother will still be genetically different from each other, since the different alleles in the F1 heterozygotes still segregate through meiosis. However, now, instead of having 4 kinds of *a/a* fish, we will only get two: *a/a, B/B* and *a/a, b/b*. Now we cross the *a/a, B/B* fish back to the *A/A, B/B* stock, inbreed once, and we're basically done, since we don't have to worry about heterozygosity at the *B* locus.

This has the potential to save many generations of inbreeding, and was one reason Streisinger thought that zebrafish genetics could advance much faster than mouse genetics. Unfortunately, it turns out that the haploid fish generated by parthenogenesis tend to have abnormalities and high death rates. Mullins et al. decided to treat zebrafish genetics the same way that genetics is done in other animal systems.

Optimizing mutagenesis

Spontaneous mutations do happen in higher organisms, just as they do in microorganisms. However, they are still rare events. In genetic studies of *E. coli* and yeast, the sheer numbers of individuals that can be grown makes the isolation of spontaneous mutations possible as long as there is a way to find them among the $>10^5$ fold excess of nonmutant cells (We will address this in chapter 4). For all metazoan genetic systems, mutagenesis is a necessary component of any large-scale mutant hunt.

In a haploid unicellular microorganism like *E. coli* or yeast, the survivors of a mutagenic treatment propagate into populations of mutated cells. Every cell in a mutant colony (or at least a vast majority of the cells) will be able to give rise to another colony with the same mutation. We have already seen how a recessive mutation in a diploid organism won't be seen as a mutant phenotype. If we start with a wild-type organism, a couple of rounds of breeding are needed to see the mutation (see Figure 3 in Mullins). Mutagenesis in metazoans is complicated by another factor known as **mosaicism** (see below). Since the requirements for several rounds of breeding mean that each screen for mutants takes about 9 months, it is clearly important that the procedure for doing the mutagenesis be something that

works. The first part of the Mullins paper deals with the question of how to optimize the mutagenesis protocol.

The general approach to mutagenesis was to treat male fish with chemical mutagens so that mutations will accumulate in the sperm of the treated fish. Figure 3-3 shows a cartoon diagram of what they did.

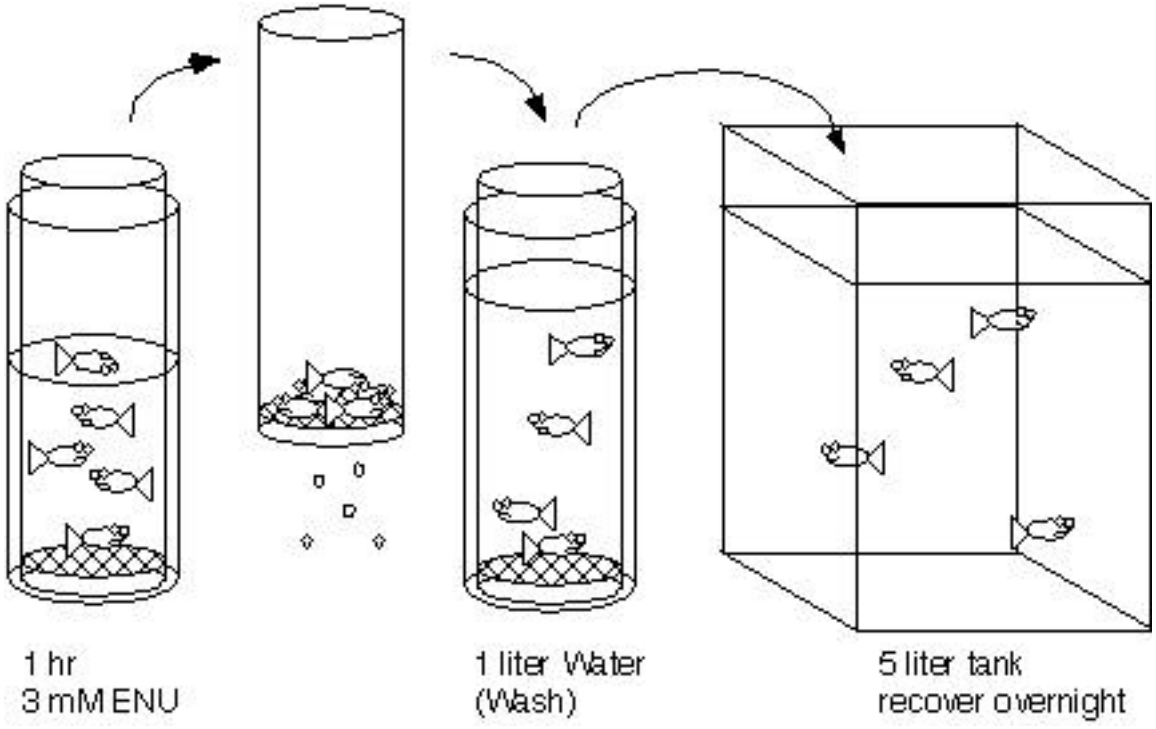


Figure 3-3. Setup for mutagenesis with ENU.

Fish were placed in a setup that consists of two plastic cylinders, one inside the other. The the outer cylinder holds water with added mutagen, and the inner cylinder has a mesh bottom so that the treated fish can be recovered by pulling out the inner cylinder and letting the water drain out the bottom. The excess mutagen is washed of by dunking the cylinder int fresh water, and then the fish are transferred into a regular tank to recover.

The treatments were repeated a couple of times, and then the fish were kept in untreated water and mated to untreated females periodically (Table 3-1).

Table 3-1		
Day	**Action**	**Comments**
1	Mutagenize	No fertile eggs
2	Mate	
3	Mutagenize	
4	Mate	
5	Mutagenize	
7	Mate	
14…	Continue to mate 2X weekly	Infertile, Abnormal and Mosaic
21…		Data in Fig. 2 of Mullins paper

Noncomplementation assay for mutations

In order to figure out what conditions are best for obtaining lots of mutants, Mullins needed to be able to determine the frequency of mutants among the embryos from the fertilized eggs produced by mating mutagenized males to females. There were two basic approaches she could have taken. She could have analyzed the F1 progeny by mating them to produce and F2 and then inbreeding the F2 to generate a population of F3 fish that would include 25% that were homozygous for any mutation that was produced. In fact, this is what they did after mutagenesis conditions were optimized, when they were looking for new mutations in unknown genes. However, as noted above, this would take total of 9 months to perform each assay.

Mullins chose an assay that takes about 2 to 3 days instead of 9 months. In order to measure relative mutation rates, it isn't necessary to isolate mutations in new genes. It's much easier to just count the frequency of mutations in a previously identified gene by setting up the matings so that the females can't provide a wild-type copy of the gene being tested.

Mullins mated the mutagenized males to females that were homozygous for one or two recessive mutations in genes that are nonessential. The mutations, which affect things like pigmentation and shape, are easily scored visual phenotypes when mutated. Figure 1 in the Mullins paper shows an example, using females homozygous for the *gol*^{b1} mutation, which causes a loss of pigmentation in embryos and adults. Consider what will happen when eggs from the *gol*^{b1} / *gol*^{b1} female are

fertilized with sperm from a mutagenized male. Most of the time (>99.7%), the sperm cell that fertilizes the egg will carry the wild-type allele at the *gol* locus, and the mutation derived from the mother will be complemented. The embryo will be pigmented.

The only way to get a nonpigmented embryo is if the paternal gene is mutated to a form that fails to complement the maternal gene. Thus, the frequency of noncomplementation measures the frequency with which mutations arise in the *gol* gene.

Mosaics

As the fish swim unhappily in the mutagen solution, the mutagen is modifying their DNA (and RNA and protein and who knows what else). The chemical reactions are clearly going on during the first day of treatment, since some of the fish die from the treatment and no fertile eggs are obtained from the early matings. As the males recover from the mutagenesis, live offspring start to appear, but they are **mosaics**, animals where different cells in the same animal are phenotypically different.

Mosaics can occur by a variety of mechanisms, but basically fall into two classes: situations where the different phenotype reflects differences in genotype at the cellular level, and those where the difference in phenotype does not reflect differences in the allelic content of the cells. The latter situation is called an **epigenetic** effect. An example of an epigenetic form of mosaicism is found in human females, where one of the two X chromosomes is inactivated in each cell. Whether the paternal or maternal chromosome is inactivated in any given cell seems to be more or less random, except to the extent that which homolog is inactivated tends to stay the same through a cell lineage.

Among the embryos produced two weeks after the mutagenesis, Mullins observed embryos where a few patches of cells lacked pigmentation. Although they didn't do any experiments to show this, it is reasonable to accept their unstated assumption that these are genetic mosaics, not epigenetic ones. However, none of these fish passed the mutant allele on when mated to homozygous partners. This indicated that the frequency of the genetic change was so low that changes in the germ line were too rare to be observed.

How could the fertilization of a haploid egg by a haploid sperm cell result in a single cell that will divide to give rise to cells with different genotypes? At first glance this seems weird; the *gol* gene in

the egg has to be mutant, and there is only one copy of the *gol* gene provided by the sperm. The paternal chromosome seems to be the key, because the mosaics are not observed when unmutagenized males are used, nor when homozygous mutant males are used (not shown in the paper). Somehow, the sperm from males treated two weeks before the mating retains both alleles. In addition, the sperm from the same males seems to change a week later, since the mosaic embryos are not found when the mating is done three weeks after treatment.

The solution to this mystery lies in remembering that although the sperm only brings in one copy of each chromosome, and each chromosome only has one molecule of DNA, the DNA is double stranded. Thus, a mosaic should occur if the two strands carried different information, instead of the normal complementary information ensured by base pairing. However, the observations are not explained by the simple idea that noncomplementary bases are found in each strand.

Consider a situation where a G:C base pair in the wild-type allele is replaced by a G:T base pair. This could occur if the C is normally methylated at the 5 position, and mutagenesis caused loss of the amino group. If this kind of **heteroduplex** DNA molecule is found in the sperm, the fertilized egg will also have a heteroduplex paternal homolog. Three things can happen. The mismatch can be repaired to the wt sequence before replication, and a wild-type embryo will result. The mismatch can be repaired to the mutant allele, and a homogeneous mutant embryo will result. This is not observed. Third, if the mismatch is not repaired before the first S phase, the two strands will go into different daughter cells in the first cell division, leading to different genotypes in the cells derived from each daughter. This explanation is qualitatively consistent with the observation of mosaics, but it is still inconsistent with the overall date. If a mismatch between normal bases was present, we'd expect to see half of the tissues in each embryo with each genotype. In fact, only small patches of mutant tissue were found, and germline transmission was not observed.

Let's look at this process in more detail for mutagenesis with 1-Ethyl-1-nitrosourea (ENU). ENU alkylates the O6 position of guanine and the O4 position of thymine of DNA. The mutations induced by ENU are mostly G:C to A:T transitions, so we will only consider what happens when G residues

are modified (Figure 3-4). Suppose a sperm cell contains a O6-ethyl G paired with C at a position where a G:C to A:T transition would cause a mutation in the *gol* gene. Like unmodified G, O6-ethyl G can basepair with C, but the modified G discriminates against T less well than the unmodified base.

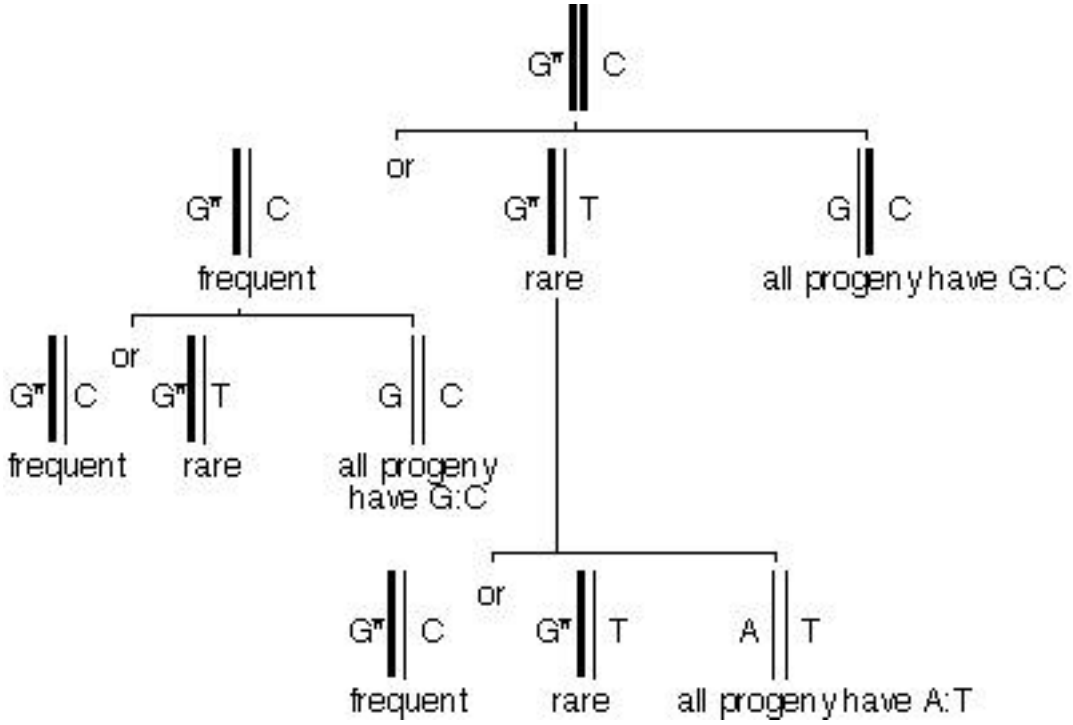


Figure 3-4. Replication of modified DNA.

After fertilization, the fertilized egg replicates its DNA and the two strands separate and are copied. The strand containing the C will give rise to a G on the opposite strand. This leads to a normal duplex DNA molecule that encodes the wild-type allele; additional rounds of replication propagate the G:C pair. The daughter cell that inherits the C strand, and all of its descendants, will thus all be genotypically wild-type. The other strand, which segregates into the other daughter cell, can have two fates. The O6-ethyl G can pair with either C or T to form one of two different kinds of molecules. If DNA polymerase incorporates a C opposite the O6-ethyl G, the daughter cell that inherits the G strand will have the same kind of O6-ethyl G:C pair as was found in the fertilized egg. On the other hand, if DNA polymerase incorporates a T opposite the O6-ethyl G, then the cell that inherited the G strand will segregate out daughters with A:T basepairs when the strand containing the O6-ethyl G and the strand containing the T separate during the next round of replication. At least half of the descendants of this

cell lineage will have a mutant genotype. Since the modified G residue is passed down the opposite arm of the lineage, the O6-ethyl G can continue to give rise to more mutant cells until it either arrives in a terminally differentiated cell or is removed by DNA repair.

This explains the origin of mosaics. The fact that the mosaics only contained small patches of mutant tissue suggests that O6-ethyl G:T basepairs are still rare, since any given G residue is either modified in the fertilized egg or not. There's no way to modify the DNA in a small fraction of the cells later in development.

As Mullins continued to mate the mutagenized males to the tester females, the mosaics disappeared, and mutant F1 progeny were found that passed their mutations down to succeeding generations. This suggests that the sperm produced at these times now contain mutant basepairs on both strands. This makes sense if we superimpose the possible fates of the DNA strands shown in Figure 3-4 onto the cell lineages of a hypothetical pathway for spermatogenesis (Figure 3-5).

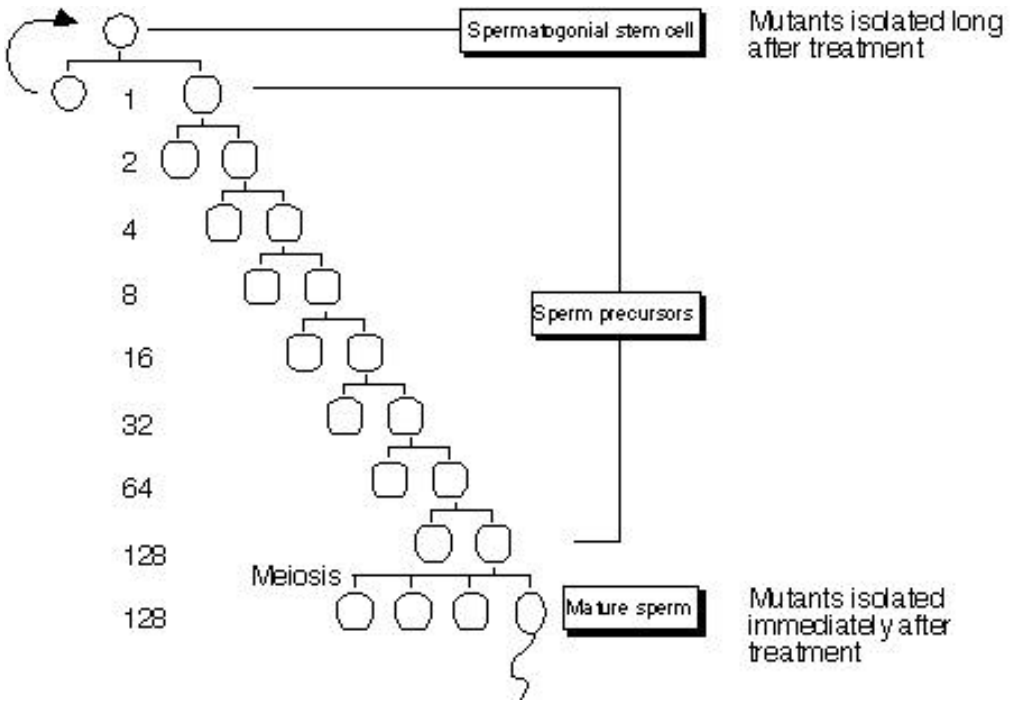


Figure 3-5. Hypothetical pathway for generation of sperm from stem cells, based on spermatogenesis in mice

Early after mutagenesis, the sperm fertilize the eggs were either already there in the males or were derived from late precursors in the pathway. Thus, the relative abundance of O6-ethyl G:C basepairs

to A:T basepairs favors the former. When the males are mated at later and later times, the sperm derive from cells that were at earlier and earlier stages of spermatogenesis when the fish were exposed to the mutagen. The males become mosaic animals with respect to their sperm precursors and their sperm, but since only a single sperm is used to fertilize each egg, the F1 progeny will not be mosaics.

Embryonic lethals - how many genes?

Having determined good conditions for mutagenesis, Mullins performed a pilot experiment to show that their protocol could be used to isolate lots of mutations in genes required for development. Note that the screen for mutations looks for mutant embryos that start to divide and then arrest or develop abnormally (all of the F3 embryos will start development normally because embryonic genes are not expressed until several cell divisions after fertilization).

The frequency of mutations induced by ENU treatment varied about 2-fold among the four genes tested by the noncomplementation assay, with an average value of 2.3×10^{-3} mutant embryos per embryo screened. This frequency can be used to estimate the number of genes essential for development in the zebrafish if they could also estimate the number of lethal mutations that were produced per haploid genome screened by the mutagenesis. Note that this is not the same as the number of mutations that were found, since breeding steps required to detect mutations will necessarily mean that some mutations will be missed.

To understand why this is the case, we need to examine how Mullins screened for the embryonic lethal mutations (Figure 3-6). Unlike the mutations used to optimize the mutagenesis protocol, they didn't know ahead of time what genes would be affected. Thus, the only way to see the phenotype of a recessive mutation, they need to get the same mutation into both copies of the gene in the same embryo. This requires that both parents from the preceding generation have the same mutation in one copy. Since fish don't reproduce asexually, the fastest way to achieve this is to inbreed the offspring of the F1 fish that first carried the mutation derived from the mutant sperm into the germ line. Thus, mutagenized males mate with wild-type females to give the F1, the F1s are crossed to each other to give F2s, and pairs of F2s are picked for inbreeding.

In order to save crosses, Mullins sometimes crossed F1 fish to each other instead of crossing each F1 to a wild-type. It is very unlikely that the two F1s will carry mutations in the same gene, so, at any given locus, the F2 will be either heterozygous for a mutations found one of the parents or homozygous for the wild-type allele. As shown in the figure, there are four possible F2 genotypes.

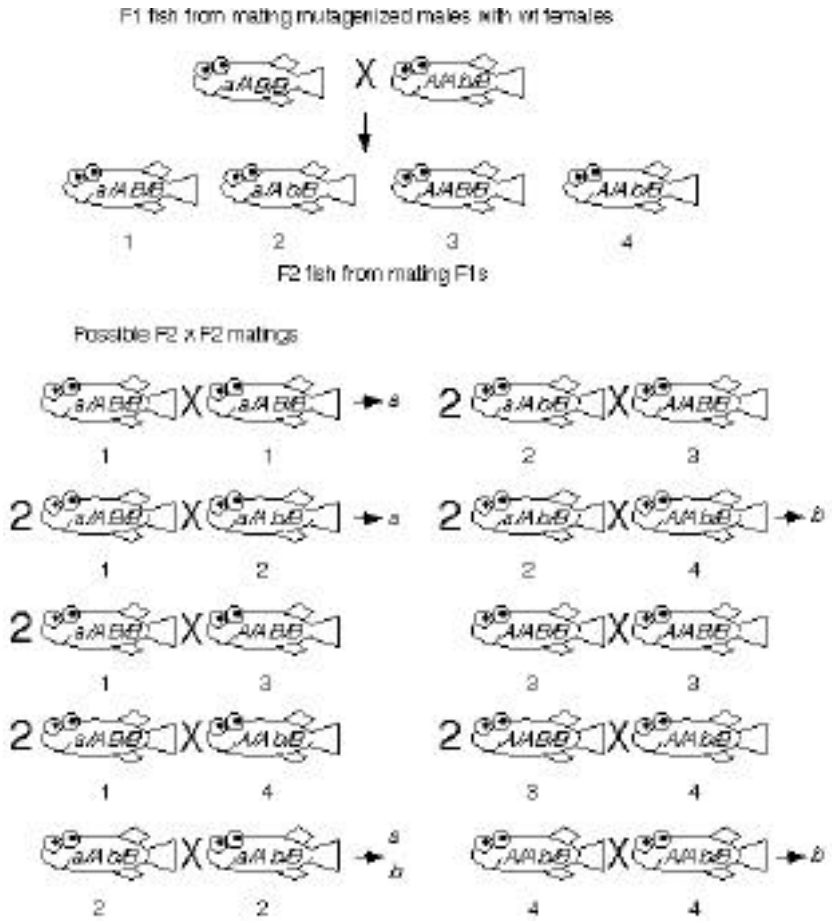


Figure 3-6. Matings used to screen for embryonic lethals.

Consider what happens each time a pair of F2 fish are picked to mate. Since the F1 fish are heterozygous for any induced mutations, only half of the F2s will have the mutant allele. Recall that both F2 parents have to carry the mutation in order to see the phenotype in the F3 generation. A mutation will be missed whenever either parent is homozygous for the wild-type allele; this will be the case in 75% of the randomly chosen pairs. For each cross there is only a 25% chance of finding each mutation. There is only a 7/16 chance of finding either mutation and a 9/16 chance of missing both. Mullins did up to 8 F2 pair matings for every F1 tested. The probability of missing both mutations falls to($9/16)^8$ = approx. 0.01. The probability of missing each mutation is on the order of 10%.

Although the details of the calculation described above and the formulas shown in the Materials and Methods are slightly different, the bottom line is that the number of matings used was sufficient to find the vast majority of the induced mutations.

Mullins finally argues that about 1.2 lethal mutations were induced in each mutagenized genome. If the probability of hitting any individual gene was 2.3×10^{-3} mutations/genome then

2.3×10^{-3} mutations/genome screened x X genes = 1.2 mutations/genome screened

so $X = 1.2 / 2.3 \times 10^{-3} = 0.5 \times 10^{3} = 500$

Mullins gets 600 by rounding down the mutation frequency. Again, within the expected precision of this kind of calculation, the point is that one would expect there to be about 500-1,000 essential genes.

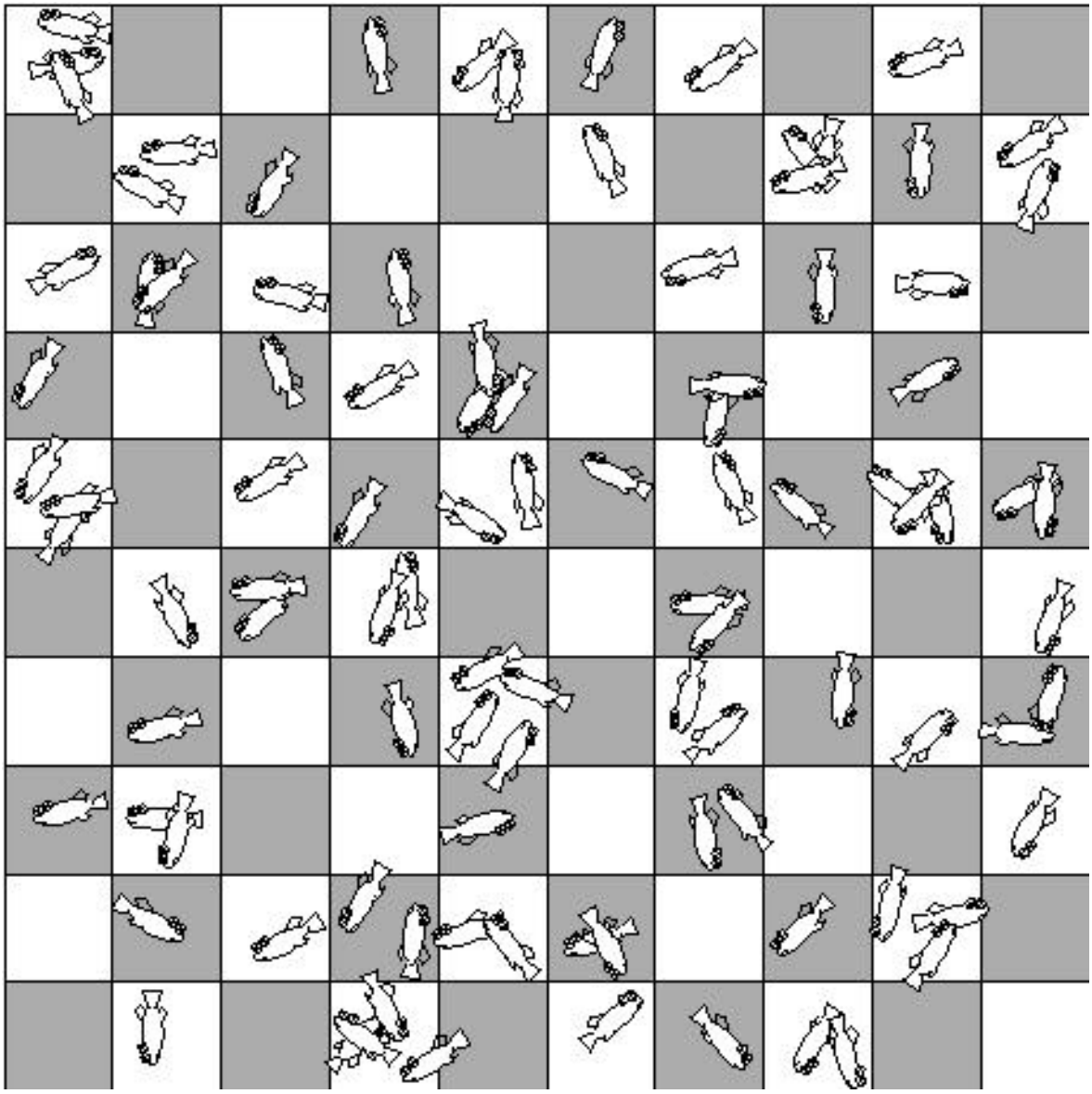
The Poisson distribution

Mullins then estimates how many genomes they would have to screen in order to find mutations in most of these genes. This calculation is an example of one of the many applications of the Poisson equation to molecular biology and genetics. The Poisson distribution allows us to calculate the expected distribution of random events.

Imagine that animal rights activists have bombed the lab where the zebrafish experiments are being done. Although the fish room was not destroyed, enough of the tanks were shattered so that 100 fish wound up on the floor, which happens to be covered by a 10 x 10 grid of tiles (Figure 3-7). The overall average number of fish per tile is one. However, in a random distribution, some tiles will have no fish and some will get more than one. The Poisson equation can be used to calculate the expected distribution of events in a series of tests:

$$P(i) = \frac{m^i e^{-m}}{i!}$$

where P(i) is the frequency of tests with *i* events and *m* is the average number of events per test. In the example, *m* is the average number of fish/tile and P(0) gives the fraction of tiles with no fish, P(1) gives the fraction with exactly 1 fish, P(2) gives the fraction with 2 fish etc.



For m=1, the probabilities for some different cases are:

i	0	1	2	3	4	5
P(*i*)	.369	.368	.184	.06	.02	.003

Thus, if the fish are randomly distributed, we'd expect about 37 empty squares, 37 with one fish,

18 with two etc. If the distribution differs dramatically from this, then it probably isn't random. For example, finding all 100 fish in one tile is very unlikely. We can calculate the precise probability of this occurring by solving for P(100), which turns out to be about 3×10^{-159}.

Mullins' calculation of how many genomes must be screened to find mutations in 95% of the essential genes demonstrates another common use of the Poisson distribution. The calculation is based on the following reasoning. As more and more mutations are isolated in essential genes, the probability of finding more than one mutation in the same gene also increases. Stated in terms of the Poisson equation, let *m* be the average number of mutations per gene. As m increases, the probabilities of classes involving multiple hits in the same gene increase and the probability of missing any individual gene decreases. To find 95% of the genes, we can look for the value of m such that the sum of all P(*i*) values for i>0 is equal to 0.95. The simplest way to do this is to remember that the sum of all of the probabilities is equal to 1. Thus, we want to solve for the value of *m* where

1-P(0) = 0.95.

It turns out that this happens at around *m*=3. The P(0) class is especially convenient for calculations of this kind because $m^0=1$ for all values of *m* and 0!=1 by definition. Thus, the general form of the Poisson equation reduces to:

$$P(0)=e^{-m}$$

The Poisson equation has lots of uses that are different formulations of the same problem. For example, you can use it to calculate how many clones in a library you need to sequence in order to cover a given percentage of the genome in a genome project, or determine what fraction of cells mixed with a virus remain uninfected. You can use it to determine the probability that the incidence of cancer in a population of experimental animals (or people) is statistically significant, or whether the variation in the grade distributions of various sections of a class are meaningful.

Where did this work go next?

The questions addressed by the Mullins and Carlson papers are similar in many respects: how can we isolate genes that encode products needed for a process of interest, and how many such genes are there? The scale of the questions could not be more different. The process Carlson sought to dissect started with the control of a single enzyme involved in a specific metabolic pathway. Mullins describes initial steps toward asking similar questions about the development of an entire vertebrate organism.

As foreshadowed in the pilot experiment and the discussion of the scale needed to find 95% of the genes, by 1996, Nüsslein-Vollhard's lab at the Max Planck institute in Tübingen, Germany and Wolfgang Driever's lab at Massachusetts General Hospital had scaled up the mutant hunt and had found over 1800 mutations with developmental phenotypes, and had shown that these identified on the order of 500 new genes. This work, which has been called a historic achievement, was described in a series of papers that took up an entire issue of the journal *Development*. They found mutants that affected everything from general cellular functions such as cell cycle progression or nuclear replication to the formation of a variety of specific tissues. This mutant collection will provide an incredible resource for developmental biologists and is a major step along the path to enlightenment for understanding vertebrate biology.

