Mutations and mutant hunts

Papers to read for this section

1. **Brenner, S., W. Dove, I. Herskowitz, and R. Thomas** (1990). Genes and development: molecular and logical themes. *Genetics* **126:**479-86. [concentrate on the section subtitled "The genetic analysis of development", which starts on p482]
2. **Carlson, M., B. Osmond, and D. Botstein**. (1981). Mutants of yeast defective in sucrose utilization. Genetics **98:**25-40
3. **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

Identifying components - Genetic and Biochemical approaches

The first part of Alberts' Path to Enlightenment involves cataloging the components involved in the process you are interested in. Biochemical and genetic approaches can be thought of as the extremes. In one kind of biochemical approach to a problem, one seeks to purify to homogeneity all of the needed components of a process, and to reconstitute the activity in vitro. Where the processes can be reconstituted, the components of the reaction can be said to be **sufficient** for the activity.

In principle, this approach is very powerful, and purified components are essential to move on to most of the later steps in the Path to Enlightenment. However, there at least three very large fundamental potential problems, other than simple failure to obtain the desired material, with the biochemical approach. First, you need an assay for the process. Some processes can be readily performed in cell free systems; others, especially those that operate at levels of organization above single cells, such as the formation of a particular kind of tissue in a developing embryo, cannot by their nature be reconstituted completely from purified components. In many cases the best one can hope for is an assay for something simpler. This leads to the second problem. A working system reconstituted in vitro may not necessarily reflect what goes on in the organism. There are examples where in vitro assays purified the "wrong" enzyme, for example, solution conditions used in vitro might give a result where an enzyme appears to be involved in synthesizing a metabolite, when its role in the cell is

actually degradation of the same metabolite. Third, even with an assay for the right activity, one might miss components essential to the cell's ability to actually perform the reaction in the complex environment of the cell.

Note that having an assay that works in vitro can provide useful information even if it is not possible to reduce it to purified components. The properties of the complete system or of partially purified subfractions can be tested by a variety of challenges. For example, consider one kind of experiment that could be used to show that protein synthesis is programmed by mRNA rather than directly from DNA. The experiment I describe is not exactly what was done, but is conceptually similar. Imagine that you could show that extracts from virus-infected cells made viral proteins that were not found in extracts from uninfected cells. Suppose that you could show that a subfraction from the infected cells, which is not active in protein synthesis by itself, could program the synthesis of a viral protein when added to a complete extract from uninfected cells. How might you show that RNA is involved? Suppose that you could show that the ability of the fraction from the infected cells to direct viral protein synthesis was sensitive to RNAse but not to DNase or proteases. This would implicate an essential RNA component even though no RNA had been purified, even as a protein-free fraction. You could say that some kind of RNA was **necessary** for the reaction.

Other kinds of treatments can be used to determine the characteristics of impure components. The stability of activities against heat inactivation is often used. The translation elongation factors EF-Ts and EF-Tu are so named because one was a temperature-stable activity and the other was a temperature-unstable activity. Both were required for elongation steps during translation. Sensitivity to chemical modification reagents such as N-ethyl maleimide (NEM) can indicate types of essential amino acid residues. An NEM-sensitive fraction is presumed to have a component containing a reactive cysteine where modification destroys the activity.

Chemical and environmental treatments will not be specific to the product of a specific gene. Antibodies or other affinity based reagents can be used to deplete an extract of a specific protein, and can thus show that a specific component is involved in the process. Note however, that antibodies are

only useful if you already have them, and making antibodies requires purified antigens. This severely limits the use of antibodies to find new components in a process.

The genetic approach is based on the idea that genes that encode components involved in a process can be identified by isolation of mutants that alter how the process behaves in the organism. Thus, if *E. coli* normally makes β-galactosidase only in the presence of lactose, mutations that cause the synthesis of β-galactosidase under inappropriate conditions will identify genes involved in the regulatory pathway that determines lactose specificity for induction. If a mutation that results in loss of function of the gene product also abolishes the process under study, it is deduced that the gene product is necessary for the activity. A screen for inactivating mutations can be thought of as analogous to treating an extract with a series of antibodies to deplete it of every possible component. Think of each mutant colony that arises from a population of mutagenized *E. coli* or yeast as a different assay where a single component has been depleted. Since hundreds to billions of colonies can be tested in parallel, depending on the nature of the screen or selection used to isolate the mutants, a genetic approach can clearly go through the candidates for possible components very quickly. This massively parallel property is part of what gives genetics its "awesome power". Genetics also derives some of its power from the fact that, since phenotypic testing can be done on anything from extracts to cells to tissues to populations of whole organisms, a much broader array of assays for a genetic change are available than can be used in a biochemical approach.

Note that a purely genetic approach has its own theoretical pitfalls (see the reading assignment in the Brenner paper). First, a loss-of-function phenotype where your favorite process is lost only tells you that there is some kind of functional connection between the gene and the process. This connection can be so indirect as to be virtually useless in terms of understanding. For example, if you were looking for mutants that were unable to ferment maltose, you would expect to get mutations in genes that encode the enzymes involved in maltose metabolism. However, an exhaustive screen for mutants with a Maltose-negative phenotype would also yield mutations in adenyl cyclase and catabolite activator protein (CAP), which are involved in the regulation of how the genes for maltose metabolism are expressed. Although those might be very interesting, they are not directly involved in converting

maltose to glucose (part of the power of genetics is its ability to steer the prepared mind toward something more interesting than the original problem).

Second, it is very difficult to assess whether or not mutations have been isolated in every necessary component. Geneticists will readily acknowledge two major reasons for this:

1) The essential gene problem: it is often difficult to isolate mutations in a gene that is essential for the viability of the organism. Mutations in genes that are essential must be found as either "leaky" mutations or conditional lethals.

2) The redundancy problem: if the genome contains more than one gene that can provide an essential activity, and if only one of the copies is sufficient to allow the process to proceed, then none of the genes in this family are likely to be identified by loss of function mutations.

Ask a typical geneticist who has spent a long time repeating his or her mutant hunt whether the study has identified all of the components of a process, and you will almost automatically get the response "well, we might have missed essential or redundant genes". However, there is a third reason that comes into play more often than most geneticists will admit unless pressed: luck. The probability of finding mutations in all of the nonessential and nonredundant genes encoding products that act in a process, or "saturating the screen", depends on how many genes there are to be found, the probability of each of them being mutated to a loss-of-function form and the number of candidates the experimenter is willing to look at before giving up. This is a real issue; the existence of additional components will sometimes be demonstrated years after a thorough genetic study by either finding more stuff that you can't get of in the biochemical assay or by a variation on the original genetic screen, such as a change in the assay or the kind of mutagenesis used, or blind luck.

I described purely biochemical and purely genetic approaches as extremes. The essence of molecular biology is in the synergy that is obtained from combining approaches. Thus, genetics provides powerful leverage for classical biochemistry by providing mutants as starting materials. Imagine that you have extracts from two different mutant strains that have inactivation mutations in genes that encode different proteins, A and B, involved in the process you are studying. "In vitro

complementation" occurs if you can mix the two extracts and get activity. The extract that is missing component A gets it from the other extract which has A but lacks B. If in vitro complementation is observed, then the ability to reactivate the first extract can be used as an assay to purify A, while the ability to reactivate the second extract provides an assay to purify B.

What is learned from the biochemistry of a process can also be useful for the genetics. Partially understanding the mechanism may provide insights into how to design a more efficient selection or screen, biochemical assays can be used as screens, and peptide sequences from purified proteins can be used to identify candidate genes.

Genetic Approaches

The first section of the course will concentrate on approaches to identifying those genes that encode products involved in processes of interest. There are a variety of approaches to this, which I would divide into two classes:

- Classical - Mutations to Genes.

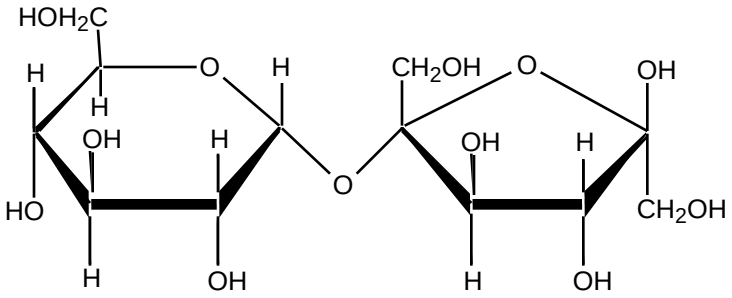
  The classical approach is basically what is described in outline above. Mutants are isolated that have discernable phenotypes, and the affected genes are identified. The majority of our time will be spent on the classical approach
- Reverse - Genes to mutations.

  A reverse genetic approach identifies candidate genes on the basis of either proteins or DNA sequences. Mutations are then targeted to or engineered into the candidate genes in order to determine the phenotype of inactivating the gene.

The goal of this section is to describe the how one goes about a classical mutant hunt as a means of identifying genes. We will examine a pair of screens for mutants: one in yeast and one in zebrafish. By comparing and contrasting these screens, I hope to illustrate most of the issues that one should think about before embarking on a screen, and what one does with mutant candidates once they have been isolated. Basic genetic ideas will be reviewed as needed for those who have not had an

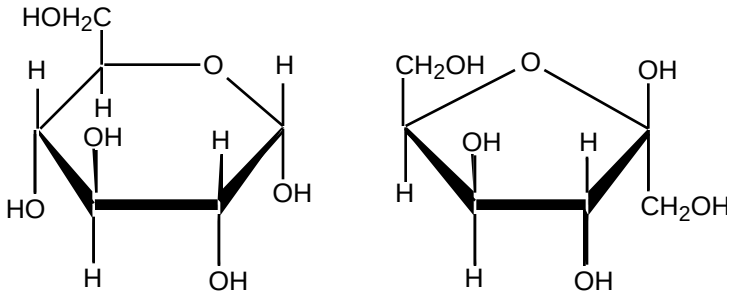
introductory genetics course.

In the reading assignment, Carlson, Osmond and Botstein set out to determine what genes in yeast are needed specifically to use sucrose as a carbon source. They already knew that sucrose was broken down into glucose and fructose by a secreted form of the enzyme invertase (the name comes from the difference in optical activity between the reactants and products; Figure 1-1). They also knew that yeast would grow happily on either glucose or fructose if either monosaccharide was used as a sole carbon source. Therefore, they expected to get mutations in the structural gene for invertase, which they already knew was a homodimeric protein. If they found mutations in more than one gene, then that would imply that there are more components involved in sucrose utilization.



Sucrose (α-D-Glucopyranosyl-(1-2)-β-D-fructofuranoside

Invertase



Glucose Fructose

Mutagenesis

In order to find mutant strains that were unable to use sucrose, they needed two things: a starting population that contained mutants and a way to find the mutants among an excess of wild-type cells. Mutations in any specific gene are rare in a culture grown under normal conditions. This makes sense,

since many mutations are deleterious and every organism that has been looked at has extensive mechanism to proofread DNA sequences and to repair errors. Spontaneous mutation rates vary in different organisms and for different genes, but the range is on the order of 10^{-4} to 10^{-6} per cell per generation. Thus, if Carlson had simply started looking at individual yeast colonies from a normal culture, she probably would have had to check more than 10,000 to a million candidates to have an even chance of finding a single mutant. This is clearly not a viable strategy if the goal is to find all of the genes. There are two ways to get around this problem. Either increase the fraction of mutants in the starting population or find a more efficient way to find them. They took the first approach by mutagenizing their starting yeast strains with ethane methane sulfonate (EMS).

EMS is an alkylating agent, which probably increases the mutation rate by both direct and indirect mechanisms due to the modification of bases in the DNA. The direct effect of EMS will be due to misreading of an altered base during replication. For example, modification of a G residue could allow it to pair better with U than is normally the case. Even if C is still the preferred partner for the modified G, the mutation rate will increase.

I would expect EMS to have an indirect effect on mutation rates as well. Severe DNA damage induces repair systems in most organisms. Inducible repair systems tend to be efficient at the expense of accuracy; certain repair enzymes are part of "error-prone repair" systems. When these systems are induced, undamaged parts of the DNA may accumulate mutations at a higher rate than normal.

The existence of indirect mutagenesis was first shown in studies of *E. coli*. Untreated phage were used to infect cells that were preirradiated with UV light. The number of mutants could then be determined by looking at the plaques formed by the progeny of the infection, and comparing the number of abnormal plaques to the number seen in a control infection. Although the phage DNA was not exposed to UV, the progeny from infections in irradiated cells contained 100-fold more mutant plaques than the phage grown in control cells. Similar results can be observed if the repair systems are constitutively induced due to mutations in the systems that normally keep them turned off.

Screening

Carlson used EMS treatments that were harsh enough to kill 70% of her starting yeast cells. The frequency of mutants among the survivors was increased, but even with heavy mutagenesis, only a small fraction of the cells were unable to grow on sucrose. To find these cells, Carlson performed a screen on single colonies that grew from the survivors. In order to do "high-throughput" screening, they used replica plating, a method developed by Lederberg for screening *E. coli* for nutritional mutants (Figure 2-2).

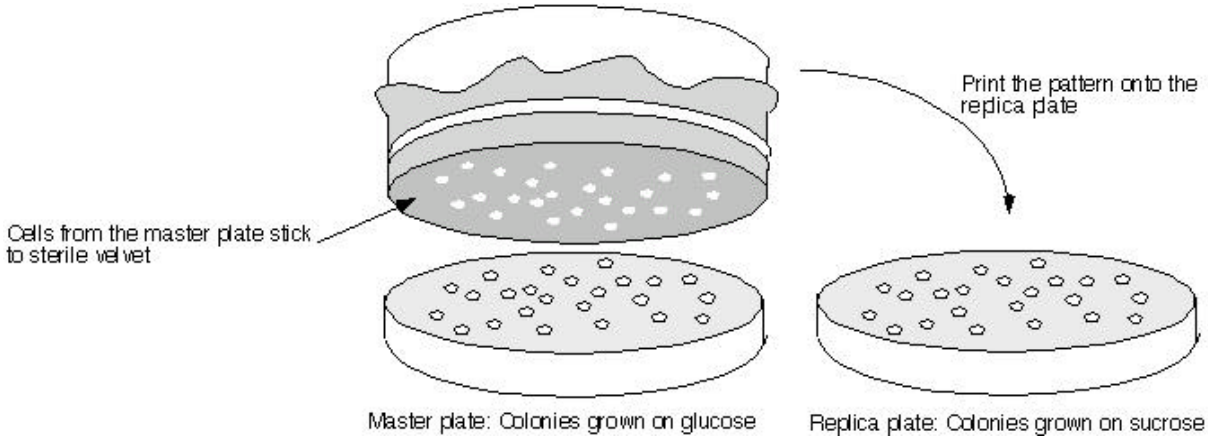


Figure 2-2 Replica plating

As simple as it looks, replica plating was a technical advance for microbial genetics because it allowed large numbers of colonies to be tested on many different kinds of media. About 200-500 colonies from each master plate can be copied to several different replica plates. Carlson et al only cared about glucose and sucrose, but other studies used replica plating to do parallel screening for a wide variety of auxotrophies or inability to utilize different carbon sources.

One of the first studies in biochemical genetics was the study of nutritional mutants of *Neurospora* (bread mold) by Beadle and Tatum. As mentioned above, this work changed the way we think about genes. Here, however, I'm bringing up their study to point out the value of replica plating. What Beadle and Tatum did, was to looked for mutations that could affect any biochemical pathways required for growth on minimal medium, and then figure out which ones were altered in any specific mutant. They treated *Neurospora* with X-rays to increase the frequency of mutants and then grew

2000 individual liquid cultures in complete medium from individual survivors of the X-ray treatment. They then used these cultures, each in its own test tube, to inoculate the same 2000 strains in minimal medium. They found 3 mutants that did not grow on minimal medium. Compare that to the 22,000 colonies that Carlson screened by replica plating.

Working on plates saved space, and moving >200 colonies at a time saved a lot of energy. Replica plating with velvet is less commonly used these days, but the idea behind replica plating is used in methods that lift colonies or plaques onto filters for hybridization or visualization with antibodies. Making large numbers of individual liquid cultures has also become more efficient with the introduction of robots that can inoculate microtiter plates (replica plating is still a lot cheaper, however).

Carlson identified mutant candidates as colonies that failed to grow on sucrose. Since the master plate required that they be able to grow on glucose, anything that was on the master and was missing on the replica could be specifically defective in sucrose utilization rather than a step later in the pathway.

Testing segregation - simple tetrad analysis

In the next step, Carlson did a genetic test to determine whether the defect in sucrose utilization was due to a change in one gene or more than one gene. To understand this test, we have to review some basic Mendelian genetics.

The mutants were isolated in haploid strains of mating type **a**. Carlson mixed each candidate with a sucrose-fermenting strain of mating type α. This can be done by spreading a patch of each strain on a plate with a toothpick or an inoculating loop. Where the two strains come in contact, they will mate and form diploids. The diploids and any remaining haploid yeasts that failed to find partners will grow.

The mixture is then patched onto sporulation media. When diploid yeast are starved for nitrogen, they go through meiosis and form dormant spores (See Figure 1-3). Individual asci can be isolated and dissected and the spores tested for their ability to grow on sucrose.

For any individual sucrose nonfermenting strain, there are the following possibilities:

1) A mutation in a single gene eliminates the ability to grow on sucrose

2) Mutations in two genes are present and

   a) both mutations are required to eliminate sucrose utilization or

   b) either mutation is sufficient to eliminate sucrose utilization

3) Variations on possibility #2 where more than two genes are involved

The probability of mutating two genes is roughly the square of the probability of mutating one gene. Thus, it is unlikely that more than one mutation was needed to make the cells into sucrose nonfermentors. However, rare events will be found if you look at enough events, and remember that they mutagenized the cells with a treatment that was harsh enough to kill 70% of the input.

In order to consider how the different possibilities will affect the tetrad analysis, we need to quickly review meiosis. For simplicity, figure 2-3 shows an idealized meiosis for a genome with two chomosomes, rather than the 16 chromosomes actually found in yeast.

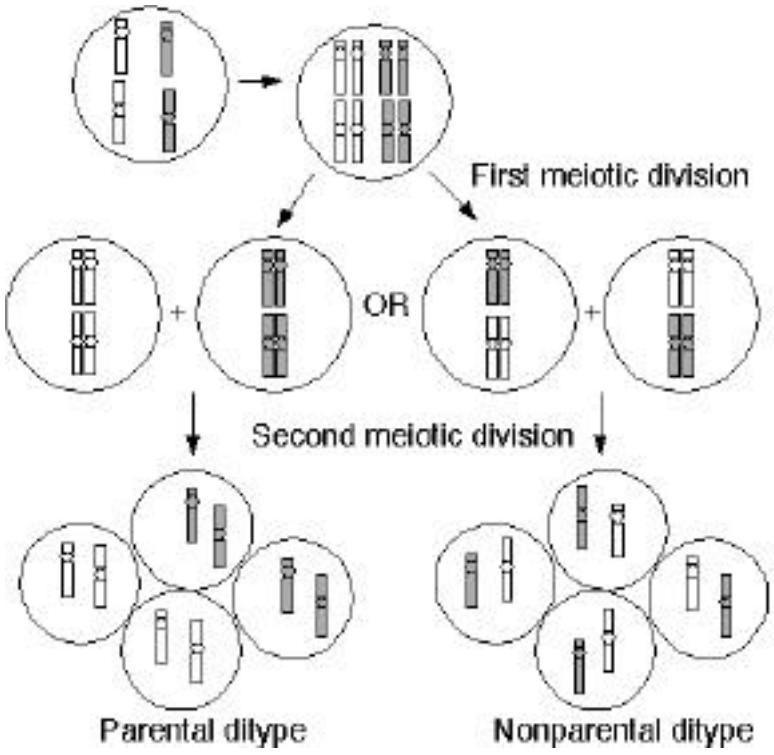


Figure 2-3 Simplified representation of meiosis. Two different chromosomes are shown.

Chromosomes from one parent are shown in white, while chomosomes from the other parent are shaded.

A diploid cell has two copies of each chromosome. In the case of yeast, one copy is from the **a** parent and the other is from the α parent. The two copies of each chromosome are called **homologs**. Meiosis begins after a round of DNA replication. The DNA strands for each chromosome stay together. Each DNA molecule is part of a **chromatid**. The two paired copies of the same chromosome are called **sister chromatids**. Meiosis involves the segregation of chromatids through two rounds of cell division. In meiosis I - the homologs pair and then move into different daughter cells. The sister chromatids stay paired. Before the first meiotic division, recombination can occur between homologs while they are paired; we will ignore recombination for now, but it is an important part of genetics that will come back later. A second cell division happens to generate the four haploid spores. During this second division, meiosis II, the sister chromatids split and migrate to different daughters

The diagram show a branch at meiosis I because there are two paths that the cell can follow in this simplified example. The two homologs of chromosome I don't care which way the two homologs of chromosome II go in the division step. Thus, the two cells that result from meiosis I can have chromosomes from the same parent; these will divide at meiosis II to give the parental ditype (PD). Alternatively chromosomes from different parents can segregate together at meiosis I, the second division gives four spores of the nonparental ditype (NPD). Although the diagram shows only two possible paths, the **independent assortment** of all 16 chromosomes means that there are many more possibilities than are shown.

Now let us return to the question of what results will be obtained from the tetrad analysis when we sporulate the diploid formed by mating a sucrose nonfermenter with a sucrose fermenter. First let's assume that there is only one mutation, in a gene we will call *SNF*. In the yeast genetic nomenclature, the wild-type version (or **allele**) is designated with capital letters, while the mutant version is given in lowercase letter. Using this notation, the diploid is *SNF/snf*, with the *snf* allele on some chromosome from the mutant parent and the *SNF* allele on the homolog from the wild-type parent. If

we follow a single chromosome through meiosis, it doesn't matter which path it takes. We will always get 2 *SNF* spores and 2 *snf* spores. This is called 2:2 segregation.

By contrast, consider what will happen if there are mutations in two genes, and both are necessary to prevent sucrose utilization. For simplicity, I will assume that the genes are on different chromosomes. Keep this assumption in the back of your mind. I will call the two genes *SNF1* and *SNF2*. The diploid is *SNF1/snf1, SNF2/snf2*. Now it matters which path is taken through meiosis. The pathway leading to the parental ditype will give 2 spores that are *SNF1,SNF2* and 2 spores that are *snf1, snf2*. These tetrads will give 2:2 segregation, just as we saw for a single gene. However, if the parental chromosomes segregate into different daughters at meiosis I, a different result will occur. The spores will reflect the nonparental ditype: 2 will be *SNF1, snf2* and 2 will be *snf1, SNF2*. All 4 of the spores will have the wild-type allele at one of the two *SNF* loci. By hypothesis, all of the spores will be able to grow on sucrose, and the segregation will be 4:0. Tetrads that do not segregate 2:2 will also be observed if you assume hypothesis 2b above, working out the expected phenotypes is left to the reader.

Independent assortment means that any given diploid cell has an equal probability of giving a PD or NPD tetrad. Carlson does not explicitly tell us how many tetrads she did for each candidate, but standard practice is to do several. Note the careful wording on page 29; she describes the ones that always segregated 2:2 as those which "could be attributed to a single nuclear mutation". In other words, 2:2 segregation does not prove that only one gene is involved, but failure to segregate 2:2 disproves the hypothesis that there is only a single mutation in a relevant gene.

Identification of nonsense mutations

Carlson did six separate experiments to get 31 mutants that segregated 2:2. Four of these turned out to be nonsense mutants. Before explaining how this was shown, we will go on another digression to review the kinds of mutations that can occur.

Mutations can be divided into classes based on different criteria. The major ways of classifying mutations are based on either the nature of the physical change in the DNA or the effect of the mutation on the structure or activity of the affected gene product.

Physical classification Point mutations are defined as mutations that change a single base-pair. Point mutations are divided into different classes based on the chemical nature of the change. **Transitions** are changes where a the mutant and wild-type base pairs have the purine and pyrimidine bases on the same strands. In other words, if a G:C base pair is changed so that the G is replaced by an A and the C is replaced by a T, then the mutation is a transition. Given these constraints, there are only two kinds of transitions G:C to A:T and A:T to G:C. Transitions are the most common type of point mutation, and the rates of transitions can be increased by a variety of mutagens. G:C to A:T transitions are significantly more common than A:T transitions among spontaneous mutations; the mutations induced by EMS are predominantly G:C to A:T transitions . A large fraction of G:C to A:T transitions are probably due to the spontaneous chemical deamination of C, which converts G:C to G:U. In *E. coli,* "hotspots" for G:C to A:T transitions are found in sequences where the C is methylated. Deamination of 5-methyl-C yields a G:T basepair, which is more resistant to repair than G:U. Mutations that change purines to pyrimidines are called **transversions**.

Mutations often affect more than one basepair; such mutations include insertions, deletions and rearrangements. Insertions and deletions can be mutations that add or remove anything from a single base pair to hundreds of kilobasepairs. Duplications are a special class of insertion mutations. Duplications and deletions can occur by common mechanisms involving repeated sequences. Other insertions are caused by **transposons**, DNA sequences that can be moved to new chromosomal locations by a variety of mechanisms. Transposons are a form of what has been called "selfish DNA" and can be thought of as parasites or symbionts at the DNA level. Transposons can be useful tools in genetic studies, as we may see later in the course. Rearrangements can occur that involve **inversion** of a DNA segment, or **translocation** of a piece of DNA from one chromosome to another chromosome.

In principle, most of these kinds of mutations are reversible, the process that gives a return to the wild-type sequence is called **reversion**. Isolating a true revertant is usually harder than isolating the original mutation. The target size is much smaller, and the kind of chemical change required is often

one that occurs at a lower frequency. Nevertheless, revertants can be isolated in systems where very large populations can be examined. With rare exceptions, deletions cannot be reverted; once a sequence is lost from the genome it is pretty much gone forever.

Effects on the structure of the gene product. Point mutations can have three kinds of effects on the primary structure of a protein encoded by a gene. Mutations that change the DNA sequence within a coding region without changing the protein sequence will be silent unless they have other effects; these would not be detected in a classical mutant hunt. However, silent mutations are important for studies involving natural or artificial evolution. A mutation can generate a detectable phenotype without changing the encoded amino acid sequence if it alters a regulatory sequence that affects the expression of the gene. Point mutations can also cause single amino acid changes by changing one codon into another. Point mutations that create nonsense codons (UAG, UAA or UGA in the mRNA) are called nonsense mutations. They lead to premature truncation of the protein due to termination of translation at the affected codon. Mutations to UAG codons are called **amber** mutations, mutations to UAA are called **ochre** mutations. Mutations that create UGA codons are called opal mutations, or sometimes they are just called UGA mutations.

The effects of nonsense mutations can be reduced by **suppressor** mutations in other genes. Genes encoding tRNAs can be mutated so that their anticodons can recognize the nonsense codon and insert an amino acid into the growing polypeptide chain. Suppressor tRNAs that recognize UAG codons are called amber suppressors, and tRNA mutations that recognize UAA codons are ochre suppressors. In *E. coli*, amber suppressors recognize only amber codons, while ochre suppressors recognize both ochre and amber codons.

If a tRNA suppressor can suppress the defect of a mutation, then that mutation is a nonsense mutation. However, the converse is not always true. For example, there are two reasons why an amber mutation would not be suppressed by an amber suppressor. First, suppression is not 100% efficient. The suppressor mutation does not change the intracellular concentration of the translational termination factors that normally release newly synthesized proteins from the ribosome. The extent of

suppression is therefore a kinetic competition between the termination factors and the mutant tRNA. Second, the amino acid inserted by each suppressor inserts is determined by the tRNA; there is no way that the cell can tell what the original codon was before it was mutated to amber. The inserted amino acid might be the same as what was found in the wild-type protein, or not. Thus, combining a particular suppressor with a particular amber mutant can lead synthesis of a protein with the normal activity, a protein that is still defective, or a protein that has a changed activity. In some cases, suppressing an amber mutation can make the mutant phenotype stronger; a recessive mutation will behave as a dominant mutation in the presence of the suppressor.

Insertions and deletions within a coding sequence can lead to truncation or extension of the protein by changing the reading frame of the mRNA. Ribosomes will enter a new reading frame at the site of the lesion, and will incorporate a different amino acid sequence from the original protein until a termination codon is reached. Nonsense suppressors do not suppress frameshift mutations; even though there may be readthrough of the stop codons, the proteins sequence is still drastically altered.

Effects on the activity of the gene product. The kind of physical change in the DNA determines how the gene product is affected, which, in turn, determines what will happen to the synthesis, degradation or specific activity of the protein encoded by the affected gene. Even before we know why the protein is affected, however, it is useful to classify mutations on the basis of their phenotypic effects on the function of the gene product.

Mutations that reduce the net activity of the gene product are called loss-of-function mutations. Occasionally you will see these referred to as **hypomorphs**. The extreme case for loss-of-function is where the activity is reduced to zero; this is the case where the gene is deleted, or for any nonsense or missense mutation that is indistinguishable from a deletion. A mutation that causes a complete loss of function is called a **null** mutation, and the null phenotype is generally the most useful for assigning the function of a gene.

Mutations can also cause a protein to be overexpressed or to become more active. Such mutations are called gain-of-function, or **hypermorphs**. Mutations can also change the activity of a gene product to something that the normal product doesn't do at all. Suppressor tRNAs are an example of

this kind of mutation. There are also examples of mutations in proteins that change the specificities of active sites so that new biochemical activities are found.

Getting back to the paper (at last!), Carlson designed some of her screens so that it would be easy to identify amber mutants. She did this by using a starting strain with genetic markers that were irrelevant to the initial screen, but which would be useful in later steps. This sort of planning ahead several steps is very important in molecular genetics. She did some of ther screens for *snf* mutants in strains that had nonsense markers in the *HIS4* and *LYS2* genes. The products of these genes are enzymes that are required for the biosynthesis of histidine and lysine. A strain that contains *his4* and *lys2* mutations will not grow unless histidine and lysine are added to the medium; it is **auxotrophic** for histidine and lysine.

Carlson took each of her *his4, lys2* double nonsense mutant strains, and plated them on medium lacking histidine and lysine. Colonies that grew up had to restore the function of both *HIS4* and *LYS2* in order to grow. There are two ways this could happen. The amber mutations in the two genes could simultaneously revert. Recall however that reversion is rare. Restoring growth in the absence of the added amino acids would require two rare events to occur; this frequency of this happening is the product of the frequencies of the two individual events. If a typical reversion event occurs at around 10^{-7}, then the double revertant will be found at 10^{-14}. A more frequent way to restore growth would be to mutate one of the tRNA genes to an amber suppressor.

Assuming that the colonies that grew contained amber suppressors, Carlson then tested each one for its ability to ferment sucrose. If the suppressed strains now grew on sucrose medium, she inferred that the original mutation that caused the inability to utilize sucrose was an amber mutation. Note that she did not apply selection for sucrose utilization. To confirm that the growth phenotypes were due to extragenic suppression, she crosed each “co-revertant” to a sucrose-fermenting strain and did another tetrad analysis. You should try to work out for yourself what kinds of segregants would occur if an amber suppressor was present vs. what would happen if both auxotrophic mutations were reversed by true reversion.

How many genes? Complementation tests

At this point (about 2/3 of the way down page 29), we know that Carlson found 31 independent mutants that were unable to utilize sucrose, and that at least 4 of them were amber mutants. If you can't explain why we know this, go back and reread the previous sections. The next questions she asked was: how many genes were affected by these mutations, and were any or all of these mutations in the previously identified *SUC2* gene?

To answer these questions, she did a series of **complementation** tests. The idea behind complementation is not hard to grasp, but it is one of the most important ideas in molecular genetics, and we will be using it later in the class. To illustrate how complementation works, we can start with Carlson's conclusions and what we know now and show how she got there from the data.

It turns out that there are many genes needed for yeast to utilize sucrose. They need invertase, which is encoded by the SUC genes, and they need a bunch of other gene products to regulate the synthesis of invertase. The other genes include a variety of snf genes; in this paper, Carlson identifies *snf1*, and in later papers she identifies 5 more *snf* genes. Here, she uses the complementation test to show that some of her mutants are in *SUC2* and others are in *snf1*.

Since both SUC2 and snf1 are needed to utilize sucrose, cells lacking either invertase or Snf1 protein will be unable to grow on sucrose medium. We know that all of her mutants from this paper are in either *SUC2* or *SNF1*. Consider what will happen when she starts making diploids from haploids derived from each mutant (Figure 2-4). Three things could happen:

Both parents could contain mutations in SUC2

Both parents could contain mutations in SNF1

One parent could be a suc2 mutant while the other was a snf1 mutant.

In the first case, the diploid will be *suc2*/*suc2*, *SNF1*/*SNF1*. Even though the cell can make Snf1 protein, there is no way to make invertase. Similarly, in the second case, the diploid can't make Snf1 since both copies of the *SNF1* gene are mutant. Neither of these diploids will be able to grow on sucrose; each is missing one of the components needed for sucrose utilization.

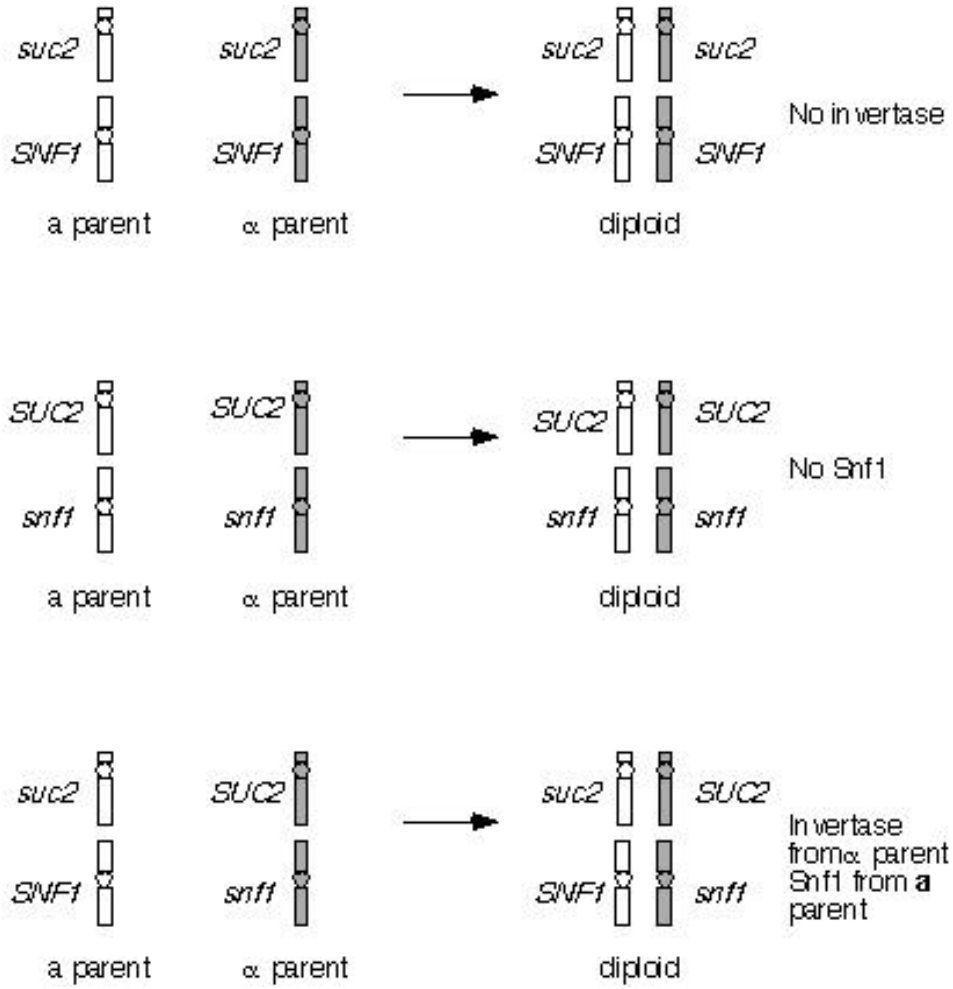


Figure 2-4. Complementation test. See text for explanation.

In the third case, however, each parent provides a different wild-type gene for the diploid. The *suc2* parent doesn't make invertase, but it can make Snf1. The *snf1* parent can't make Snf1, but it contains the wild-type information to make functional invertase. The wild-type genes from the two strains each provide what is missing in the other; in other words, they **complement** each other. Note that when we are in a hurry we will say that *suc2* and *snf1* complement each other; in fact, we don't mean that the mutant alleles provide the functions to utilize sucrose, it's what is not mutated that counts.

Table 1 in the Carlson paper shows the results for the complementation tests. As a control, they used a strain that was the same as the starting strain, except that it lacked the SUC2 gene (indicated by the *suc2*^{0} in the first column). The mutations could be divided into two classes based on whether or not they complemented *suc2*^{0}. The majority of the alleles, including the four amber mutants, behaved as if they were in *SUC2*. Those that complemented are presumed to be in new genes required for sucrose utilization, several of them, designated Group 2 in the paper, complemented *suc2*^{0}, but did not complement one another. A group of mutants that don't complement one another is known as a **complementation group**. Carlson named the gene defined by this complementation group *SNF1*.

Since the paper contains real rather than idealized data, you should note that complementation was observed within the group 1 alleles. The four group 1B alleles complemented *suc2-437*, the single group 1C allele when complementation was assayed by the ability to utilize sucrose. Nevertheless, Carlson et al classified these alleles as mutations in *SUC2*. To understand why, examine the overall pattern of complementation. Neither class 1B nor class 1C complemented any of the other *SUC2* mutations, including the control allele, which is presumed to be a null. Any model for what is going on has to explain these observations. If class 1B and class 1C mutations really are in SUC2 (the data certainly support the idea that they are), then this is a case on **intragenic complementation**. In the Discusson section of the paper, Carlson states that the intragenic complementation suggests that the *SUC2* gene product is oligomeric. As it turns out, *SUC2* encodes invertase, which is homodimeric, so Carlson was right. However, the paper does not really explain how she came to make this suggestion.

Intragenic complementation suggests that a gene encodes a multimeric protein based on the following kind of model (Figure 2-5). Imagine a dimeric protein where the symmetry of the dimer places residues A and A' close in space but not tightly packed in the dimer interface. Residues B and B' are similarly disposed. Now imagine that a mutations introduce a bulkier side chain at positions A and B. In the strain with a mutation at position A only, all of the monomers will have the larger sidechain at A, and the bulk of the larger side chain will prevent formation of stable homodimers. Similarly, increasing the size of the side chain of B will prevent the B and B' positions from

approaching each other. In cells that are heterozygous for the two mutations, however, less bulk is introduced at each position, and dimers are able to form. These active heterodimers give a wild-type phenotype, so intragenic complementation is observed.

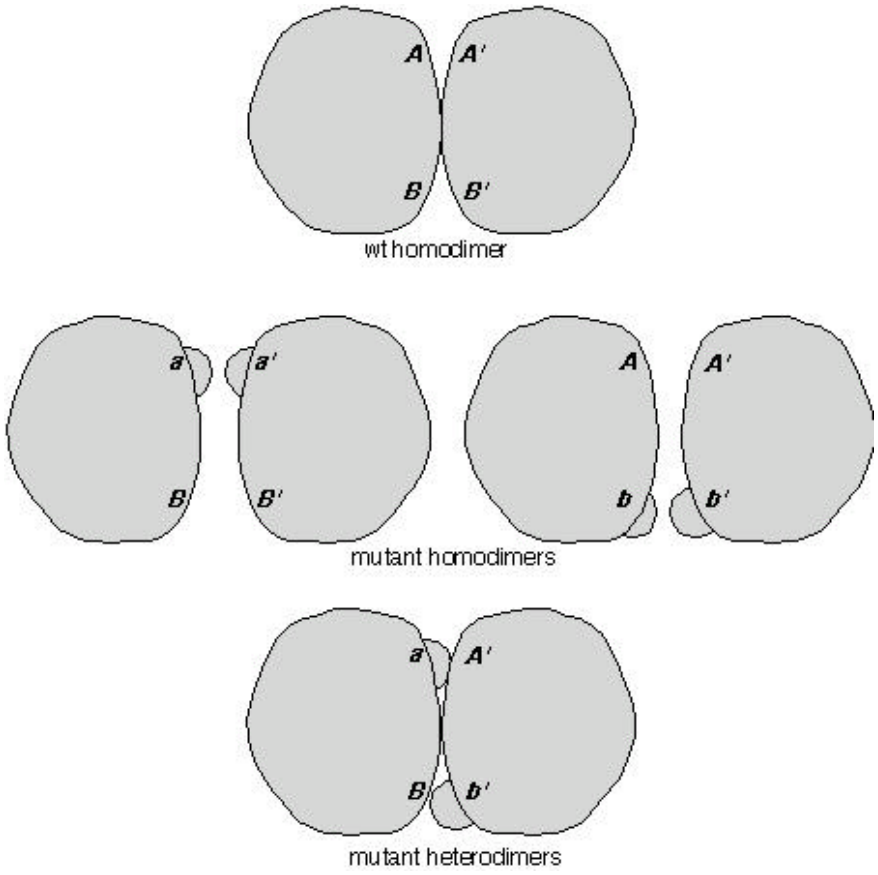


Figure 2-5. A model for intragenic complementation. See text for details

This is only one kind of mechanistic model for how intragenic complementation could occur. Here's a completely different model that works for a monomeric enzyme (Figure 2-6). Imagine that the reaction catalyzed by the enzyme occurs in two steps, where R represents the reactant, I a diffusible intermediate, and P is the product. The mutation at site A affects the active site for the first step in the pathway, while the mutation at site B affects the second step. Haploid cells with mutations at A can't make I, but if I was present, they could convert it to P. Haploid cells with mutations at B can make I,

but can't convert it to P. Heterozygotes will make both forms of the enzyme, so the one form will convert R to I and the other will take I to P. Intragenic complementation is observed without physical contact between the protein molecules.

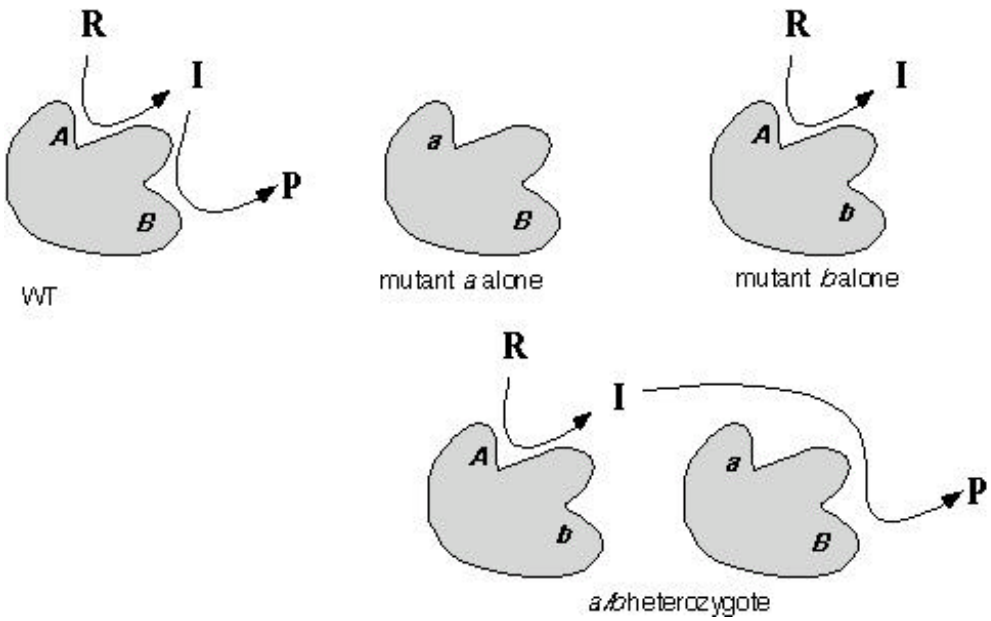


Figure 2-6. Alternative model for intragenic complementation. See text for details

These two models illustrate both a strength of genetics, a weakness of geneticists and a theme that will recur at various times in the course. The intragenic complementation observed by Carlson indicates that something interesting is going on with those particular combinations of alleles. Reasoning such as that used to illustrate Figure 2-5 suggests that the genetic interaction between the alleles could reflect a physical interaction between the mutant proteins. This suggestion will seem even more reasonable if there are precedents in the literature. However, Figure 2-6 presents an equally reasonable model which is independent of the oligomerization state of the affected protein. What does this mean for the idea that observing intragenic complementation suggests oligomerization?

It is difficult to come up with a genetic experiment that will distinguish between the two models in the absence of additional data from physical biochemistry. As molecular biologists, we should have no problem with that; genetics and biochemistry are both in our armory of approaches to problem solving.

There is nothing from the data that says that either of these models is correct. There couldbe other models that fit the observation of intragenic complementation that we just haven't thought of yet; falsifying the second model does not prove the first one. The burden of proof in science is not set up so that models are proven by the lack of imagination of the experimenters or the reviewers of his or her publications.

It turns out, of course, that SUC2 encodes invertase, and that invertase is an oligomeric protein. Does that mean that something like the model shown in Figure 2-5 is actually the case? Not really.

Recombination and segregation patterns

A diploid made from a suc2 mutant and a snf1 mutant is able to utilize sucrose, since complementation is observed. Such a diploid will have the genotype *SUC2/suc2*, *snf1/SNF1*. This means that its genotype fits hypothesis 2b on page 2-10: it contains two mutations on different chromosomes, each of which affects a gene needed for sucrose utilization. Earlier, I asked you to work out how such a diploid would behave in a tetrad analysis (page 2-12); now I will give away the answer and you can see if you had it right.

With respect to SUC2 and SNF1, each tetrad has an equal probability of having either a PD or NPD arrangement of spores. In the NPD tetrads, the chromosomes carrying the SUC2 and SNF1 loci will separate at meiosis I so that each daughter contains one chromosome from each parent. This means that one daughter will carry both mutations and the other daughter will carry none. As we have seen above, this leads to 2:2 segregation. In the PD tetrads, the *suc2* and *snf1* mutations would go into different cells at meiosis I, and you would expect 2 *suc2* mutant spores and 2 *snf1* mutant spores. None of the spores would grow on sucrose. Since the PD and NPD lineages are equally probable, we would expect half the tetrads to segregate 2:2 and half to segregate 0:4. This is not what was observed. Among five tetrads analyzed, Carlson never observed any where all of the spores failed to grow on sucrose. What happened to the PD tetrads?

In describing segregation patterns, I have been ignoring something that is very important: **recombination**. Recombination can be defined as a process that involves breaking and joining two

DNA sequences such that the products of recombination contain DNA strands where one part comes from atoms that orignated in one parent molecule, and another part has atoms from the other parent molecule. For example, in molecular cloning, the combination of a vector and insert is a **recombinant** DNA molecule because there is a junction between the vector and insert DNA sequences within each strand of double stranded DNA. This is an example of an artificial kind of recombination. The recombination that we are concerned with here is **homologous recombination**, which is also called crossing over(Figure 2-7).

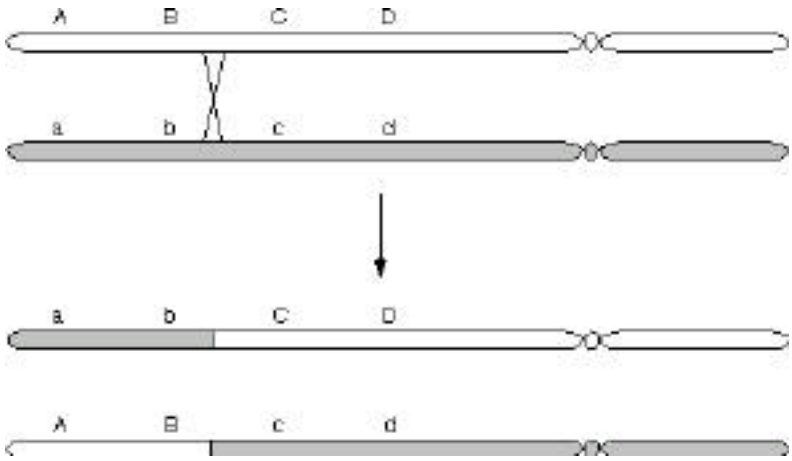


Figure 2-7 Homologous recombination between two chromatids.

The molecular mechanism of homologous recombination will not be addressed here. The important things to remember about homologous recombination are illustrated in Figure 2-7. Homologous recombination can occur anywhere the sequences match between two DNA molecules. In the illustration, the white and grey chromatids are from homologs that are paired in meiosis I. Since the chromatids are from homologs, their sequences will be the same everywhere except where there is a genetic difference between the parents. In other words, the only differences between the grey chromatid and the white chromatid will be at base-pairs that represent different alleles. These allelic differences may be in genes or they may be in intragenic regions; many of the differences will have no phenotype that can be detected by looking at the colonies or testing their ability to grow under different

conditions. Some loci are indicated on the chromatids by different letters; upper and lower case indicates different alleles.

In the illustration, a crossover (indicated by the crossing lines between the chromatids) occurs where homologous recombination breaks the DNA in both chromatids between the B and C loci. The chromatids then trade the segments carrying the DNA to the left of the break and rejoin the segments. The result is to restore two intact chromatids, but genetic markers are now attached to each other differently. The order of the loci is the same, but which alleles are on the same DNA molecule has changed. You can determine the order of markers on the products by starting at the left end of one of the chromosomes, following one of the lines to cross over to the other chromosome, and then continuing rightward until you reach the end.

Note that homologous recombination can also occur in bacteria, which never go through meiosis, or in mitotic eukaryotic cells (we'll see that in the next subsection). Homologous recombination between viruses can occur, or between recombinant DNA molecules and chromosomal DNA in a wide variety of organisms. This will become important later in the course, especially when we discuss knocking out genes (section XX, not written yet).

What does this have to do with the segregation of *SUC2* and *SNF1*? Consider what will happen when crossovers occur during meiosis I. Figure 2-8 shows what will happen if there is a crossover between homologs carrying *SUC2*, but the kind of result you will get is similar if the crossover occurs on the pair of homologs carrying the *SNF1* locus.

Pairing of the homologs and recombinatin occurs after replication. Recall that at this point each homolog is comprised of a pair of sister chromatids. A single recombination event between homologs will only involve one chromatid from each partner. As a result, you end up with pairs of chromatids where the sisters, which remain attached until meiosis 2, are no longer identical. For the DNA that is distal to the centromere, the chromatids are heterozygous for any allelic differences between the parents.

In Figure 2-8, a recombination between two chromatids on the upper homolog generates sister chromatids that differ at the SUC2 locus. The pairs can segregate in either a PD or NPD lineage, but

now the genetic composition of the spores that result from meiosis II has changed. Both lineages give a the same phenotypic pattern among the spores, 3 spores that will give rise to strains that can't use sucrose, and 1 that can.

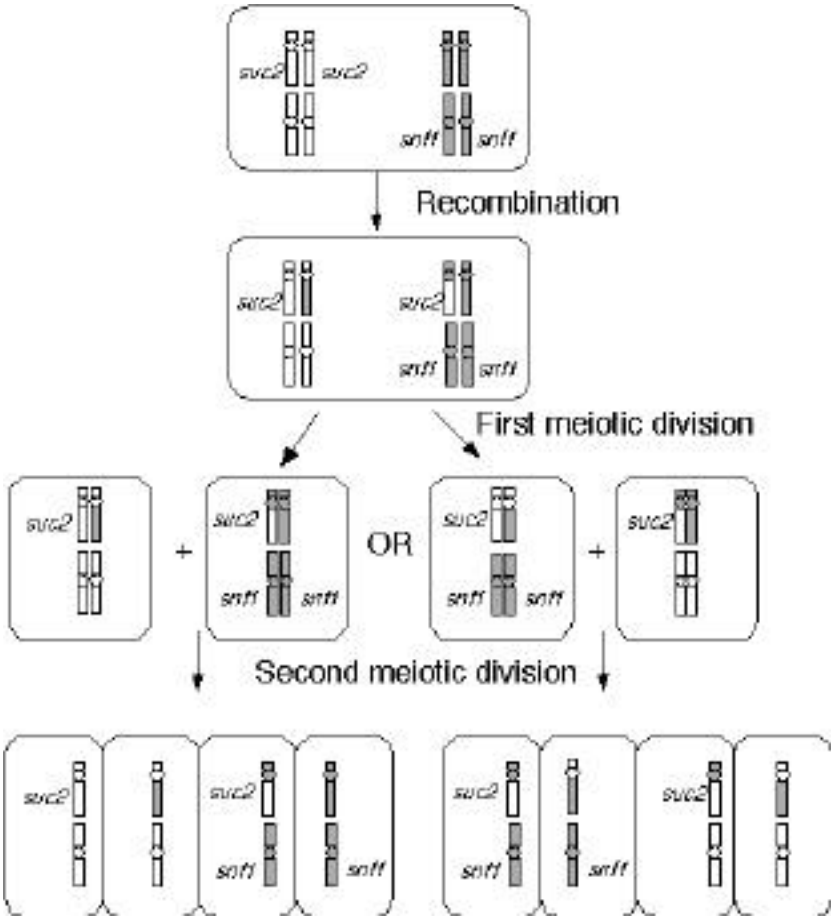


Figure 2-8. Segregation of markers during meiosis with a crossover event.

Carlson looked at 5 tetrads and always found at least one spore that germinated and grew on sucrose, and she interpreted this result to mean that *SUC2* and *SNF1* were not tightly linked. We've just seen what happens in the extreme case of markers being unlinked. In the absence of recombination, we would see 2:2 and 0:4 (putting the sucrose utilizers first) tetrads. Recombination allows the formation of 1:3 tetrads. The fraction of each kind of tetrad will depend on the likelihood of recombination between *SUC2* and the centromere and between *SNF1* and the centromere.

What would happen if *SUC2* and *SNF1* were on the same chromosome? In the absence of recombination, all of the copies of that chromosome would contain one of the mutations or the other. Since there would not be any way in which a wild-type chromosome could be inherited by any of the spores, none of the spores would give rise to cultures that could utilize sucrose.

Recombination between the parental chromosomes could generate one chromosome with both mutations and the other with both wild-type alleles (Figure 2-9). The *SUC2, SNF1* chromosome would then give rise to a spore that could grow on sucrose. Crossovers occur more or less randomly along chromosomes, but the probability of a crossover at any specific base-pair is relatively low. Thus, the probability of a crossover between two markers on the same chromosome is proportional to their separation. If the markers are very close to one another, recombinant chromosomes are not observed without examining enormous numbers of progeny (see below). If two markers are so close together so that it is difficult to observe recombination between them, they are said to be **tightly linked**.

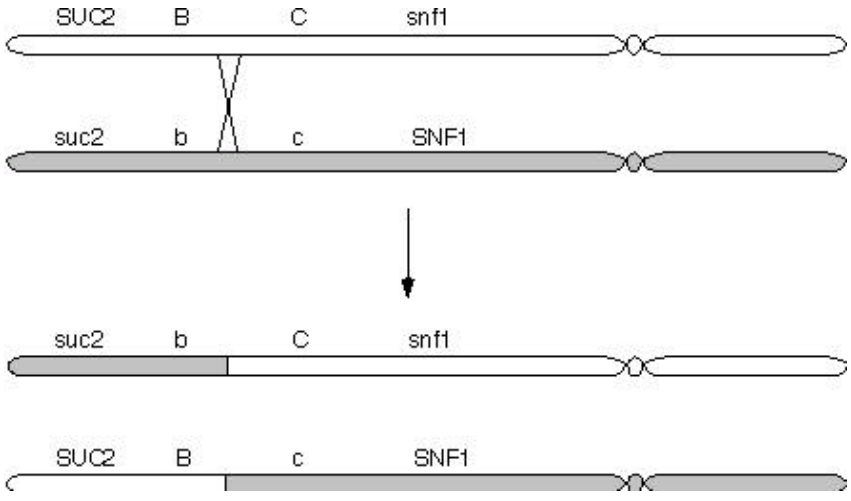


Figure 2-9 Crossover between two mutant chromosomes based on the hypothesis that *SUC2* and *SNF1* might be linked.

If SUC2 and SNF1 were tightly linked, they would behave as if there was little or no

recombination between them. Since Carlson found that wild-type spores could be found in all of the tetrads she examined, she could safely conclude that the two genes were not tightly linked.

Linkage and mapping

As noted above, the probability of a crossover event between two markers on the same DNA molecule increases as they get further apart. This means that the distance between two markers can be measured by quantitating the number of recombination events. For this, you need a way to tell that the recombination happened.

The measurement of genetic distances by looking for recombinants goes back to early experiments in *Drosophila* genetics. Mutant flies were isolated that had visually obvious phenotypes such as altered eye or body color or changes in the shapes, number or positions of body parts. If the genes that were affected by the mutations were on the same chromosomes, then the parental genotypes tended to stay together, just as in the hypothetical example shown above. If the one of the parental flies had white eyes and normal wings, while the other had normal eyes and miniature wings, fly geneticists could determine whether or not the mutations were linked or not and measure the distance, by counting the number of recombinant flies from an appropriate set of breeding experiments. In this case, which was examined by Sturtevant, about a third of the flies had recombinant chromosomes, which gave rise to either wild-type flies or white-eyed flies with miniature wings. We will examine this kind of mapping between genes in more detail when we get to map-based cloning.

The point I want to make about the different genes examined is that the fraction of recombinants is large enough so that it can be determined simply by looking at random organisms from the appropriate experiment, and counting all of the organisms in a given phenotypic class. This works as long as the markers are far enough apart that one can get a statistically significant number of recombinants from scoring a reasonable number of organisms.

Carlson sought to examine how the mutations within *SUC2* were organized. Since these mutations were already known to be in the same gene based on the complementation tests above, we

would expect the genetic distance between any pair of markers to be small, and she has to examine a lot of cells to see any recombination events at all.

The kind of fine structure intragenic mapping of mutations was pioneered by Seymour Benzer in the mid 1950s. Using phage T4 as the model system, Benzer realized that high resolution genetic maps could be determined if the system met two conditions: you need to have 1) an organism that grows to very large populations (preferably quickly and in a manageable amount of lab space), and 2) an efficient way to find and count the rare recombinants among a sea of nonrecombinants. In the T4 experiments, Benzer had mutations in a gene called rII, which was required for growth on one strain of E.coli, K-12, but was not required in the other, B. Benzer could infect B with a mixture of two rII mutants and recover the phage that grew. From this population, he could determine the total phage concentration by plating dilutions on B, and determine the number of recombinants by plating the same phage stock on K-12. Each recombination event between two rII mutants will give two products of the crossover, one wild-type phage genome and the one double mutant. The wild-type phage will grow on K-12, while the parents and the double mutants will not. This made it easy to detect events as rare as 1 in 10^9. This high level of sensitivity allowed Benzer to obtain recombinants between mutations at adjacent base-pairs, and allowed him to answer some very fundamental questions about the structure of genes. Benzers papers are very elegant and beautifully reasoned, but the phage genetics and mathematical analysis are beyond the scope of this digression. However, if you would like to read a less technical description of the work, he wrote a Scientific American article that appeared in January of 1962.

While Carlson was following in Benzer's footsteps by mapping mutations within *SUC2*, the details of the experiment were different, and the resolution obtained by Benzer was not needed. Carlson looked for recombination between the two homologs in diploid cells. To stimulate recombination, she irradiated the cells with UV light using a sunlamp. UV light causes DNA damage, and both prokaryotes and eukaryotes induce recombination enzymes in order to repair the damage. Crossovers that occur between the locations of the two mutations tested will generate two products; the *SUC2* gene on one chromosome will have both mutant alleles, while the other chromosome will have

no mutations. The wild-type chromosome will be able to make the *SUC2* gene product, invertase, and will be able to grow on sucrose.

Thus, to do the experiment, we infer that Carlson did something like this:

- Start with a diploid made from a cross between two *suc2* alleles. Grow a liquid culture.
- Irradiate the culture with the sunlamp, taking aliquots out at different times.
- For each timepoint:

  - plate part of in onto glucose to determine how many survivors there were.

  - plate part of it onto sucrose to determine how many recombinants there were.

