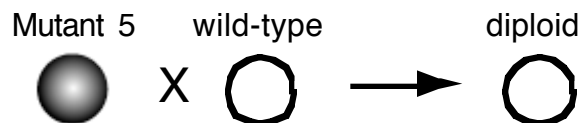
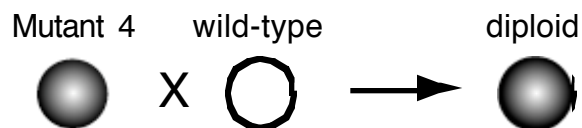
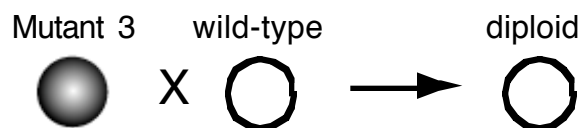
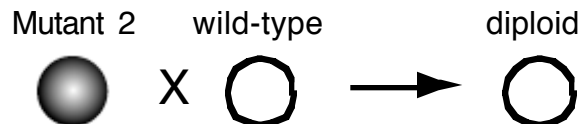
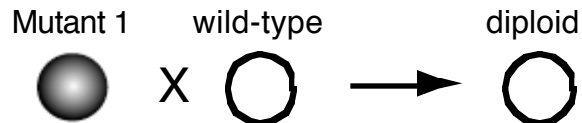


7.03 Problem Set 1

Due before 5 PM on Thursday, September 18

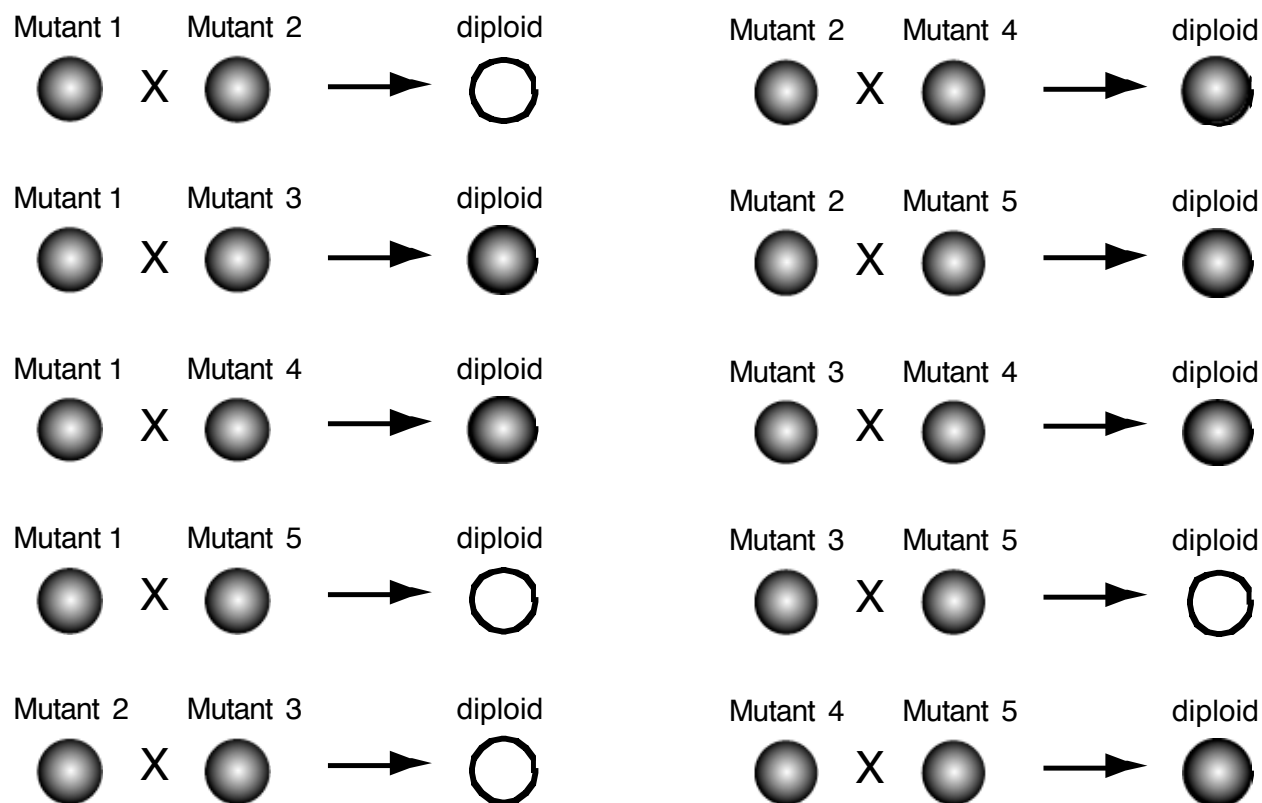
Hand in answers in recitation section or in the box outside the class

1. You have isolated a set of five yeast mutants that form dark red colonies instead of the usual white colonies of wild-type yeast. You cross each of the mutants to a wild-type haploid strain and obtain the results shown below

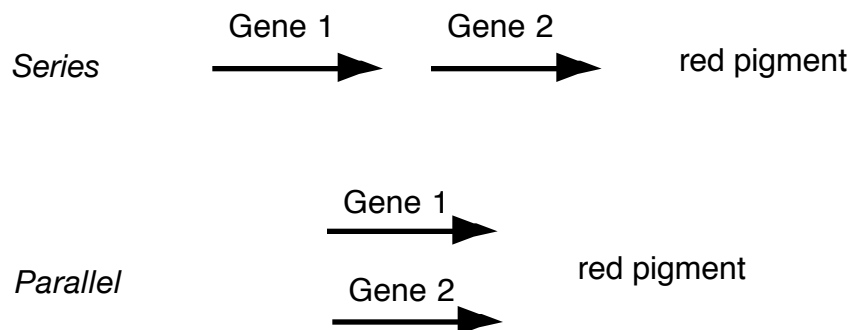


(a) What do these results tell you about each of the mutants?

(b) Next you cross each haploid mutant strain to a different haploid mutant of the opposite mating type. From the results shown below deduce as much as you can about which mutations lie in the same gene. Clearly state any remaining ambiguities and suggest some general ways that the ambiguities might be resolved.



2. In this problem we will explore some of the many ways that mutations in two different genes can interact to produce different Mendelian ratios. Consider a hypothetical insect species that has red eyes. Imagine mutations in two different unlinked genes that can, in certain combinations, block the formation of red eye pigment yielding mutants with white eyes. In principle, there are two different possible arrangements for two biochemical steps responsible for the formation of red eye pigment. The two genes might act in *series* such that a mutation in *either* gene would block the formation of red pigment. Alternatively, the two genes could act in *parallel* such that mutations in *both* genes would be required to block the formation of red pigment.



Further complexity arises from the possibility that mutations in either gene that lead to a block in enzymatic activity could be either dominant or recessive.

a) Such considerations yield the following six possibilities:

- 1) Pathways in series with recessive mutations in both genes.
- 2) Pathways in series with a recessive mutation in one gene and a dominant mutation in the other.
- 3) Pathways in series with dominant mutations in both genes.
- 4) Pathways in parallel with recessive mutations in both genes.
- 5) Pathways in parallel with a recessive mutation in one gene and a dominant mutation in the other.
- 6) Pathways in parallel with dominant mutations in both genes.

First let's consider the possible outcomes from these possibilities in crosses between a wild type insect with red eyes and a true breeding white eyed strain with mutations in both genes. For each of the six possible cases outlined above, determine the phenotype of the F1 progeny (either red or white eyes) and the expected phenotypic ratio of red to white eyed insects in the F2.

b) You obtain a true-breeding white-eyed mutant, which you cross to wild type. All of the progeny have normal looking red eyes. Crosses among these F1 insects yield 24 progeny; 5 with white eyes and 19 with red eyes. Determine whether this data is consistent with each of the six possibilities outlined in part (a). For your answer also consider the additional possibilities that the true-breeding white eyed phenotype is due to: 7) a recessive mutation in a single gene, or 8) a dominant mutation in a single gene. Use the table below of chi-squared probabilities for any statistical tests.

<i>p</i> value:	.995	.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005
df = 1	.000	.000	.016	.46	2.7	3.8	5.0	6.6	7.9
df = 2	.01	.05	.21	1.4	4.6	6.0	7.4	9.2	10.6
df = 3	.07	.22	.58	2.4	6.3	7.8	9.3	11.3	12.8

3. Genes that control coat color in mammals represent some of the best early examples of genes with multiple alleles exhibiting different phenotypes. A classic example is the C gene in hamsters. Three of the known alleles of this gene are C, c^k and c^a . Each supplies a different amount of black pigment to the coat hair. The phenotypes of three different homozygous strains are as follows.


<u>Genotype</u>	<u>Phenotype</u>
C C	Black fur
$c^k c^k$	Sepia colored fur
$c^a c^a$	Albino fur


(a) Based on the concepts of dominance, recessivity, and complementation, describe a set of crosses and the interpretation of their outcomes that would have allowed early mammalian geneticists to conclude that c^a and c^k are actually alleles of the same gene rather than recessive mutations in two different coat color genes. (To answer this question you will need to make some reasonable assumptions about the coat color of heterozygous strains)

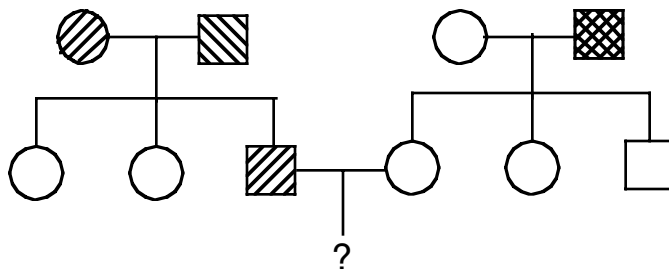
(b) Now describe a different set of crosses and their interpretation that would have allowed geneticists to come to the conclusion that c^a and c^k were likely to be alleles of the same gene using the concept that alleles of the same gene will be tightly linked.

4. In the following pedigree shows the segregation of two different recessive traits.

 = female showing trait 1 only

 = male showing trait 2 only

 = male showing both traits



- (a)** Assuming that both traits are due to linked autosomal genes that are 10 cM apart, calculate the probability that the indicated child will have both recessive traits.
- (b)** Derive a general equation relating the probability that the child will have both traits expressed as a function of the distance between the markers in cM.
- (c)** Assuming that both traits are 10 cM apart on the X chromosome, what is the probability that a male child will have both recessive traits?
- (d)** Assuming that both traits are 10 cM apart on the X chromosome, what is the probability that a female child will have both recessive traits?

7.03 Problem Set 1 Solutions

1.

a. Crossing each yeast haploid mutant to wild-type will tell you whether the mutation is recessive or dominant to wild-type. If the diploid is wild-type phenotype, then the mutation is recessive to wild-type. If the diploid is mutant phenotype, then the mutation is dominant to wild-type. Therefore, **mutants 1, 2, 3 and 5 are recessive to wild-type and mutant 4 is dominant to wild-type.**

b. These crosses are actually complementation tests, allowing us to determine whether the two mutants crossed have mutations in the same gene or in different genes. Mutants that do not complement each other (progeny have mutant phenotype) are said to be in the same complementation group and therefore in the same gene. Based on non-complementation of the recessive mutations, we can conclude **that mutants 1 and 3 form one complementation group and are mutations in the same gene (gene A).** Likewise, we can conclude that **mutations 2 and 5 form a second complementation group and are mutations in a second gene (gene B).** We are unable to determine whether mutant 4 represents a mutation in a new gene or a mutation in one of the other two genes because it is a dominant mutation.

The first ambiguity is whether mutant 4 has a mutation in gene A or B, or whether it represents a unique gene. To determine this, you would cross mutant 4 to one mutant from each of the two complementation groups, generating a diploid. You would then sporulate and look at the segregation pattern of the white/red phenotypes in the resulting haploids. If any of the haploids form white wild-type colonies, then gene 4 is not likely to be in the same gene as the mutant it was crossed with because a crossover is likely to have occurred.

2.

a. (wild-type genes are represented as “+” and mutant as “-“. Also, when one mutation is dominant, it is mutation 1.)

possibility 1: **F1 will have red eyes** because mutations are recessive and F1 is heterozygote: 1-/1+, 2-/2+.

To determine expected F2 phenotypic ratio, we must determine the probability that F2 will have one of the eye colors:

$P(\text{F2 white eyes}) = p(\text{homologous for 1- and not for 2-}) \times p(\text{homologous for 2- and not for 1-}) + p(\text{homologous for both})$

$p(\text{F2 white eyes}) = (1/4)(3/4) + (1/4)(3/4) + (1/4)(1/4) = 7/16$

therefore, F2 will be **7 white: 9 red**

possibility 2: **F1 will have white eyes** because there is a dominant mutation in series. Even though only one of the mutant phenotypes will be expressed in the F1 heterozygote, the F1s will have white eyes because both wild-type gene products are required to make red color.

$$\begin{aligned} p(\text{F2 red eyes}) &= p(\text{homologous for 1+ and (homologous or heterozygous for 2+)}) \\ p(\text{F2 red eyes}) &= (1/4)(1/4 + 1/2) = 3/16 \end{aligned}$$

therefore, F2 will be **13 white: 3 red**

possibility 3: **F1 will have white eyes** because heterozygous for two dominant mutations in series.

$$\begin{aligned} p(\text{F2 red eyes}) &= p(\text{homologous for 1+ and homologous for 2+}) \\ &= (1/4)(1/4) = 1/16 \end{aligned}$$

therefore, F2 will be **15 white: 1 red**

possibility 4: **F1 will have red eyes** because F1 is heterozygous for recessive mutations

$$\begin{aligned} p(\text{F2 white eyes}) &= p(\text{homozygous for 1- and homozygous for 2-}) \\ &= (1/4)(1/4) = 1/16 \end{aligned}$$

therefore, F2 will be **1 white:15 red**

possibility 5: **F1 will have red eyes** because even though there is a dominant mutation, it is in parallel, so can be compensated for by the wild-type allele of the other gene.

$$\begin{aligned} p(\text{F2 white eyes}) &= p(\text{homologous for 2- (and homologous or heterozygous for 1-)}) \\ &= (1/4)(1/2 + 1/4) = 3/16 \end{aligned}$$

therefore, F2 will be **3 white: 13 red**

possibility 6: **F1 will have white eyes** because both mutations are dominant in parallel.

$$\begin{aligned} p(\text{F2 white eyes}) &= p(\text{(heterozygous or homozygous for 1-) and (heterozygous or homozygous for 2-)}) \\ &= (1/2 + 1/4)(1/2 + 1/4) = 9/16 \end{aligned}$$

therefore, F2 will be **9 white: 7 red**

b. to determine which of the possibilities is consistent with the observed data, we must do a chi-squared test using the expected ratios found in part (a). The $df = 1$. we can reject hypothesis if $p < 0.05$

$$\chi^2 = \sum [(O-E)^2/E]$$

possibility 1: expected ratio is 7 white: 9 red, so out of 24 total F2's we would expect 10.5 white and 13.5 red.

$$\chi^2 = (5-10.5)^2/10.5 + (19-13.5)^2/13.5 = 5.12$$

by using the provided table and $df=1$, you get $0.025 > p > 0.01$ therefore we can **reject possibility 1.**

possibility 2: expected ratio is 13 white: 3 red, so out of 24 total F2's we would expect 19.5 white and 4.5 red.

$$\chi^2 = (5-19.5)^2/19.5 + (19-4.5)^2/4.5 = 57.5$$

by using the provided table and $df=1$, you get $0.005 > p$ therefore we can **reject possibility 2.**

possibility 3: expected ratio is 15 white: 1 red, so out of 24 total F2's we would expect 22.5 white and 1.5 red.

$$\chi^2 = (5-22.5)^2/22.5 + (19-1.5)^2/1.5 = 217.78$$

by using the provided table and $df=1$, you get $0.005 > p$ therefore we can **reject possibility 3.**

possibility 4: expected ratio is 1 white: 15 red, so out of 24 total F2's we would expect 1.5 white and 22.5 red.

$$\chi^2 = (5-1.5)^2/1.5 + (19-22.5)^2/22.5 = 8.71$$

by using the provided table and $df=1$, you get $0.005 > p$ therefore we can **reject possibility 4.**

possibility 5: expected ratio is 3 white: 13 red, so out of 24 total F2's we would expect 4.5 white and 19.5 red.

$$\chi^2 = (5-4.5)^2/4.5 + (19-19.5)^2/19.5 = 0.068$$

by using the provided table and $df=1$, you get $0.9 > p > 0.5$ therefore we **cannot reject possibility 5.**

possibility 6: expected ratio is 9 white: 7 red, so out of 24 total F2's we would expect 13.5 white and 10.5 red.

$$\chi^2 = (5-13.5)^2/13.5 + (19-10.5)^2/10.5 = 12.23$$

by using the provided table and $df=1$, you get $0.005 > p$ therefore we can **reject possibility 6.**

possibility 7: if true-breeding white eyed phenotype was due to a recessive mutation in a single gene, we would expect $1/4$ of the F2 progeny to have the white eyed phenotype.

$$\chi^2 = (5-6)^2/6 + (19-18)^2/18 = 0.222$$

by using the provided table and $df=1$, you get $0.9 > p > 0.5$ therefore we **cannot reject possibility 8.**

possibility 8: if true-breeding white eyed phenotype was due to a dominant mutation in a single gene, we would expect $3/4$ of the F2 progeny to have the white eyed phenotype.

$$\chi^2 = (5-18)^2/18 + (19-6)^2/6 = 37.5$$

by using the provided table and $df=1$, you get $0.005 > p$ therefore we can **reject possibility 9**.

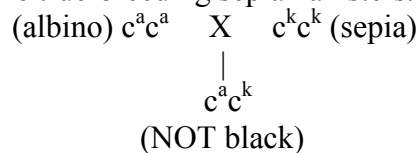
3.

a. First, we would determine whether c^a and c^k were dominant or recessive alleles by crossing the true-breeding sepia and albino hamsters to true-breeding black hamsters.



From the resulting F1 progeny phenotype, we can conclude that the c^a and c^k alleles are both RECESSIVE to the wildtype C allele.

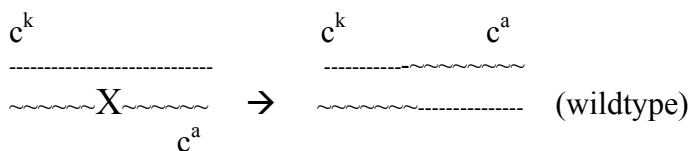
Since both c^a and c^k are recessive, we can perform a COMPLEMENTATION TEST to determine whether the different mutations lie on the same gene by mating the true-breeding albino hamsters to the true-breeding sepia hamsters.



From the resultant F1 phenotype of this cross, we note that c^a and c^k do NOT complement each other. This indicates that c^a and c^k are different alleles of the same gene.

b. We would mate the true-breeding $c^a c^a$ hamsters to the true-breeding $c^k c^k$ hamsters to produce heterozygous $c^a c^k$ F1 progeny.

We would then cross heterozygous $c^a c^k$ hamsters to other heterozygous $c^a c^k$ hamsters. From the fraction of F2 progeny that have the wildtype black phenotype, we can determine whether or not c^a and c^k are alleles of the same gene.



When crossing over occurs, an allele with neither mutation is formed. The passage of this wildtype chromosome onto an offspring will result in a hamster with black fur.

If c^a and c^k are mutations on separate genes, they will be further away from each other, hence making crossing over between them more likely. The presence of some black F2 hamsters will indicate that crossing over had occurred and that the mutations are on different genes.

More specifically, when the mutations are on different genes, the ratio of the F2 hamsters should be as follows:

9 black (heterozygous wt/mutant): 3 sepia: 3 albino: 1 black (double mutant)

If c^a and c^k are alleles of the same gene, they will be very close to each other and crossing over will rarely occur between them. The absence (or rare occurrence) of black F2 hamsters in a large population will indicate that crossing over did not (or rarely) occurred. This indicates that the mutations are on the same gene.

More specifically, when the mutations are on the same gene, the ratio of the F2 hamsters should be as follows:

1 sepia: 2 not black (double mutant): 1 albino

4. Remember that the traits are recessive!

a. A = wt B = wt
 a = trait 1 b = trait 2

Dad's genotype = aaBb

*He's affected with trait 1, so he must be aa

*He's not affected with trait 2, but he must have gotten one b from his father who was bb.

So he must be heterozygous for trait 2 with Bb

Dad's chromosomes look like this:

a B (from mother)
a b (from father)

Mom's genotype = AaBb

*Her father is aabb, so she must have gotten ab alleles from him. However, she is not affected, and so she must be heterozygous for both traits.

Mom's chromosomes look like this:

A B (from mother)
a b (from father)

The genes are 10 cM apart, so crossing over between them will occur 10% of the time.

For example, Mom's gametes will exist in the following ratios

A B (Parental) 45%
a b (Parental) 45%
a B (crossover) 5%
A b (crossover) 5%

Otherwise, the two alleles on one chromosome will remain linked 90% of the time.

(Note that for the dad's chromosome, when crossing over between the alleles occurs, the allele combinations remain the same)

$$\begin{aligned} p(\mathbf{aabb\ child}) &= p(\mathbf{a\ b} \text{ chromosome from mom}) \text{ and } p(\mathbf{a\ b} \text{ chromosome from dad}) \\ &= \frac{1}{2} \times (0.9) \quad \times \quad \frac{1}{2} \\ &= (0.45) \times (0.5) = \mathbf{0.225} \end{aligned}$$

b. $cM = (100) \times$ (fraction of gametes that crossover between gene 1 and gene 2)
 fraction of gametes that crossover = $cM/100$
 fraction of gametes that will NOT crossover = $1 - (cM/100)$

$$\begin{aligned} p(\mathbf{aabb \ child}) &= p(\underline{a \ b} \ \text{chromosome from mom}) \text{ and } p(\underline{a \ b} \ \text{chromosome from dad}) \\ &= (1/2) \times (1 - [cM/100]) \times (1/2) \\ &= \mathbf{(1/4) \times (1 - [cM/100])} \end{aligned}$$

c. $X^a = \text{trait 1} \quad X^b = \text{trait 2} \quad X^{ab} = \text{traits 1 and 2} \quad X^+ = \text{wt}$

Dad's genotype = $X^{a+}Y$

Mom's genotype = $X^{ab}X^{++}$

*Mom's father was $X^{ab}Y$, so she received the X^{ab} chromosome from him, but must be heterozygous for both traits since she's unaffected

$$\begin{aligned} p(\mathbf{X^{ab}Y \ child}) &= p(X^{ab} \ \text{from mom}) \\ &= \mathbf{(1/2) \times (0.9) = 0.45} \end{aligned}$$

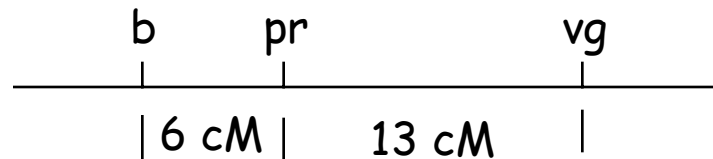
$$\begin{aligned} \text{d. } p(\mathbf{X^{ab}X^{ab} \ child}) &= p(X^{ab} \ \text{from mom}) \text{ and } p(X^{ab} \ \text{from dad}) \\ &= (.45) \quad \times \quad 0 \\ &= \mathbf{0} \end{aligned}$$

7.03 Problem Set 2

Due before 5 PM on Thursday, September 25

Hand in answers in recitation section or in the box outside the class

1. Consider a portion of an autosome in *Drosophila*, which carries the following three recessive markers: black body (b^-), purple eyes (pr^-), and vestigial wings (vg^-). The corresponding wild type alleles of each gene are designated b^+ , pr^+ , and vg^+ . A genetic map of this portion of the chromosome is shown below:



The measured two-factor distance between b and pr is 6 cM, and the distance between pr and vg is 13 cM.

- (a) Imagine that you want to set up a cross to verify these map distances and you have in the lab a true-breeding strain with black body and vestigial wings. What type of true breeding fly would you want to mate to in order to carry out a three-factor cross with b , pr , and vg ?
- (b) For the cross described in (a) what type of strain would be the best to mate to in order to score the gamete genotypes passed to the next generation?
- (c) What would be the rarest phenotypic class(es) produced from the cross in part (b)?
- (d) Write out the eight possible genotypes produced in the cross described in part (b). If 1,000 progeny are produced in this cross, how many of each type would you expect?
- (e) What would you expect the measured distance between b and vg to be in a two-factor cross between these markers? Explain why this distance is less than the sum of the b – pr and pr – vg distances (i.e. 19 cM)?

2. You have isolated two different yeast mutants that will not grow on medium that lacks arginine. You call these mutants $arg1^-$ and $arg2^-$.

a) Mating of either $arg1^-$ or $arg2^-$ to wild type produces diploids that can grow without arginine. Mating of $arg1^-$ to $arg2^-$ produces a diploid that also can grow on medium without arginine. What do these results tell you about the $arg1^-$ and $arg2^-$ mutations?

b) Sporulation of the diploids produced by the mating of $arg1^-$ to $arg2^-$ yields tetrads of the following types:

<u>type 1</u>	<u>type 2</u>
1 Arg ⁺ , 3 Arg ⁻	4 Arg ⁻

Out of 20 tetrads, 2 are type 1, and 18 are type 2. Categorize each of the tetrad types as parental ditype (P), tetratype (T), or nonparental ditype (N). If the markers are linked give the distance between them in cM. Does this result tell you anything new about the gene(s) mutated in $arg1^-$ and $arg2^-$?

c) Now you isolate a third mutant called $arg3^-$. When this mutant is mated to wild type the resulting diploids cannot grow without arginine. When $arg3^-$ is mated to $arg1^-$ the resulting diploid cannot grow without arginine. What do these results tell you about $arg3^-$ and its relationship to $arg1^-$?

d) When the diploid produced by mating $arg3^-$ to $arg1^-$ is sporulated tetrads of three types are produced.

<u>type 1</u>	<u>type 2</u>	<u>type 3</u>
1 Arg ⁺ , 3 Arg ⁻	4Arg ⁻	2 Arg ⁺ , 2 Arg ⁻

Out of 20 tetrads, 12 are type 1, 5 are type 2, and 3 are type 3. Categorize each of the tetrad types as parental ditype (P), tetratype (T), or nonparental ditype (N). What does this result tell you about the relationship between $arg3^-$ and $arg1^-$? Give any relevant distances in cM.

e) Based on these results deduce the relationship between $arg3^-$ and $arg2^-$. In a mating of $arg3^-$ to $arg2^-$ what would the phenotype of the resulting diploid be? If this diploid were sporulated and 60 tetrads were examined how many tetrads of each type would you expect?

3. You have isolated two temperature-sensitive mutations in phage ϕ . These phage mutants are called **ts-1** and **ts-2**. Each mutant will form plaques at 30°C but not at 42°C. You cross **ts-1** to **ts-2** phage by coinfecting *E. coli* at the permissive temperature of 30°C. When the resulting phage lysate is plated at 30°C you count 10^5 plaques per ml of phage lysate, but when the same phage lysate is plated at 42°C, there are 300 plaques per ml.

(a) What is the distance between the **ts-1** and **ts-2** mutations in m.u.?

You next cross a **ts-1** phage strain to a **ts-2** phage strain that also carries a **cl⁻** mutation, which gives plaques with clear centers. When the resulting lysate is plated out at 42°C and 100 plaques are examined, 85 are clear and 15 have normal turbid centers.

(b) If the phage produced from this cross were plated at 30°C, what fraction of the plaques would you expect to have clear centers?

(c) Draw a map showing the relative order of **cl⁻**, **ts-1** and **ts-2** as well as all of the distances you can calculate in m.u.

7.03 Problem Set 3

Due before 5 PM on Thursday, October 16

Hand in answers in recitation section or in the box outside the class

1. (a) The three stop codons are 5'UAG3', 5'UAA3', and 5'UGA3'. Explain why mutagens that specifically induce G•C to A•T mutations will generally produce TAG (amber) and TGA nonsense mutations more frequently than TAA (ochre) mutations.

(b) You are trying to isolate tRNA mutations that will suppress an amber (TAG) mutation. To increase the frequency of such mutations you use a mutagen that produces transition mutations (i.e. C•G to T•A and T•A to C•G base changes). Which tRNA genes could in principle be altered by the mutagen to give the desired suppressor mutation? For your answer give the sequence of the codon normally recognized by the tRNA. In addition, write out the DNA base sequence of the segment of the tRNA gene that codes for the anticodon sequence for both the wild-type and suppressor alleles of the gene. Please show both strands of the DNA and indicate the 5' and 3' ends of each strand indicating which strand is used as the template for transcription of the tRNA molecule.

2. The diagram below shows the F factor and a portion of the *E. coli* chromosome that has three different insertion sequences (IS) of the same type as is carried on F.

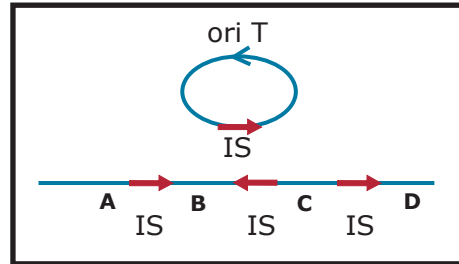


Figure by MIT OCW.

(a) Draw a diagram of the three different Hfrs that can be formed by recombination between the IS on F and each of the IS sequences on the chromosome. For your answer, include the positions of each of the markers (**A**, **B**, **C**, and **D**) and state which of these markers would be transferred early. (You will need to take into account the orientation of the IS and origin of transfer (ori T) on F as well as the orientations of the different IS sequences in the chromosome.)

(b) For each of the three Hfrs in part (a) consider the outcome of further crossovers between IS sequences flanking the integrated F factor (there should be five different relevant IS by IS recombinations for each of the three Hfrs). For each possible recombination event state whether the product is F, an F' or an Hfr. Also provide information about which markers will be transferred on the F's and which markers will be transferred early by the Hfrs.

3. Wild type *E. coli* metabolizes the sugar lactose by expressing the enzyme β -galactosidase. You have isolated a mutant that you call $lac1^-$, which cannot synthesize β -galactosidase and cannot grow on lactose (Lac^-).

(a) You have a wild type (Lac^+) strain carrying a Tn5 insertion known to be near several Lac genes on the *E. coli* chromosome. You grow P1 phage on this strain and use the resulting phage lysate to infect the $lac1^-$ strain, selecting for kanamycin resistance (Kan^r). Among 100 Kan^r transductants, you find that 82 are Lac^- and 18 are Lac^+ . Express the distance between Tn5 and the $lac1^-$ mutation as a cotransduction frequency.

(b) You isolate a second Lac^- mutation, which you designate $lac2^-$. Using the same P1 lysate as in part (a) you infect the $lac2^-$ strain, selecting for Kan^r transductants. In this case, all 100 Kan^r transductants that are examined are Lac^- . What does this result tell you about the relationship between the $lac1^-$ and $lac2^-$ mutations?

(c) Next, you isolate a mutation that constitutively expresses abnormally high levels of β -galactosidase, which you designate lac^C . Preliminary P1 transduction experiments indicate that lac^C is linked to the Tn5 insertion described in part (a). To map lac^C relative to $lac1^-$ you set up two reciprocal crosses. In the first cross you grow P1 on a strain that carries the Tn5 insertion and the $lac1^-$ mutation. You then use this lysate to infect a lac^C mutant and select for Kan^r . From 100 Kan^r transductants examined, 20 are Lac^- , 76 express β -galactosidase constitutively and 4 show normal β -galactosidase expression. In the second cross you grow P1 on a strain that carries the Tn5 insertion and the lac^C mutation. You then use this lysate to infect a $lac1^-$ mutant, and select for Kan^r . From 100 Kan^r transductants examined, 81 are Lac^- and 19 express β -galactosidase constitutively. Draw a genetic map showing the relative positions of the Tn5 insertion and the $lac1^-$ and lac^C mutations. Express any measured distances as cotransduction frequencies.

(d) A third Lac^- strain, called $lac3^-$, is linked to the Tn5 insertion. From a strain carrying the Tn5 insertion and $lac3^-$ mutation you isolate an F' that carries a region of the chromosome that includes both Tn5 and the linked Lac region (You know from mapping experiments that this region includes the sites of both the $lac1^-$ and $lac3^-$ mutations). You introduce this F' into an F^- strain carrying $lac1^-$ by selecting for Kan^r . These merodiploids express β -galactosidase normally. What does this result tell you about the relationship between the $lac3^-$ and $lac1^-$ mutations? If the merodiploid were Lac^- , what could you conclude about the relationship between the $lac3^-$ and $lac1^-$ mutations? (Be sure to consider all the possibilities.)

4. In your study of a new bacterial species you have identified a transducing phage that you call Px. In addition you have worked out methods to make random transposon insertions into the bacterial genome. You have generated two different transposon insertion collections one with 10^5 random Tn5 (Kan^r) insertions and the other with 10^5 random Tn10 (Tet^r) insertions. You grow Px phage on the mixed collection of Tn5 insertions and use the resulting phage lysate to infect the mixed collection of Tn10 insertions. You select 10,000 Kan^r transductants and find that 80 of them are Tet^S . Use this information to estimate the total size of the bacterial genome assuming that both Tn5 and Tn10 insert randomly and that the average size of a fragment recombined into the recipient genome during Px transduction is 55 kbp. (Tn5 is about 5 kbp and Tn10 is 10 kbp.)

7.03 Fall 2003
Problem Set #3 Solutions
Issued Friday, October 17, 2003

1. (a) We are analyzing mutagens that specifically induce G·C→A·T mutations in DNA. Therefore, we must determine the potential double stranded DNA sequences that will encode stop codons after going through this specific mutation.

We will start with 5'UAG3'. The double stranded DNA that corresponds to 5'UAG3' is:
3'ATC5' template strand
5'TAG3' coding strand

We need to figure out what specific double stranded DNA sequences could have undergone a G·C→A·T mutation to become the sequence above. To do this, just work backwards and change the AT base pairs in the above sequence into GC base pairs, one pair at a time. If you do this, you will get the following double stranded DNA sequences (the mutated bases are boldfaced):

3'**G**T**C**5' template strand
5'**C**A**G**3' coding strand (codes for Gln)

3'**A**C**C**5' template strand
5'**T**G**G**3' coding strand (codes for Trp)

These two sequences each underwent one G·C→A·T mutation to become DNA that encodes the 5'UAG3' stop codon. We have chosen to make only one mutation per three base pairs because the likelihood of a mutagen acting on two consecutive base pairs (or two base pairs out of three) is extremely small. We can neglect those extremely rare events since this problem is asking why something "generally" happens.

Now, repeat the process for the other two stop codons. The double stranded DNA that encodes 5'UGA3' is:

3'ACT5' template strand
5'TGA3' coding strand

And the DNA possible sequences that could have been mutated to become the above sequence are:

3'**G**C**T**5' template strand
5'**C**G**A**3' coding strand (codes for Arg)

3'**A**C**C**5' template strand
5'**T**G**G**3' coding strand (codes for Trp)

Lastly, repeat for 5'UAA3'. Its corresponding DNA is:

3'ATT5' template strand
5'TAA3' coding strand

And the DNA possible sequences that could have been mutated to become the above sequence are:

3'**GTT**5' template strand

5'**CAA**3' coding strand (codes for Gln)

3'**ACT**5' template strand

5'**TGA**3' coding strand (codes for UGA stop codon)

3'**ATC**5' template strand

5'**TAG**3' coding strand (codes for UAG stop codon)

Notice that for UGA and UAG, **the two candidate sequences for mutation both encode amino acids**, whereas for UAA, **only one of its candidate sequences encodes an amino acid**. The other two correspond to stop codons. When G·C→A·T mutagens are introduced, they mutate random base pairs along the DNA sequence. **Given that there are more targets for the creation of TGA and TAG nonsense mutations, these two mutations will occur more frequently than TAA mutations.**

(b) The amber TAG mutation has the following mRNA codon, tRNA anticodon, and corresponding DNA coding for the anticodon portion of the tRNA:

mRNA codon 5'UAG3'

tRNA anticodon 3'AUC5'

DNA encoding anticodon portion 5'**TAG**3' ← template strand for tRNA transcription
3'**ATC**5'

This DNA sequence is also the sequence that codes for the anticodon portion of amber suppressing tRNA alleles. In other words, we want the mutagen to change normal DNA sequences into this sequence, so that we can have amber suppressing alleles. To find which tRNA genes can be altered to become the DNA sequence above, we can work backwards like in part (a). Now, since we have a G·C↔A·T mutagen, we can change AT pairs into GC pairs, and vice versa. Again, we will change one base pair at a time. If we do this, we will get the following three DNA sequences (the mutated bases are in boldface):

DNA encoding anticodon portion 5'**CAG**3' ← tRNA template (for tRNA^{Gln})
3'**GTC**5'

DNA encoding anticodon portion 5'**TGG**3' ← tRNA template (for tRNA^{Trp})
3'**ACC**5'

DNA encoding anticodon portion 5'**TAA**3' ← tRNA template (for UAA stop tRNA)
3'**ATT**5'

If you follow the flow of genetic information, you will see that these three sequences code for the anticodon portion of tRNA^{Gln}, tRNA^{Trp}, and the tRNA that recognizes the UAA stop codon. **So the gln and trp tRNA genes can be mutated to become amber**

suppressors. The codons normally recognized by tRNA^{Gln} and tRNA^{Trp} are 5'CAG3' and 5'UGG3', respectively.

We gave full credit to those who ended the solution here. But since the problem asked which genes could "in principle" be altered to become amber suppressors, we can consider cases where two or three base pairs from each codon were altered by the mutagen. In that case, we would arrive at the following DNA sequences (with the corresponding mRNA codons also listed):

DNA encoding anticodon portion 5'CGA3' ← tRNA template (for tRNA^{Arg})
 3'GCT5'
 mRNA codon 5'CGA3'

DNA encoding anticodon portion 5'CGG3' ← tRNA template (for tRNA^{Arg})
 3'GCC5'
 mRNA codon 5'CGG3'

DNA encoding anticodon portion 5'CAA3' ← tRNA template (for tRNA^{Gln})
 3'GTT5'
 mRNA codon 5'CAA3'

DNA encoding anticodon portion 5'TGA3' ← tRNA template (for UGA stop tRNA)
 3'ACT5'
 mRNA codon 5'UGA3'

So assuming the mutagen can change two or three base pairs in the same triplet, we arrive at additional tRNA genes that can be mutated to become amber suppressors. These are **two tRNA^{Arg} genes and another tRNA^{Gln}**.

2. (a) ⇨ = IS, ► = OriT

Hfr #1: A ⇨►⇨B ⇨ C ⇨ D A transferred early
 Hfr #2: A ⇨B ⇨◄◄ C ⇨ D C (and then D) transferred early
 Hfr #3: A ⇨B ⇨ C ⇨►⇨D C (then B then A) transferred early

(b) label the order of IS from A ⇨ ... ⇨D
 1 4

Hfr #	Recombination Between IS #s	Recombinatory Result	Product	Early
1	1 & 2	⇨►	F	---
1	1 & 3	A ⇨B ⇨◄◄ C ⇨ D	Hfr	C, D
1	1 & 4	⇨►⇨B ⇨ C	F'	C, B

1	2 & 3 (or 3&4*)	A \Rightarrow \blacktriangleright \Rightarrow B \Leftarrow C \Rightarrow D	Hfr	A
1	2 & 4	A \Rightarrow \blacktriangleright \Rightarrow D	Hfr	A
2	1 & 2 or 3 & 4	A \Rightarrow B \Leftarrow \blacktriangleleft \Leftarrow C \Rightarrow D	Hfr	C, D
2	1 & 3	A \Rightarrow \blacktriangleright \Rightarrow B \Leftarrow C \Rightarrow D	Hfr	A
2	(1&4*)	\Rightarrow B \Leftarrow \blacktriangleleft \Leftarrow C	F'	C, B
2	2 & 3	\blacktriangleleft \Leftarrow	F	---
2	2 & 4	A \Rightarrow B \Leftarrow C \Rightarrow \blacktriangleright \Rightarrow D	Hfr	C, B, A
3	2 & 3 (or 1&2*)	A \Rightarrow B \Leftarrow C \Rightarrow \blacktriangleright \Rightarrow D	Hfr	C, B, A
3	1 & 3	A \Rightarrow \blacktriangleright \Rightarrow D	Hfr	A
3	1 & 4	B \Leftarrow C \Rightarrow \blacktriangleright \Rightarrow	F'	C, B
3	2 & 4	A \Rightarrow B \Leftarrow \blacktriangleleft \Leftarrow C \Rightarrow D	Hfr	C, D
3	3 & 4	\blacktriangleright \Rightarrow	F	---

* denotes recombinatory events between two IS sequences NOT flanking the OriT
(The p-set asked only for events between sequences that flank the OriT. Therefore these are extraneous answers)

3. (a) The donor strain is Lac⁺ and the recipient strain is Lac⁻. Therefore, in the Lac⁺ Kan^r transductants, *lacI*⁺ was cotransduced with Tn5. So the distance between Tn5 and the *lacI*⁻ mutation is:

$$(18/100) \cdot 100\% = 18\%$$

(b) Since none of the 100 Kan^r transductants were Lac⁺, we can conclude that Tn5 was never co-transduced with *lac2*⁺. This indicates that the distance between *lac2*⁻ and Tn5 is at least one phage head (10⁵ bp). We know from part (a) that Tn5 and *lacI*⁻ are within one phage head since their cotransduction frequency was 18%. But since we do not know the relative order of the three markers, we cannot say whether the two *lac* mutations are within one phage head. If Tn5 were between *lacI*⁻ and *lac2*⁻, then the distance between the two mutations would be more than one phage head. However, if Tn5 were not the middle marker, we cannot say whether *lacI*⁻ and *lac2*⁻ are one phage head apart. **So the only conclusion we can draw from this data is that *lacI*⁻ and *lac2*⁻ are very far from each other. Therefore, they are unlinked, and are mutations in two different genes.**

(c) The best way to solve this type of problem is to draw out the two crosses. We will consider only two possible orders, instead of three, because the order where Tn5 is in the middle is impractical for three-factor cotransduction experiments. It is likely that Tn5 insertion between the two markers of interest (which are close together in this case) will occur in a coding region, causing an additional mutation that would skew the experimental data.

There are two reciprocal crosses and two possible orders. So we will have to draw four diagrams:

(1- and 1+ will denote the mutant and WT loci of *lac1*, respectively. Similarly, C- and C+ will denote the mutant and WT loci of *lac^c*)

	Order #1 (Tn5, <i>lac1</i> , <i>lac^c</i>)	Order #2 (Tn5, <i>lac^c</i> , <i>lac1</i>)																																								
Cross #1 (Tn5, 1- crossed to C-)	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">1-</td> <td style="text-align: center;">C+</td> <td style="text-align: center;">C+</td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">1+</td> <td style="text-align: center;">C-</td> <td></td> </tr> </table>	Tn5	1-	C+	C+	~~~~~				X	X	X	X	-----					1+	C-		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">C+</td> <td style="text-align: center;">1-</td> <td style="text-align: center;">1-</td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">C-</td> <td style="text-align: center;">1+</td> <td></td> </tr> </table>	Tn5	C+	1-	1-	~~~~~				X		X		-----					C-	1+	
Tn5	1-	C+	C+																																							
~~~~~																																										
X	X	X	X																																							
-----																																										
	1+	C-																																								
Tn5	C+	1-	1-																																							
~~~~~																																										
X		X																																								

	C-	1+																																								
Cross #2 (Tn5, C- crossed to 1-)	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">1+</td> <td style="text-align: center;">C-</td> <td style="text-align: center;">C-</td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td></td> <td style="text-align: center;">X</td> <td></td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">1-</td> <td style="text-align: center;">C+</td> <td></td> </tr> </table>	Tn5	1+	C-	C-	~~~~~				X		X		-----					1-	C+		<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="text-align: center;">Tn5</td> <td style="text-align: center;">C-</td> <td style="text-align: center;">1+</td> <td style="text-align: center;">1+</td> </tr> <tr> <td colspan="4" style="text-align: center;">~~~~~</td> </tr> <tr> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> <td style="text-align: center;">X</td> </tr> <tr> <td colspan="4" style="text-align: center;">-----</td> </tr> <tr> <td></td> <td style="text-align: center;">C+</td> <td style="text-align: center;">1-</td> <td></td> </tr> </table>	Tn5	C-	1+	1+	~~~~~				X	X	X	X	-----					C+	1-	
Tn5	1+	C-	C-																																							
~~~~~																																										
X		X																																								
-----																																										
	1-	C+																																								
Tn5	C-	1+	1+																																							
~~~~~																																										
X	X	X	X																																							

	C+	1-																																								

The data given will allow us to determine which of the two possible orders is correct. As in any three factor cross, we determine order by looking for the rarest class. In this case, the rarest class shows normal B-gal expression. The genotype of this class is Tn5, 1+, C+.

From the data, we see that cross #1 produced four such transductants while cross #2 produced none. This is the key observation that allows us to determine order.

We look back to our diagrams. The X's indicate the crossovers that needed to occur to give the Tn5, 1+, C+ genotype (which gives normal B-gal expression). Assuming order #1 is correct, we would get more normal B-gal transductants from cross #2, since only two crossovers are required as opposed to four. Assuming order #2 is correct, we would get more normal B-gal transductants from cross #1. This is what the data shows. There are four normal B-gal transductants from cross #1 and none from cross #2. **Therefore, order #2 is the correct order:**

_____ Tn5 _____ *lac^c* _____ *lac1-* _____
 [-----18%-----]
 [-----24%-----]

The Tn5 to *lac1-* distance is 18% as calculated in part (a).

The Tn5 to *lac^c* distance is found in the data for the first reciprocal cross. The 76/100 Kan^r b-gal constitutive transductants from this cross received Tn5 but did not receive C+. Therefore, P(no cotransduction between Tn5 and *lac^c*) = 76%

We can see that: **$P(\text{cotransd. btw Tn5 and } lac^c) = 1 - P(\text{no cotransd. btw Tn5 and } lac^c) = 100\% - 76\% = 24\%$**

The distance between *lac^c* and *lacI* is very small. This is indicated by the fact that there are very few transductants that exhibit normal b-gal expression (crossing over between *lac^c* and *lacI* needs to occur to produce transductants with normal b-gal expression in both crosses). The exact distance between *lac^c* and *lacI* cannot be calculated since we only selected for Kan^r transductants. Therefore, all data is relative to the Tn5 marker.

(d) The F' plasmid carries Tn5 and *lac3-* while the chromosome has a *lacI-*. When selecting for Kan^r, you are selecting for cell that have successfully taken in the F plasmid, which confers resistance to kanamycin.

If there merodiploids express b-gal normally, then you can conclude that the ***lacI-* and *lac3-* mutations lie in different genes (they complement each other).**

If the merodiploid were Lac⁻, then you can draw one of two conclusions, which are indistinguishable without further experimentation: **(1) the two mutations lie in the same gene, or (2) one (or both) of the mutations is dominant to wild type.**

4. A Kan^R, Tet^R transductant results when recombination of the Tn5 occurs in the genome in a place not overlapping with the insertion site of Tn10. This happens 9920/10000 times, or 99.2% of the time. Similarly, A Kan^R, Tet^S transductant results when recombination of Tn5 occurs in the genome in a place overlapping with the insertion site of Tn10, thus disrupting the Tet^R gene. This happens 80/10000 times, or 0.8% of the time.

We are told that the average size of a recombined fragment is 55kbp, this means that a phage head can fit 55 kbp of DNA, which is larger than either Tn10 or Tn5. We can find the size of the genome by seeing that 55kbp is proportional to the size of the genome just as 0.8% is proportional to 100% (0.8% recombination tells us the phage is holding 0.8% of the genome). However, before we can use 55kbp to find the genome size, we must subtract the 5kbp of Tn5. We subtract off 5kbp because there is no Tn5 in the genome, so there is no possibility of recombination within this region. Therefore, it is not counted in our measurement of recombination.

Therefore: $80/10,000 = 0.008 =$ the probability that the 50 kbp phage DNA with Tn5 recombined into a spot overlapping Tn10.

$0.008/50 = 1/(\text{total genome length}) \rightarrow$ from this we get a **total genome length of 6250kbp.**

note: A common mistake would be to subtract the size of the Tn10 off of the final answer. This is not correct, because it was never included in the 6250kbp in the first place, as there is no Tn10 in the 55kbp of DNA and there was never a chance of recombination.

7.03 Problem Set 4

Due before 5 PM on Thursday, October 23

Hand in answers in recitation section or in the box outside the class

1. (a) You have isolated two mutations in the Lac operon that cause constitutive expression of Lac genes. You designate these mutants Lac1⁻ and Lac2⁻. Making use of an F' that carries the Lac operon with the LacY gene mutated, you construct strains that you test for both β-galactosidase activity and Lac permease activity with results shown below.

	β-galactosidase activity		Lac permease activity	
	- IPTG	+ IPTG	- IPTG	+ IPTG
Lac 1 ⁻ Lac Z ⁻ / F' LacY ⁻	-	+	+	+
Lac 2 ⁻ Lac Z ⁻ / F' LacY ⁻	+	+	+	+

Classify each mutation as dominant or recessive and as cis- or trans-acting, giving the experimental result that allows you to arrive at each conclusion. Finally, deduce what type of Lac mutation best fits the properties of Lac 1⁻ and of Lac 2⁻.

(b) The first Lac promoter mutations that were isolated were “leaky” mutations that decreased, but did not entirely eliminate, the promoter function. The table below shows the behavior of such mutants giving the quantity of β-galactosidase activity and Lac permease activity produced.

	β-galactosidase activity		Lac permease activity	
	- IPTG	+ IPTG	- IPTG	+ IPTG
Lac ⁺ (wild type)	2	100	2	100
LacP ⁻	1	8	1	10

The LacP⁻ mutations map very close to LacO^c nevertheless researchers concluded that LacP and LacO represent functionally distinct sites on the DNA. Explain how the results shown above would lead to such a conclusion.

(c) You have isolated two mutations that show decreased expression of the Lac operon. However, unlike like the promoter mutations described in part (b) these mutations don't respond to the inducer IPTG. These mutations, designated Lac3⁻ and Lac4⁻, are evaluated for the quantity of β-galactosidase and permease activity expressed with or without IPTG:

	β-galactosidase activity		Lac permease activity	
	<u>- IPTG</u>	<u>+ IPTG</u>	<u>- IPTG</u>	<u>+ IPTG</u>
Lac 3 ⁻	12	12	12	12
Lac 3 ⁻ / F' Lac I ⁺ LacZ ⁻	1	12	na	na
Lac 4 ⁻	8	8	9	9
Lac 4 ⁻ / F' Lac I ⁺ LacZ ⁻	8	8	na	na

(na = not assayed)

Mapping experiments reveal that Lac3⁻ and Lac4⁻ are different short deletions located in the region before the start of the LacZ gene. Given the data shown above suggest which genetic element(s) in addition to part of the promoter has been deleted in each mutant. Explain your reasoning.

2. You are studying a new strain of *E. coli* that can utilize the disaccharide melibiose very efficiently. You find that utilization depends on the enzyme melibiase, which is encoded by the gene Mel1. Mel1 is not expressed unless melibiose is present in the growth medium.

(a) You have isolated a mutation that causes constitutive melibiase activity, which you designate MelA⁻. P1 phage mapping experiments using a Tn5 insertion linked to Mel1 show that MelA⁻ is not linked to Mel1. Moreover you find that when an amber suppressor is introduced into a MelA⁻ mutant, normal melibiase regulation is restored. Classify the MelA⁻ mutation in terms of its basic genetic properties explaining the rationale behind your conclusions. Based on these properties make a proposal for the type of regulatory functions affected by the MelA⁻ mutation.

(b) Next you isolate a mutation, designated MelB⁻, which gives uninducible melibiase activity. Mapping experiments show that MelB⁻ is linked to Mel1. Using an F' factor that carries the chromosomal region surrounding Mel1, you perform the following genetic tests:

	melibiase activity	
	<u>- melibiose</u>	<u>+ melibiose</u>
wild type (Mel1 ⁺)	-	+
Mel1 ⁻	-	-
MelB ⁻	-	-
MelB ⁻ / F' Mel ⁺	-	+
Mel1 ⁻ / F' Mel ⁺	-	+
MelB ⁻ / F' Mel1 ⁻	-	+

Describe the basic genetic properties of the $MelB^-$ mutation, explaining the rationale for your conclusions, and make a proposal for the type of regulatory functions affected by the $MelB^-$ mutation.

(c) Diagram two possible models for regulatory pathways for $Mel1$ that can explain the behavior of the $MelA^-$ and $MelB^-$ mutations. For each model include a role for the inducer melibiose.

(d) You next construct a $MelA^- MelB^-$ double mutant, which gives the following behavior:

		melibiase activity	
		<u>- melibiose</u>	<u>+ melibiose</u>
$MelA^-$	$MelB^-$	-	-

Which of your two models is consistent with this new data?

(e) Next, you isolate a third mutant, $MelC^-$, which gives constitutive melibiase expression. The $MelC^-$ mutation is closely linked to $Mel1$ and $MelB^-$. Genetic tests of the $MelC^-$ mutation yield the following:

		melibiase activity	
		<u>- melibiose</u>	<u>+ melibiose</u>
$MelC^-$		+	+
$MelC^- / F' Mel^+$		+	+
$MelC^- Mel1^- / F' Mel^+$		+	+
$MelC^- / F' Mel1^-$		+	+

As above, classify the $MelC^-$ mutation in terms of its basic genetic properties and explain how you arrived at your conclusions.

(f) A $MelB^- MelC^-$ double mutant shows uninducible melibiase activity. Assuming that $MelC^-$ mutations affect the same gene as $MelB^-$ mutations, propose two different possible mechanisms to explain the behavior of $MelC^-$. Your answer should include a diagram showing the entire pathway for $Mel1$ regulation indicating the function of each of the elements affected by the $MelA^-$, $MelB^-$, and $MelC^-$ mutations and the inducer melibiose. Finally, propose some type of experiment(s) that would allow you to distinguish the two possible mechanisms.

7.03 Problem Set 6

Due before 5 PM on Thursday, November 20, 2003

Hand in answers in recitation section or in the box outside the class

1. In humans, albinism (unpigmented skin, hair, and eyes) is due to an enzymatic deficiency, and it is an autosomal recessive trait. Suppose that in a small country of one million people (“Generation 1”), there are 500 *aa* albinos and 9000 *Aa* heterozygous carriers.

(a) Estimate q , the frequency of allele *a*, in Generation 1. Show your calculations.

(b) Estimate p , the frequency of allele *A*, in Generation 1. Show your calculations.

(c) In the next generation of 1 million individuals (Generation 2), what are the expected numbers of *aa* albinos and *Aa* carriers? (Assume random mating and all other Hardy-Weinberg conditions.) Show your calculations.

(d) Has the frequency of allele *a* changed between Generations 1 and 2? Briefly justify your answer.

(e) Was Generation 1 in Hardy-Weinberg equilibrium? Briefly justify your answer.

What is the probability that a child will be albino if:

(f) Both parents are non-albino members of Generation 2?

(g) One parent is a non-albino member of Generation 2, and the other parent is a non-albino member of Generation 1.

(h) Both parents are members of Generation 2, and one parent is albino and the other is non-albino.

2. Consider a heritable autosomal disease with an incidence in the population of 1 per thousand. On average, individuals with the disease have 80% as many children as the population average.

In answering the various parts of this question, assume that mating is random. State any additional simplifying assumptions that you employ, and show your calculations.

- (a) What mutation rate would be required to maintain the observed incidence of the disease in the population if the disease is dominant?
- (b) What mutation rate would be required to maintain the observed incidence of the disease in the population if the disease is recessive?
- (c) Now assume that the mutation rate is zero, that the disease is recessive, and that the disease allele is maintained in the population by heterozygote advantage. Calculate the heterozygote advantage, and restate its meaning in biological terms.
- 3.** In this question we will consider the interaction of selection and inbreeding in determining the incidence of autosomal recessive diseases. Consider a gene in which recessive mutations occur at a rate of 10^{-5} . Assume a selective disadvantage S of 0.4 in homozygotes for the recessive allele.

In answering the various parts of this question, show your calculations (unless none are required), and state any additional simplifying assumptions that you employ.

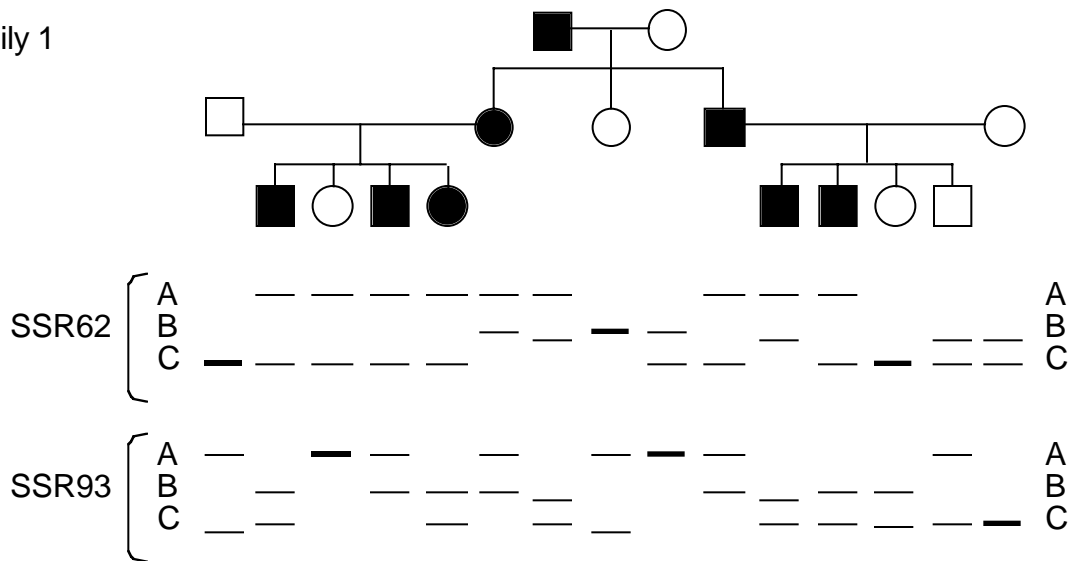
- (a) Calculate q , the frequency of the recessive allele. Also calculate the incidence of the disease. Assume random mating.
- (b) Now assume that, for thousands of generations, 10% of all children have been products of first-cousin matings (the remaining 90% being products of random matings). Calculate the steady-state value of q . Also calculate the incidence of the disease at steady state. (Hint: first modify the equation $\Delta q_{\text{sel}} = -Sq^2$ [from lecture 26] to reflect inbreeding's effects on the incidence of homozygotes for the recessive allele.)
- (c) Now assume that the population described in part (b) suddenly and completely ceases all inbreeding. Calculate the incidence of disease in the first generation conceived with no inbreeding.
- (d) Would q be expected to rise, fall, or remain unchanged during the first 10 generations after the cessation of inbreeding described in part (c)? Briefly justify your answer. (No calculations needed.) What numerical value would q approach after thousands of generations with no inbreeding?

7.03 Problem Set 7

Due before **4 PM** on **Friday**, December 5, 2003. Hand in answers in recitation section or in the box outside the class.

1. You are conducting genetic linkage studies of an autosomal dominant disease. You are focused on two SSR markers that may be linked to each other and to the disease. Here is a family in which some individuals are affected:

Family 1



Calculate LOD scores for linkage at $\theta = 0.1$ between:

(a) The disease and SSR62

(b) The disease and SSR93

(c) SSR62 and SSR93

(d) If SSR62, SSR93, and the disease gene are all located in the same region of the same chromosome, what is their most likely order? Briefly justify your answer.

2. You are running a human assisted reproduction clinic and providing state-of-the-art genetic diagnostic services. A married couple who already had a child with cystic fibrosis approach you because they wish to have another child, but only if they can be assured that the child will not have cystic fibrosis. You genotype the woman and discover that she is a heterozygote for Del508, the most common mutation causing cystic fibrosis. You suggest that the couple consider first polar body testing, in which several unfertilized oocytes (each with its first polar body) are retrieved from the woman, the first polar bodies are removed, and PCR tests are conducted on DNA from each of the first polar bodies. The couple agrees, and you obtain the following results:

Oocyte	PCR test of first polar body DNA for presence of	
	Del508	wild-type sequence
#1	+	+
#2	+	-
#3	+	-
#4	-	+
#5	+	+
#6	-	+

(a) Propose an explanation for the observation that the polar bodies for oocytes #1 and #5 test positive for both Del508 and the wild type sequence, while the polar bodies for the other oocytes do not. (Do not appeal to nondisjunction.) Illustrate your explanation with a sketch of meiosis in oocyte #1.

(b) Given the couple's desire to have a child without cystic fibrosis, which oocytes would you employ in vitro fertilization? Briefly explain or illustrate your answer.

(c) Which additional oocytes would you consider fertilizing if you could then 1) retrieve the second polar body for PCR testing and 2) selectively return "unaffected" embryos to the woman's uterus? Briefly explain or illustrate your answer.

(d) The cystic fibrosis gene is located on human chromosome 7. Would polymorphisms within the cystic fibrosis gene be suitable for use in determining the parental origin of the extra chromosome in cases of trisomy 7? Would they be suitable for use in determining whether nondisjunction had occurred in Meiosis I or Meiosis II? Briefly justify your answers.

7.03 Problem Set 5

Due before 5 PM on Thursday, November 13

Hand in answers in recitation section or in the box outside the class

1. You are studying regulation of the yeast enzyme glutamine synthetase (GS), which is encoded by the *GLN1* gene. You have isolated two mutants, designated *gln2*⁻ and *gln3*⁻, that give decreased GS activity. Mating of either *gln2*⁻ or *gln3*⁻ haploids to wild type produces heterozygous diploids that show normal amounts of GS expression. When you cross either a *gln2*⁻ or *gln3*⁻ haploid to a *gln1*⁻ strain the resulting diploids show normal expression of GS.

(a) From these experiments classify the *gln2*⁻ and *gln3*⁻ mutations in terms of their basic genetic properties explaining the rationale behind your conclusions. Based on these properties make a proposal for the types of regulatory functions affected by the *gln2*⁻ and *gln3*⁻ mutations.

(b) Diagram two different linear models and one parallel model that could explain the effect of the *gln2*⁻ and *gln3*⁻ mutations on the regulation of *GLN1*.

The *GLN1* gene shows a rather complex regulation in response to different amino acids. When either glutamate (*glu*) or glutamine (*gln*) is added to the medium, the amount of GS expression diminished and when both glutamate *and* glutamine are added to the medium GS expression is shut off completely. The effects of different mutants on the response to *glu* and *gln* are shown below.

	Units of GS activity			
	<u>-</u>	<u>+ glu</u>	<u>+ gln</u>	<u>+glu& gln</u>
wild type	100	50	50	0
<i>gln1</i> ⁻	0	0	0	0
<i>gln2</i> ⁻	50	50	0	0
<i>gln3</i> ⁻	50	0	50	0

(c) Which of the models from part (b) best fits these experimental results? Diagram a complete model for the regulation of *GLN1* that includes the effects of *glu* and *gln*.

(d) Based on your model for part (c) how would you expect a *gln2*⁻ *gln3*⁻ double mutant to behave?

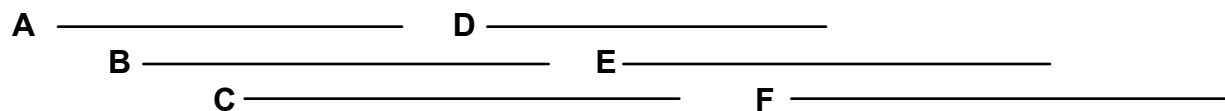
Next, you decide to evaluate the promoter for the GLN1 gene. To do this you first fuse the promoter region to the LacZ coding sequence and then place this hybrid gene on an appropriate yeast plasmid. You find that cells carrying the hybrid gene express activity under the same conditions that GS is expressed in wild type cells, meaning that the promoter region you have selected contains all of the necessary cis-acting sequences for normal regulation. The figure below shows the effect of different 50 bp deletions in the promoter region on the amount of β -galactosidase activity expressed by the reporter gene.

	-300	-250	-200	-150	-100	-50	+1	Units of β -galactosidase			
								-	+ glu	+ gln	
wt	—————							LacZ	100	50	50
1	—	—————						LacZ	50	50	0
2	—————		—————					LacZ	100	50	50
3	—————			—————				LacZ	50	0	50
4	—————				—————			LacZ	50	0	50
5	—————					——		LacZ	100	50	50
6	—————						-	LacZ	0	0	0

(e) Describe the cis-acting elements in the GLN1 promoter that are evident from these experiments, giving both their position and as much of their function as you can deduce.

(f) How many units of β -galactosidase would you expect to be expressed from deletion 1 in a $gln2^-$ mutant? How many units of β -galactosidase would you expect to be expressed from deletion 4 in a $gln2^-$ mutant?

2. You obtain 6 BACs (of **known** order, as shown below) and 7 STSs (of **unknown** order) that derive from a region of mouse chromosome 16 whose genomic sequence has not yet been finished.



By PCR (using 20-bp primers at either end of each STS), you test each of the 6 BACs for the presence (+) or absence (-) of each of the 7 STSs. You obtain the following results:

		STSs						
		1	2	3	4	5	6	7
BACs	A	+	-	+	-	-	-	+
	B	+	+	-	-	+	-	+
	C	+	+	-	+	+	-	-
	D	-	+	-	+	+	-	-
	E	+	-	-	+	+	-	-
	F	+	-	-	-	+	+	-

(a) Construct a physical map of this region that shows the order of the STSs, and the placement of the STSs within the BACs. (Hint: Consider the possibility that one or two of the STSs are present twice in the region spanned by the 6 BACs.) Are there any ambiguities within your map? If so, describe them.

(b) You determine the DNA sequences of PCR products obtained at STS5 using BACs B, C, D, E, and F as PCR templates. The DNA sequences of PCR products obtained using BACs B and C are identical, but they differ by several nucleotide substitutions from the DNA sequences of PCR products obtained using BACs E and F. Oddly, PCR product obtained using BAC D appears to be a mixture of two sequences – one corresponding to the BAC B/ BAC C sequence and the other corresponding to the BAC E/ BAC F sequence. Briefly account for these findings, and explain how a single PCR assay could amplify two different sequences.

(c) Do the findings presented in part (b) confirm, refine, or refute your answers to part (a)? Briefly explain your answer.

(d) Would you expect the DNA sequence of PCR product obtained at STS5 using mouse genomic DNA as template to more closely resemble that obtained using BAC B, BAC D, or BAC F? Briefly explain your answer.

- (e) Is there a second STS at which you would like to sequence PCR products obtained using BACs as templates? If so, which BACs would you test in this way, and what sequencing results might you predict for each of the BACs tested?
- (f) How would you use the sequence information presented in part (b) to design two new STSs (with new PCR primer pairs) to replace STS5? (Call the new ones STS51 and STS52. STS51 should be present (+) in BAC B, and STS52 should be present (+) in BAC F.)
- (g) Would you expect STS51 and STS52 to be present in BAC D? In mouse genomic DNA?

3. Many mouse genes are “tissue-specific,” that is, they are present throughout the body but are expressed in only one of the animal’s many tissue types. (Other mouse genes are expressed throughout the body, or in multiple tissues.) Geneticists can study the regulation of a mouse gene by fusing the gene’s promoter region to the LacZ coding sequence and injecting the construct to create a transgenic mouse. Fusion of the mouse amylase promoter to LacZ yielded a P_{amylase} -LacZ construct.

- (a) Would microinjection of the P_{amylase} -LacZ construct into the male pronucleus of a fertilized egg likely result in integration of the construct into the amylase gene? Briefly explain your answer.
- (b) Mice heterozygous for the resulting P_{amylase} -LacZ transgene displayed LacZ expression exclusively in the pancreas. Would you expect homozygotes for the transgene to also display LacZ expression in the pancreas? Elsewhere? Briefly explain your answer.
- (c) You are surprised to observe that mice homozygous for the transgene insertion display a serious heart defect. (Heterozygotes have normal hearts.) Suggest a possible explanation.
- (d) Propose an experimental test of your hypothesis from part (c) using gene targeting in mice. Draw a labeled diagram of the targeting construct that you would use, and outline any breeding experiments required to test your hypothesis.
- (e) Propose how you might use LacZ in a gene targeting experiment in mice to test whether the amylase gene is expressed exclusively in the pancreas. Draw a labeled diagram of the targeting construct that you would use, and outline any breeding experiments required to test your hypothesis.

7.03 Fall 2003 Problem Set 6 Key

- 1a) q , the frequency of allele a , is the number of a alleles divided by the total number of alleles. Since there are 1,000,000 individuals in the population, there are 2,000,000 alleles in the population. Each aa homozygote has two a alleles, while each Aa heterozygote has one a allele. We can calculate q as follows:

$$q = f(a) = ((2(500) + 9000) / 2000000) = \mathbf{0.005}$$

- b) $p + q = 1$, so $p = 1 - q = 1 - 0.005 = \mathbf{0.995}$

- c) One generation of random mating is sufficient to bring a population into Hardy-Weinberg equilibrium, regardless of whether the previous generation was in Hardy-Weinberg equilibrium.

$$\text{Therefore } f(aa) = q^2 = (0.005)^2 = 2.5 \times 10^{-5},$$

$$\text{and the total number of } aa \text{ individuals is } 2.5 \times 10^{-5} \times (1000000) = \mathbf{25}$$

Likewise, the total number of Aa individuals is

$$1000000(2pq) = 1000000(2 \times 0.995 \times 0.005) = \mathbf{9950}$$

- d) no. In the absence of selection and mutation, there will be no change in the frequency of each allele from generation to generation.

In generation one, there are $9000 + 2(500) = 10000$ a alleles

In generation two, there are $9950 + 2(25) = 10000$ a alleles

- e) no. If generation one was in Hardy-Weinberg equilibrium, then the number of homozygous aa albino individuals would be equal to q^2 . However, $q^2 * 1000000 = 1000000(0.005)^2 = 25$, but there were 500 albino individuals in generation one.

- f) The probability that an unaffected individual is a carrier is $p(Aa) / (p(Aa) + p(AA))$. For f and g , we can assume that the probability of an unaffected individual being a carrier is just $f(Aa)$, as the number of aa albino individuals is extremely small.

The probability of an unaffected individual being a carrier is

$$f(Aa) = 2pq = 2(0.005)(0.995) = 9.95 \times 10^{-3}$$

The chances that both parents are carriers and both pass on their a allele is

$$(1/4)(9.995 \times 10^{-3})^2 = \mathbf{2.48 \times 10^{-5}}$$

- g) The frequency of Aa carriers in generation one is $9000 / 1000000 = 9 \times 10^{-3}$
The chances that both parents are carriers and both pass on their a allele is
 $(1/4)(9.995 \times 10^{-3})(9 \times 10^{-3}) = \mathbf{2.24 \times 10^{-5}}$

- h) The albino parent is aa , and must therefore pass on an a allele.

The probability that the other parent passes on an a allele is

$$(1/2)(9.995 \times 10^{-3}) = \mathbf{5.00 \times 10^{-3}}$$

7.03 Fall 2003 Problem Set 6 Key p 2

2a) For a rare dominant disease you can assume that all affected individuals are heterozygous. Therefore the frequency of heterozygotes is $f(Aa)=1/1000$, and each of those individuals has one q allele, so $q=1/2000$.
For a dominant disorder with selective disadvantage S and mutation rate μ , $q=\mu/S$.
Therefore $1/2000=\mu/0.2$, and $\mu=0.0001$

b) For a recessive mutation, the incidence of disease is equal to q^2 . Therefore, $q=(1/1000)^{1/2}=0.032$, plugging that into $q^2S=\mu$: $(1/1000) \times 0.2 = 2 \times 10^{-4}$
Final answer: $\mu=2 \times 10^{-4}$

c) Plugging in the value for q calculated in part b into $q^2S=h$, we find that $h=0.032 \times 0.2 = 6.4 \times 10^{-3}$. Heterozygous advantage is phenomena where the heterozygotes of a detrimental allele are better equipped to survive, thus maintaining the detrimental allele in the population. Individuals heterozygous at this particular locus are more fit by a factor of $1+h$. Final answer: $h=6.4 \times 10^{-3}$

3. a) Given μ and S we can calculate q using the equation $q = \sqrt{\mu/S} = \sqrt{\frac{10^{-5}}{0.4}} = 0.005$.

The incidence of the disease in the population would be

$q^2 = (0.005)^2 = 2.5 \times 10^{-5}$ therefore, one in forty thousand people would be affected by the disease.

b) At steady state, $\Delta q = 0$, and $\Delta q_{sel} = \Delta q_{mut}$. Let us call the number of affected individuals n_h^{total} ; n_h^{total} is the sum of homozygotes arising from inbreeding and from random mating. The number of homozygotes arising from inbreeding are described by the equation:

$$n_h^{inbreeding} = Fq$$

where F is the inbreeding coefficient (in this case $1/16$) and q is the allele frequency. The number of homozygotes arising from random mating is described by the familiar equation: $n_h^{random} = q^2$. Adding the respective weights, as random mating corresponds to 90% and cousin marriage corresponds to 10% of the mating, we get a final equation of:

$$n_h^{total} = 0.9n_h^{random} + 0.1n_h^{inbreeding} = 0.9q^2 + 0.1Fq$$

We plug this modified in the n_h^{total} , into the Δq_{sel} equation to get

$$\Delta q_{sel} = -S(0.9q^2 + 0.1Fq)$$

Setting this equal to $\Delta q_{mut} = \mu = 10^{-5}$ and plugging in values for F and S , we get a final equation:

$$\Delta q = \Delta q_{mut} - \Delta q_{sel} = \mu - S(0.9q^2 + 0.1Fq) = 0$$

$$10^{-5} - 0.4[0.9q^2 + 0.1(\frac{1}{16})q] = 10^{-5} - 0.36q^2 - 0.0025q = 0$$

$$q = 0.002839 \quad \text{or} \quad q = -0.009784$$

7.03 Fall 2003 Problem Set 6 Key p 3

As we cannot have negative allele frequencies, final answer is $q=0.00284$. How does this allele frequency compare to part a? To find the incidence of the disease, we plug in our q value into the $n_h^{total} = 0.9q^2 + 0.1Fq$ equation. This gives us: $n_h^{total} = 0.00002574$. Final answer: incidence=1 in forty thousand. This is the same number obtained due to random mating-why do you think this is the case?

The allele frequency is reduced because inbreeding results in a greater proportion of homozygotes, because homozygotes are selected against, those alleles fall out of the population. Why is the incidence the same? Eventhough the allele frequency is low, much more of the inbreeding population is homozygous at this locus relative to the randomly mating population, resulting the same incidence of disease at equilibrium

c) The allele frequency can be calculated by

$$q_{new} = q_{old} + \Delta q \approx q, \text{ after one generation}$$

$$\Delta q = -Sq^2 + \mu = -0.4q^2 + 10^{-5}$$

Assuming that our inbred population was at steady-state, we can use a $q = 0.00284$, giving us

$$\Delta q = -0.4(0.00284)^2 + 10^{-5} = 6.77 \times 10^{-6}$$

$$q_{new} = 0.00284 + 6.77 \times 10^{-6} = 0.00285$$

$$q_{new}^2 = (0.00285)^2 = 8.104 \times 10^{-6}$$

Incidence of disease right after cessation of inbreeding is approximately one in 125,000. This is much lower than at Hardy-Weinberg equilibrium.

d) The allele frequency would be expected to rise after cessation of inbreeding until $q=0.005$. This is due to the fact that the mutation rate is creating more alleles than are being selected out in the population. The allele frequency, q , approaches 0.005 as time approaches infinity.

7.03 Fall 2003
Problem Set #7
Solutions Key

1. (a) Let's denote the disease allele as - and the wild-type allele as +. The "grandfather's" genotype (top row) at the disease locus must be +/-, since not all of his offspring (the "parents") are affected by the autosomal dominant disease. Therefore, all of the affected individuals on this pedigree are heterozygous (+/-) at the disease locus.

We can determine the phase relationship between the disease and SSR62 in each of the affected parents (second row) by looking at the genotypes of the grandparental generation and determining what alleles each grandparent passed on to affected parent. At SSR62, the affected mother received allele B from the unaffected grandmother. She also received a + at the disease locus from the grandmother, whose genotype at the disease locus is +/+. Therefore, the affected mother must have received the rest of her alleles, SSR62 A and the disease allele -, together from the affected grandfather. So we know the phase relationship in the affected mother looks like the following:

SSR62	Disease
A	-
B	+

Similarly, the phase in the affected father can also be found by looking at what alleles he received from the grandparental generation. The phase in the affected father looks like the following:

SSR62	Disease
A	-
C	+

Now we look at the offspring generation (bottom row). In all cases, SSR62 A is in phase with -, the disease allele. In the four offspring on the left side, the three affected offspring received SSR62 A and - from the affected mother. These three belong to the parental (non-crossover) class. But the unaffected daughter received SSR62 A and + from the affected mother, which only could have happened if there were a crossover event between the SSR62 locus and disease locus. Therefore, she belongs to the recombinant (crossover) class. On the right, the affected father gave his two affected sons SSR62 A and -, and gave SSR62 C and + to his two unaffected children, so all four children on the right are in the parental class.

Among the eight offspring, there **are seven parentals and one recombinant**, so the LOD score calculation looks like the following:

$$\text{LOD}_{\theta=0.1} = \log \left[\frac{(0.45)^7 (0.05)}{(0.25)^8} \right] = 1.088$$

We are not finished yet. We must also consider the three parents in the middle row. However, we do not know the phase of the affected grandfather, so the LOD score calculation for the three middle row parents will involve the phase unknown formula. The two possible phases for the affected grandfather are:

SSR62	Disease	
A	-	Phase I
B	+	

SSR62	Disease	
B	-	Phase II
A	+	

In the middle row, the two affected parents in got SSR62 A from the affected grandfather, while the unaffected female got SSR62 B from the affected grandfather. So **if we assume phase I, there are three parentals. If we assume phase II, there are three recombinants.** The LOD score calculation would look like the following:

$$\text{LOD}_{\theta=0.1} = \log \left\{ \frac{[0.5(0.45)^3 + 0.5(0.05)^3]}{(0.25)^3} \right\} = 0.465$$

The total LOD score for linkage between SSR62 and the disease in this family would be the sum of the two scores, since the events are independent of each other.

$$\text{LOD}_{\theta=0.1} = 1.088 + 0.465 = 1.553$$

(b) Using the same reasoning as in part a, we can determine the phase relationship between SSR93 and the disease in the affected mother and father:

Affected mother:

SSR93	Disease	
B	-	
A	+	

Affected father:

SSR93	Disease	
B	-	
A	+	

Now that we know the phase, we must look at the eight offspring. The only recombinant is the unaffected daughter on the right side, who received SSR93 B and + from the

affected father. All the other seven offspring are parentals, since they received from the affected parents either SSR93 B and - or SSR93 A and +.

Again, **there are seven parentals and one recombinant**. So the LOD score is the same as in part (a):

$$\text{LOD}_{\theta=0.1} = \log [(0.45)^7(0.05) / (0.25)^8] = 1.088$$

Again, we must also consider the three parents in the middle row. The two possible phases for the affected grandfather are:

SSR93	Disease	
B	-	Phase I
C	+	

SSR93	Disease	
C	-	Phase II
B	+	

In the middle row, the two affected parents both got SSR93 B from the affected grandfather, while the unaffected female got SSR93 C from the affected grandfather. So **if we assume phase I, there are three parentals. If we assume phase II, there are three recombinants**. The LOD score calculation would look like the following:

$$\text{LOD}_{\theta=0.1} = \log \{ [0.5(0.45)^3 + 0.5(0.05)^3] / (0.25)^3 \} = 0.465$$

The total LOD score for linkage between SSR93 and the disease in this family would be the sum of the two scores, since the events are independent of each other.

$$\text{LOD}_{\theta=0.1} = 1.088 + 0.465 = 1.553$$

(c) To calculate linkage between the two SSR's, we disregard whether a person is affected or unaffected and turn our attention exclusively to the SSR genotypes. We are no longer restricted to following the alleles of the two affected parents. Therefore, we can consider the passage of alleles from all four parents and the two grandparents to their respective offspring independently:

(i) Left side middle row father: This father is homozygous for SSR62 C. Therefore, **he is uninformative for the LOD score calculation**, since crossing over between the SSR62 and SSR93 loci will make no difference in the alleles that are passed on (i.e. we cannot distinguish between recombinant and parental classes).

(ii) Left side middle row mother: This mother is heterozygous at both SSR loci, so she is statistically informative. We can also determine her phase by looking at what alleles she received from the grandparents. From the grandmother, she definitely received SSR93 A and SSR62 B, so these two alleles are in phase. Therefore, she received her other alleles, SSR62 A and SSR93 B, together from the grandfather. Her phase looks like the following:

SSR62	SSR93
A	B
B	A

Looking at the four offspring on the left, we can see that they all received SSR62 C from their father and SSR62 A from their mother. At SSR93, three offspring received SSR93 B from their mother. They are parentals, since SSR62 A and SSR93 B are in phase. The unaffected daughter received SSR93 A and SSR62 A from her mother. This could only have happened if there were a crossover between the two SSR loci, so she is a recombinant. With **three parentals and one recombinant**, the LOD score calculation looks like the following:

$$\text{LOD}_{\theta=0.1} = \log [(0.45)^3(0.05) / (0.25)^4] = 0.067$$

(iii) Right side middle row father: This father received SSR93 A and SSR62 C from the grandmother, so these two alleles are in phase. Therefore, he received his other alleles, SSR62 A and SSR93 B, together from the grandfather. His phase looks like the following:

SSR62	SSR93
C	A
A	B

Looking at the four offspring on the right, we can see that there are again three parentals and one recombinant. The first two sons both received SSR62 A and SSR93 B from their father and the last son received SSR62 C and SSR93 A from the father. So these three are the parentals. The daughter received SSR62 C and SSR93 B from the father. She is a recombinant. With **three parentals and one recombinant**, the LOD score is the same as above in (ii):

$$\text{LOD}_{\theta=0.1} = \log [(0.45)^3(0.05) / (0.25)^4] = 0.067$$

(iv) Right side middle row mother: This mother is homozygous for SSR93 C, so she is uninformative for the same reason as the left side father.

(v) Top row grandfather: The phase relationship between the two SSRs is unknown in this individual. The two possible phases are:

SSR62	SSR93	
A	B	Phase I
B	C	

SSR62	SSR93	
B	B	Phase II
A	C	

In the middle row, the three parents (left to right) received the following SSRs from the grandfather: SSR62A SSR93B, SSR62B SSR93C, and SSR62A SSR93B. So **if we assume phase I, there are three parentals. If we assume phase II, there are three recombinants.** The LOD score calculation would look like the following:

$$\text{LOD}_{\theta=0.1} = \log \left\{ \frac{[0.5(0.45)^3 + 0.5(0.05)^3]}{(0.25)^3} \right\} = 0.465$$

(vi) Top row grandmother: She is uninformative because she is homozygous at the SSR93 locus.

We can add the LOD scores for the left side mother, right side father, and the grandfather because they passed on alleles to their offspring independently (for the same reason we can add the LOD scores from different families). So the final LOD score for linkage between SSR62 and SSR93 in this family is:

$$\text{LOD}_{\theta=0.1} = 0.067 + 0.067 + 0.465 = 0.599$$

(d) From now on, the alleles for SSR93 will be known as A', B', and C'. To determine the relative order of the three markers, we must look at the two individuals who were considered recombinants in parts (a) and (b). In part (a), the recombinant individual is the unaffected female in the last row (second from left). She received the following alleles from her affected mother: A, A', +. In the affected mother, the three possible allele orders are (phase known):

62	93	disease	
A	B'	-	order 1
B	A'	+	

93	62	disease	
B'	A	-	order 2

A' B +

62 disease 93
A - B' order 3

B + A'

Since we know that the mother gave the unaffected daughter A, A', +, we look at how many crossovers are required to give these three alleles together. In orders 1 and 3, one crossover is needed. But in order two where SSR62 is in the middle, a double crossover is needed. **Since a single crossover is much more likely than a double crossover, we can rule out order 2 and conclude that SSR62 cannot be the middle marker.** However, we cannot distinguish which of the two other orders is more likely, so we need to look at another recombinant individual (the one from part b).

We now shift our attention to the unaffected daughter on the right side of the family (bottom row, second from right). She got the following alleles from her affected father: C, B', +. In her affected father, the three possible orders were:

62 93 disease
A B' - order 1

C A' +

93 62 disease
B' A - order 2

A' C +

62 disease 93
A - B' order 3

C + A'

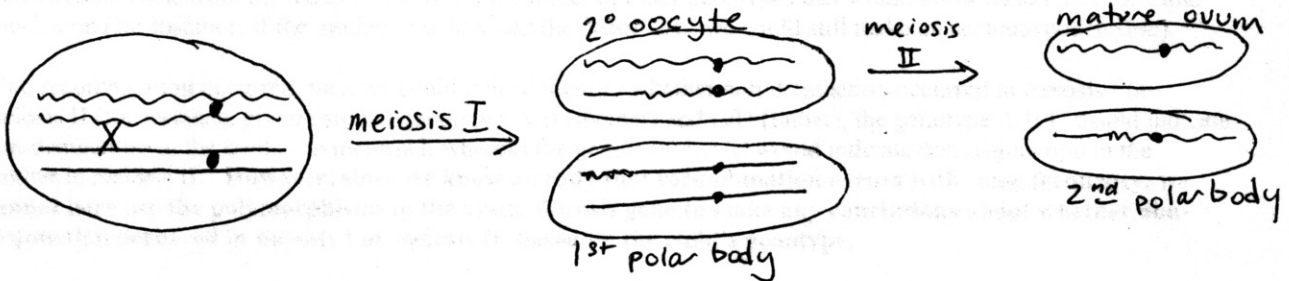
Now we use the same logic to eliminate order 1 so **SSR93 cannot be the middle marker** (i.e. there would have had to been a double crossover in order 1 for the father to give C, B', and + alleles together to the daughter).

This analysis leaves order 3 as the most likely order:

SSR93 disease SSR62

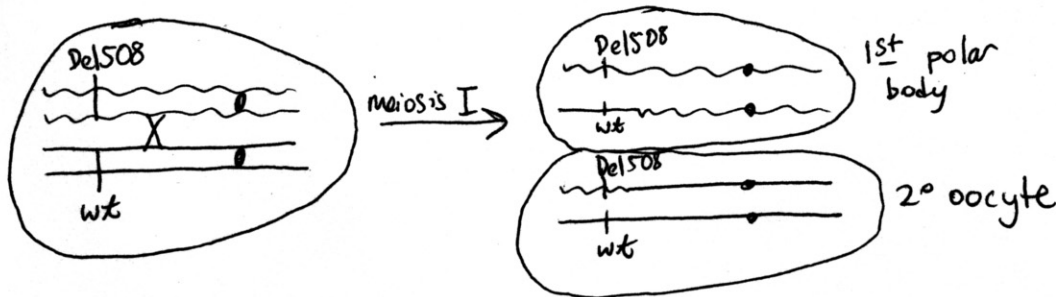
Note: the best solution to this question (the one described above) was very subtle and required a familiarity with material introduced earlier in the course, namely the superiority of three factor crosses in determining genetic order and distance. A common incorrect solution to this problem involved using the LOD scores calculated in earlier parts to predict order. This method is weaker because LOD scores essentially measure two factor distances (they are two factor crosses between two markers).

2) During ovulation, the primary oocyte divides into a polar body and a secondary oocyte. This division is the meiosis I division. As in any meiosis I division, it follows any recombination events, and it leads to the separation of homologous chromosomes via the centromeres. Next, the secondary oocyte undergoes meiosis II, which is the separation of sister chromatids, via the centromeres. This produces a second polar body and the mature ovum.

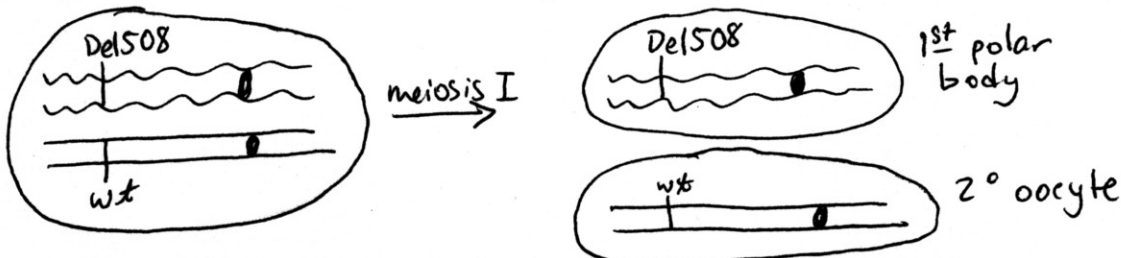


What we are observing in these data are the genotypes of the polar bodies. These polar bodies are being tested because they will not go on to become the egg (mature ovum) that would be fertilized by the sperm.

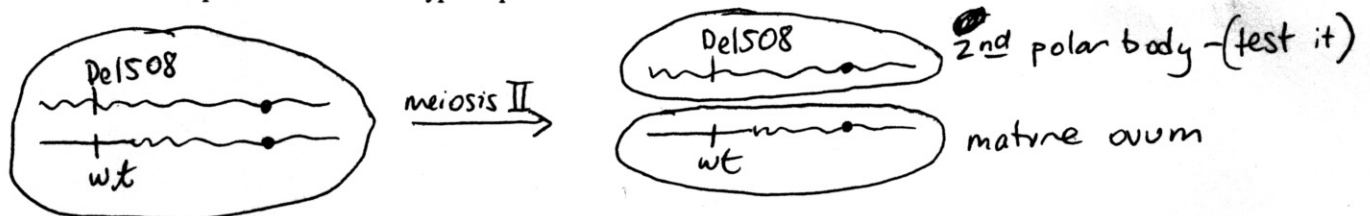
a) The polar bodies of oocytes #1 and #5 test positive for both the Del508 and the Wildtype sequence because a **recombination** event occurred between the centromere and the site of the cystic fibrosis gene prior to Prophase I. The other oocytes test positive for only one of the two sequences because no such recombination event occurred.



b) Since the secondary oocyte will go on through further development (eventually to make the mature ovum), and the polar body that we are testing will not develop further, we want the polar body to carry both copies of Del508. If the polar body carries both copies of Del508, we can assume that the secondary oocyte will have both of the wildtype copies. Therefore, we would choose **oocytes #2 or #3** to use in *in-vitro* fertilization.



c) The Meiosis II division creates the mature ovum and a second polar body, from the secondary oocyte. If we could PCR test the second polar body and then return the mature ovum to the women's uterus, then we would choose an ovum where the second polar body tested positive for the Del508 (and the first polar body was already known to be positive for both sequences). If the second polar body of **oocytes #1 or #5** tested positive for the Del508, then we would feel confident in returning the mature ovum from these oocytes to the women's uterus, knowing that the mature ovum was positive for the wildtype sequence.



d) **If we were interested in identifying the parental origin of an additional chromosome 7, we could use polymorphisms in the cystic fibrosis gene, assuming each parent has a unique set of polymorphisms.** Suppose the mother is A/B and the father is C/D. Then, if the child were A/B/C, we would know that the additional chromosome came from the mother. There are a number of other genotypes that would allow us to make the same conclusion (for instance, if the mother was A/A and the father C/C, we could still make an accurate conclusion).

If no recombination occurred, then we could indeed identify whether non-disjunction occurred in meiosis I or meiosis II (for instance, given parental genotypes A/B (mother) and C/D (father), the genotype A/B/C would indicate non-disjunction in the mother in meiosis I, whereas the genotype A/A/C would indicate non-disjunction in the mother in meiosis II. **However, since we know already that recombination occurs with some frequency, we cannot here use the polymorphisms in the cystic fibrosis gene to make any conclusions about whether non-disjunction occurred in meiosis I or meiosis II, based on the child's genotype.**

7.03 Problem Set 4 Key

1. a. We can think of lac permease activity (from LacY⁺) and beta-galactosidase activity (from LacZ⁺) as reporters for Lac operon expression from the bacterial chromosome and the F', respectively. In other words, lac permease activity tells us whether the operon on the bacterial chromosome is regulated, constitutive, or uninducible; and beta-galactosidase activity tells us whether the operon on the F' is regulated, constitutive, or uninducible.

Lac1- is a dominant, cis-acting mutation. It is dominant because lac permease activity is still constitutive in the merodiploid with one wild-type allele and one Lac1- mutation. Lac1- is cis-acting because it can only confer constitutive expression on the reporter in cis with itself (LacY⁺), while the reporter is trans (LacZ⁺) is regulated.

Lac2- is a dominant, trans-acting mutation. It is dominant because lac permease activity is still constitutive in the merodiploid with one wild-type allele and one Lac2- mutation. Lac2- is trans-acting because it can confer constitutive expression on both the reporter in cis (LacY⁺) and the reporter is trans (LacZ⁺).

Lac1- best fits the properties of LacO^C mutations (cis-acting dominant) and Lac2- best fits the genetic characteristics of LacI^d mutations (trans-acting dominant).

- b. LacP- mutations result in decreased expression of Lac operon genes, nevertheless, expression is still regulated by IPTG. A LacO^C mutation, on the other hand, would result in unregulated expression of Lac operon genes. In other words, the activity of both beta-galactosidase and lac permease would be independent of IPTG in LacO^C. Because the promoter is required for expression and the operator is involved for regulation of the operon, this implies that LacP and LacO represent functionally distinct sites on the DNA.

It is possible to imagine mutations in the operator that would give results like the LacP- mutation, thus confounding the conclusion we just drew. For example, there could be a mutation in LacO that results in a greater binding affinity with the Lac repressor. In such a state, expression of the operon's genes can almost be completely off in the absence of IPTG and only slightly induced in the presence of IPTG (like the LacP- mutations). We can, however, differentiate this from LacP- with further experimentation. With a LacO mutation we can increase expression of the operon by increasing the amount of IPTG because the operator is again the region regulated by the inducer. On the other hand, LacP- there will an insensitivity to IPTG inducer pass a particular point because its function is dependent on binding to RNA polymerase. In any case, this scenario is unlikely, as it is much more difficult to get a mutation in LacO that produces a higher binding affinity with Lac repressor than it is to get a mutation that produces a lower binding affinity.

c. **In Lac3-, a part of LacI is deleted in addition to the promoter.** From the data, we can deduce that the wild-type allele of Lac3- is a trans-acting factor, as beta-galactosidase activity becomes regulated when the wild-type allele (brought in on F') is in trans with wild-type LacZ. The wild-type LacI gene product can act to regulate the expression of both the chromosomal and the F' operon. The Lac3- mutation would result in a non-functional gene product (LacI-, which is consistent with the data.

In Lac4-, a part of LacO is deleted in addition to the promoter. From the data, we can deduce that the wild-type allele of Lac4- is a cis-acting factor, as beta-galactosidase activity is still unregulated when the wild-type allele (from the F') is in trans with the wild-type LacZ. The Lac4- mutation would result in an operator that the Lac repressor can no longer bind, so expression is constitutive (LacO^C), although at the decreased overall level of expression compared to wild-type owing to the promoter deletion.

Lac4- could possibly represent a LacI^d like mutation as judged from the data, but we know that such mutation affect the N-terminus of the protein product (the 5' part of the coding strand), so this is unlikely because the majority of the LacI gene would have to be deleted to reach the 5' region.

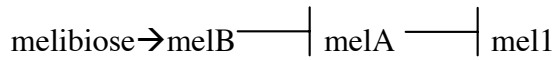
2. a. **We can classify MelA- as a recessive, loss of function mutation, whose wild-type allele codes for a trans-acting factor.** Because an amber suppressor can restore the wild-type phenotype, the MelA- mutation is most likely an amber mutation. Amber suppressors are only about 5% effective, thus the mutant allele is most likely recessive because 5% wild-type protein is unlikely to suppress the mutant phenotype for a dominant mutant allele. Because the wild-type allele codes for a functional protein product, we know that it is a trans-acting factor. We also know that MelA and Mel1 are different genes because they are unlinked.

MelA is likely a negative regulator of Mel1 expression. The MelA- mutation causes constitutive activity, which is consistent with loss of function of a negative regulator.

b. **MelB- is a recessive mutation whose wild-type allele codes for a trans-acting factor that is not Mel1.** We know that MelB- is recessive because MelB-/F' Mel+ results in regulated melibiose activity, rather than uninducible. Mel1- is also a recessive mutation using the same reasoning as with MelB-. With two recessive mutation, a complementation test is then performed (MelB-/F' Mel1-). The result is that MelB- and Mel1- complement each other, meaning that they are mutations in different genes. This strain (MelB-/F' Mel1-) also allow us to deduce that MelB's wild-type gene product is trans-acting because we observe regulated activity when wild-type MelB is placed in trans with wild-type Mel1.

MelB is likely a positive regulator of Mel1 expression. The MelB- mutation causes uninducible activity, which is the result of a recessive mutation in a positive activator.

2c) We know that melibiose and MelB both have a net positive effect on MelI activity, and MelA has a net negative effect on MelI activity. Therefore there are two possible models:



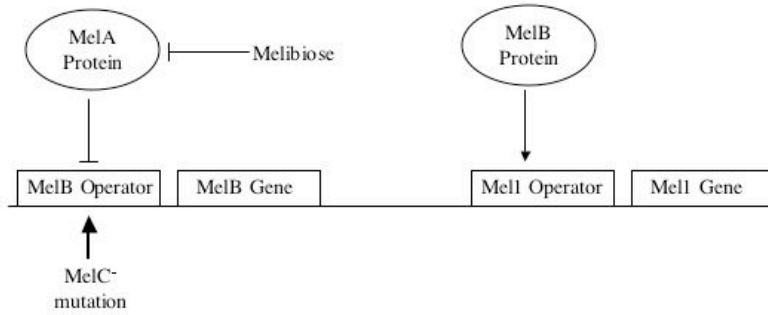
There could be other models consistent with this data, but these two models are the simplest.

2d) The phenotype of the MelA⁻ MelB⁻ double mutant is identical to the MelB⁻ mutant. Since MelB⁻ is epistatic to MelA⁻, the correct pathway must be:

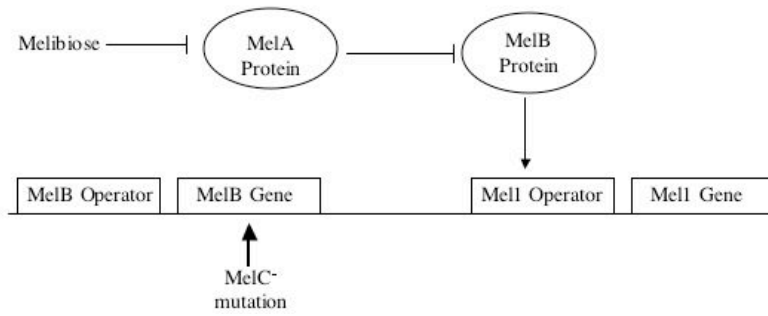


2e) Because MelC⁻ / F⁺ Mel⁺ has melibiase activity in both the presence and absence of melibiose, we know that MelC⁻ is dominant and constitutive. Furthermore, because MelC⁻ causes constitutive melibiase activity in either cis or trans to MelI⁺, it is a trans-acting factor. Therefore the two possibilities consistent with this data are MelC⁻ being either a dominant negative mutation in a negative regulator (repressor^d) or a super activator mutation (activator^s).

2f) If we assume that MelB⁻ and MelC⁻ are mutations in the same activator gene, then MelB⁻ must cause a loss-of-function in the activator, while MelC⁻ must cause a gain-of-function that renders the activator unresponsive to the MelA negative regulator protein. One possibility is that the MelA protein normally binds to the MelB protein, and the MelC⁻ mutation affects the protein such that MelA can no longer bind. Alternatively, MelA may be a repressor that binds to an operator in front of the MelB gene, and the MelC⁻ mutation is a mutation in the operator of MelB, causing constitutive MelB expression. The easiest way to distinguish these two possibilities from each other would be by sequencing the MelB gene and the upstream sequences in both wt and MelC⁻ bacteria. By comparing the sequences from wt and MelC⁻ it would be possible to determine where the MelC⁻ mutation lies. If the mutation lies in the operator for the MelB gene then we would know that MelA is a repressor that binds to an operator in front of MelB. If the mutation lies in the MelB gene itself then we would know that MelA protein negatively regulates the MelB protein.



MelC⁻ is a mutation in MelB operator



MelC⁻ is a mutation in MelB protein coding sequence

Answer Key Problem Set 5

1. a) Genetic properties of *gln2*- and *gln3*-:

Both are **uninducible**, as they give decreased glutamine synthetase (GS) activity.

Both are **recessive**, as mating them with wildtype produces normal GS activity.

Both are **trans-acting**, as when either mutation is crossed to *gln1*-, complementation takes place; if either *gln2*- or *gln3*- were a cis-acting element regulator of *gln1*, we would not see complementation, as the two mutations would be in the same gene. (i.e. the promoter is considered to be part of the gene it acts on.) Since we see complementation in both cases, we know that both mutations must be trans-acting.

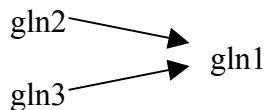
Therefore both *gln2*- and *gln3*- are mutations in genes for **positive regulators**.

b) Linear models:

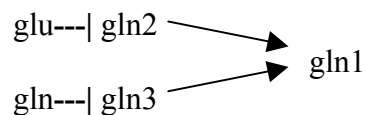
$gln2 \rightarrow gln3 \rightarrow gln1$

$gln3 \rightarrow gln2 \rightarrow gln1$

Parallel model:



c) The **parallel** model from part b) best fits the experimental results, as we are told that the addition of glutamate (*glu*) and glutamine (*gln*) independently regulate the GS activity (i.e. the results of the addition of each is different from the addition of both).



d) We would expect the *gln2*-*gln3*- double mutant to be **uninducible**.

e) The cis-acting elements in the promoter are:

- **Deletion 1** (-300 to -250): **GLN2 function** (activity pattern looks like *gln2*- when deleted), **UAS** (Upstream Activation Sequence -involved in activation)
- **Deletions 3, 4** (-200 to -100) **GLN3 function** (activity pattern looks like *gln3*- when deleted), **UAS** (Upstream Activation Sequence -involved in activation)
- **Deletion 6** (-50 to 0) TATA sequence

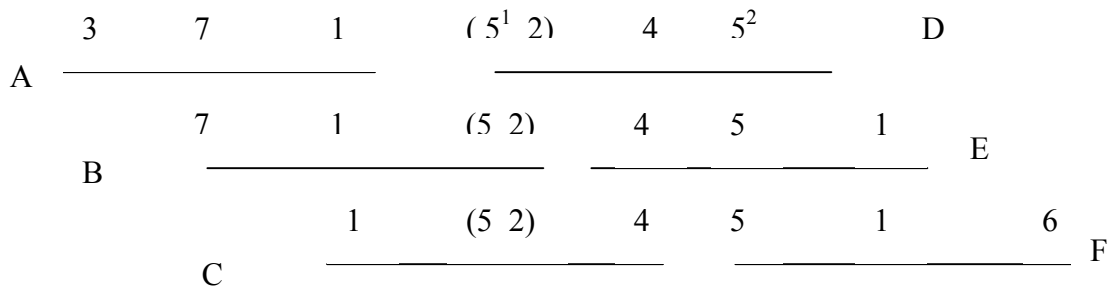
f) We would expect to see no **change in beta-galactosidase expression in Deletion 1** in a *gln2* mutant, as, in e), we determined that the region deleted in 1 is necessary for GLN2 function. (i.e. deleting both the cis-acting element and *gln2* will look the same as deleting either one.)

We would expect **0 units of betagalactose from Deletion 4** (or low baseline level), as deletion 4 is necessary for GLN3 function, and we know that when both the *gln2* and *gln3* pathways for activating GLN1 are mutated neither pathway can activate GLN1.

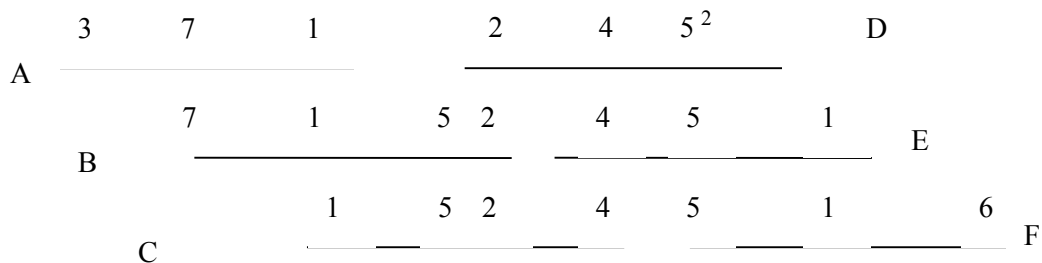
2.

- a) We can use the information from the order of the BACs to figure out where each STS is. For instance, BAC A is positive for STS 3, which is not present in any other BAC; therefore it must be in the region of A that does not overlap with B or C. There are three ambiguities in the data, resulting in three possible maps. Most of the ambiguity lies in whether BAC D contains both STS 51 and STS 52, as PCR only provides the qualitative answer that STS 5 is present; we don't get quantitative data as to how many versions of STS5 are present on the BAC. Case 1 displays the possibility that BAC D contains both STS5s. Given this case, we cannot determine the order of STS51 and STS 2 (thus the parenthesis). Case 2 displays the possibility that STS51 is not on BAC D. If this assumption is true, then we can determine the order of STS51 and STS2. Case 3 displays the possibility that BAC D contains only STS51. If this is the case, we cannot determine the relative order of STS52 and the second STS1.

CASE 1



CASE 2

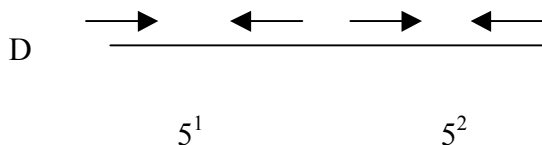


Case Three:



Therefore there are quite a few ambiguities in our map. The first is that we do not know the order of STS 51 relative to STS2. It is possible that STS 2 comes before STS 51 or STS 51 comes before STS 2. Also, we do not know if D is carrying STS 51 and 52 or if it is carrying just 52. Consider the possibility that D contains only STS 52; we would have to assume that STS 2 came after STS51, and the STS 5 we are detecting comes from STS 52. In addition, there is a third possibility, which is that the STS 5 contained on BACD is only STS51. If that is the case, then we don't know the order of STS5 with respect to the later STS1. We need an assay to distinguish between the possibilities.

b) . PCR just amplifies DNA to which primers bind. One could imagine that the PCR primers could anneal to both sites and amplify both sites. This would result in a heterogeneous mixture of STS51 and STS52.



However, PCR alone does not help us distinguish between the three possibilities. This sequencing assay is our assay to distinguish between the three cases. Sequencing can tell us whether BACD contains both STS 51 and STS52, just STS51 or just STS52. Given these data, it looks as if BAC D is carrying both STS5s, which allows us to throw out cases two and three. There is a remaining ambiguity—we still do not know the order of STS51 relative to STS2.

c) depending on your answer it could confirm/refute. Most answers should be refined, as we can now distinguish between the three cases. There are some remaining ambiguities, but the map is significantly more refined after the sequencing assay of STS5.

d) You would expect both sequences to be present in the genome. BACs are derived from the mouse genome and are therefore a reflection of what is present in the genome. If we find to STS 5 in twice our BACS, there are probably at least two STS 5 on the mouse chromosome. Only BAC D contains both STS5s.

e) STS1 is present more than once; BACs A, B, C would have the same sequence E, F would have the same sequence. It is possible that the STS1 from BACs A, B,C have a different

sequence than those from STS1 from BACs E, F. Primers that distinguish between these two STS1 would be convenient as we would not have to be concerned about an ambiguity in our data in the future like there was with STS5 and BAC D.

- f) Design primers that include the non-homologous sequence in STS51 and STS52. These primers should specifically amplify either STS51 or STS52.
- g) Yes STS51 and STS52 should both be present in BAC D and the mouse genome {refer to (d)}

- 3. a. The integration of Pamyase-LacZ into the amylase gene is unlikely to occur when the construct is microinjected into the male pronucleus of a fertilized egg.

Integration of the construct into the endogenous amylase gene locus could occur via homologous recombination or by chance non-homologous insertion into the endogenous amylase gene. Homologous recombination requires homologous sequences to the targeted locus on both sides of the LacZ gene. The microinjected construct has homology on only one side of the LacZ transgene, thus, making homologous recombination an unlikely event in this case.

Although random insertion of a construct occurs with greater frequency than homologous recombination, the chance of a transgene inserting into a particular locus is exceedingly rare. The amylase gene is on the order of a couple of kilobases (kb), while the mouse genome roughly 3000 megabases (Mb). This makes the likelihood of randomly inserting into the amylase locus approximately 1 in 10^6 .

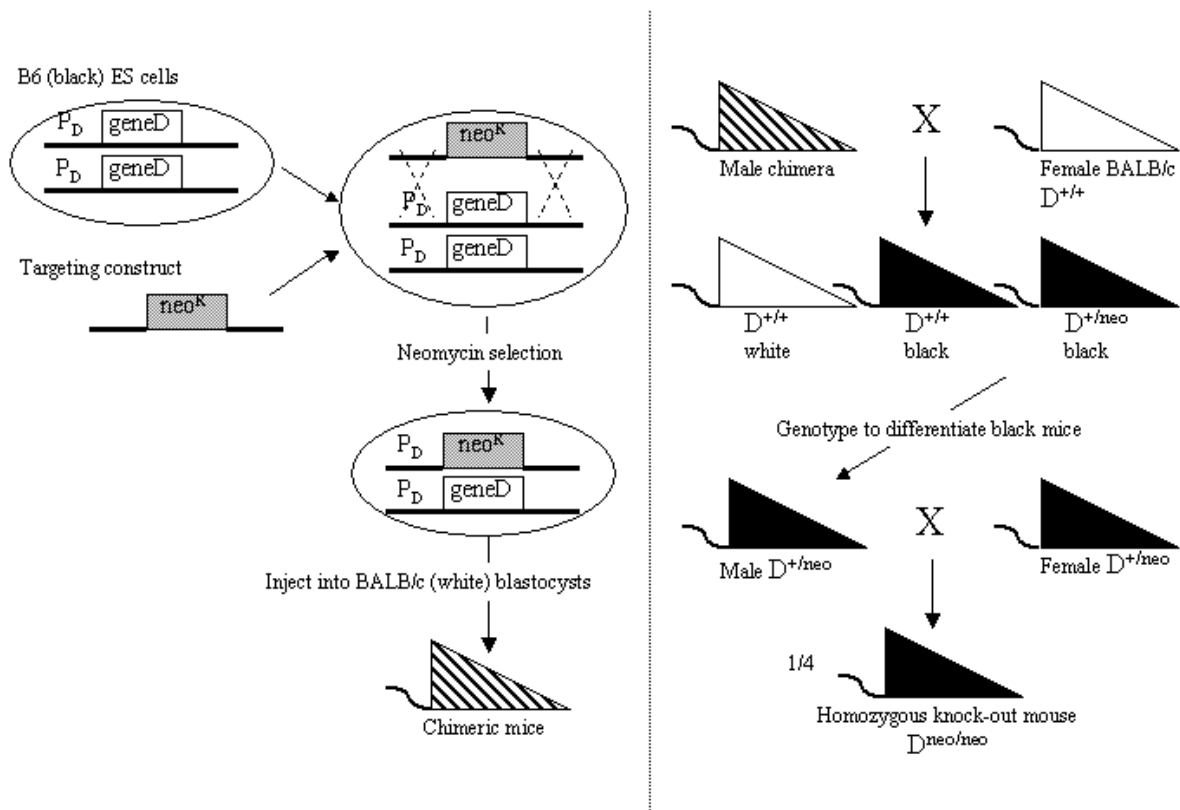
- b. We would expect mice homozygous for the transgene to display LacZ expression in the pancreas as the heterozygous mouse does, assuming that the homozygote is a result of a cross between two heterozygous mice of the same transgenic line. In the homozygote the regulation of both copies of the transgene would most likely be equivalent to the regulation of the transgene in the heterozygote.

Expression of the LacZ transgene in the homozygote may also be observed in other organs. This can occur if expression from one copy of the transgene (as in the heterozygote) is just below the level of detection. Two transgenes in a single cell (as in the homozygote) could produce just enough LacZ to be detectable.

- c. A possible explanation for the heart defect in mice homozygous for the transgene insertion is that the insertion disrupted a gene (let's call it gene D) that is haplosufficient. A single undisrupted copy of gene D could be sufficient to allow development of a normal heart, but when both copies are disrupted the mouse develops the heart defect.

Another possibility is the LacZ overexpression (with two copies of the transgene) causes the defect. When only a single LacZ transgene is present there is not enough expression to disrupt the normal development or the functioning of the heart.

d. We can test the possibility that disruption of both copies of a gene D causes the heart defect. First, we would want to clone the locus (gene D) into which the transgene integrated. This can be done by anchored PCR and sequencing using the sequence of the integrated transgene to design primers. Next, we can construct a targeting construct to knock-out gene D, in which we basically swap the coding region of gene D for a drug resistance gene (e.g. neomycin). This targeted deletion of gene D can be selected for by the presence of the drug resistance gene after ES cells in culture are transfected with the targeting construct. The ES cells containing the correctly targeted locus can then be injected into blastocysts to generate chimeric mice. The chimeras can then be mated to wild-type mice to produce progeny that are heterozygous for the targeted deletion of gene D. (We can use a combination of coat color determination and PCR analysis to figure out the genotype of the mice.) We can generate mice homozygous for the deletion of gene D by breeding heterozygous mice. If our hypothesis is correct, then we would expect to see a heart defect in the homozygous knock-out mice and normal hearts in the heterozygous mice.



e. To use LacZ as a reporter for the expression of the endogenous amylase gene, we would want to put LacZ under the control of the same regulatory elements. Because it is difficult to determine the boundaries of promoter regions, we would want to insert LacZ in place of the amylase gene in the endogenous locus. To accomplish this, we would need a targeting construct that contains the LacZ gene flanked by sequences homologous

to amylase upstream and downstream sequences. The resulting heterozygous knock-out (knock-in) mice (after ES cells targeting, blastocyst injection to generate chimera, and mating of chimeras to wild-type mice) can be assayed to determine if LacZ is expressed in the pancreas. The expression of LacZ is likely to reflect amylase expression, but will not certainly do so.

