7.03

Problem Sets

2002

7.03 Problem Set 1

Due before 5 PM on Thursday, September 19, 2002
Hand in answers in recitation section or in the box outside the class

1. You and your lab partner have isolated 20 new mutant yeast strains that are defective in synthesis of threonine, an amino acid. These Thr⁻ mutants do not grow on minimal medium, but they do grow on minimal medium supplemented with threonine. Ten of your Thr⁻ mutants (numbered 1 through 10) were isolated in a strain of mating type **a** (MAT **a**). The other 10 Thr⁻ mutants (numbered 11 through 20) were isolated in a strain of mating type α (MAT α). You and your lab partner cross each of the MAT **a** strains to each of the MAT α strains, and you include crosses to the appropriate wild-type strains. Your experimental observations are shown in the table below, where (-) indicates diploids that did not grow on minimal medium and (+) indicates diploids that did grow on minimal medium.

(a) Unfortunately, when all the data were collected and the plates discarded, your dog munched on your notebook, swallowing some data. From the data that remains, see if you can reconstruct the full table.

strains of mating type α

	wild-type	1	2	3	4	5	6	7	8	9	10
wild-type	+		+	+			+		+	-	+
11			-				+	+			
12	+	-			+		-			-	
13		-	+	+	+						+
14	-			-			-		-		-
15			-	+		-			+		
16		+		-		+				-	-
17		+			-	+			-	-	
18	+		+			+	-	-			+
19			+					+	-		
20	+					-	+	+		-	

strains of mating type **a**

(b) Which mutations are recessive?

(c) Which mutations are dominant?

(d) Which mutations do you know to be in the same gene?

(e) Based on these experiments, what is the minimum number of genes required for threonine synthesis?

(f) What is the maximum number of genes that these 20 mutants could represent?

2. Being a well-rounded geneticist, you also maintain a colony of chickens. In a cross between a female chicken from a true-breeding strain with three toes per foot and a male from a true-breeding strain with seven toes per foot, all of the F_1 progeny have five toes per foot.

(a) Based on the information that you have at this stage, is it possible that a single gene determines the differences in toe count among the two parental strains and progeny? If so, what toe counts should appear in the F_2 generation and at what frequencies?

(b) In fact, when F_1 chickens are crossed among themselves, the following F_2 progeny are produced: 46 chickens with five toes per foot, 30 chickens with four toes per foot, 28 chickens with six toes per foot, 8 chickens with seven toes per foot, and 7 chickens with three toes per foot. Propose a genetic model to account for the existence of the 5 phenotypic classes and their observed frequencies. In your answer, provide the genotypes of the parental chickens, the F_1 chickens, and each class of F_2 chickens.

(c) Use the Chi-square test to show that the observed frequencies fit with the expected frequencies based on your model. For your answer, give the observed and expected phenotypic ratios, the degrees of freedom, your calculated value for χ^2, and a rough estimate of the p value.

3. Your friends Ben and Jane are contemplating having children and seek your genetic counsel. Ben and Jane are unrelated, but both have younger brothers with the same, <u>extremely rare</u> genetic trait. (In fact, Ben and Jane met at a support group meeting for children and families of children with this rare disorder.) Ben's parents and Jane's parents are all unaffected.

(a) Assume that the trait is autosomal recessive with complete penetrance. What is the probability that Ben and Jane's first child will be affected with the trait? If not affected, what is the probability that the child is a carrier?

(b) Assume that the trait is autosomal recessive with 80% penetrance. What is the probability that Ben and Jane's first child will be affected?

(c) Assume that the trait is X-linked recessive with complete penetrance. What is the probability that Ben and Jane's first child will be affected? What is the probability that the first child will be a carrier?

(d) Assume that the trait is autosomal dominant with 80% penetrance in heterozygotes and 100% penetrance in homozygotes. What is the probability that Ben and Jane's first child will be affected? If not affected, what is the probability that the child is a carrier?

———————————

Answers: 7.03 Problem Set 1

1. You and your lab partner have isolated 20 new mutant yeast strains that are defective in synthesis of threonine, an amino acid. These Thr⁻ mutants do not grow on minimal medium, but they do grow on minimal medium supplemented with threonine. Ten of your Thr⁻ mutants (numbered 1 through 10) were isolated in a strain of mating type **a** (MAT **a**). The other 10 Thr⁻ mutants (numbered 11 through 20) were isolated in a strain of mating type α (MAT α). You and your lab partner cross each of the MAT **a** strains to each of the MAT α strains, and you include crosses to the appropriate wild-type strains. Your experimental observations are shown in the table below, where (-) indicates diploids that did not grow on minimal medium and (+) indicates diploids that did grow on minimal medium.

(a) Unfortunately, when all the data were collected and the plates discarded, your dog munched on your notebook, swallowing some data. From the data that remains, see if you can reconstruct the full table.

strains of mating type α

	wild-type	1	2	3	4	5	6	7	8	9	10
wild-type	+	+	+	+	+	+	+	+	+	-	+
11	+	+	-	+	+	-	+	+	+	-	+
12	+	-	+	+	+	+	-	-	+	-	+
13	+	-	+	+	+	+	-	-	+	-	+
14	-	-	-	-	-	-	-	-	-	-	-
15	+	+	-	+	+	-	+	+	+	-	+
16	+	+	+	-	+	+	+	+	+	-	-
17	+	+	+	+	-	+	+	+	-	-	+
18	+	-	+	+	+	+	-	-	+	-	+
19	+	+	+	+	-	+	+	+	-	-	+
20	+	+	-	+	+	-	+	+	+	-	+

(Row labels at left: strains of mating type **a**)

Note: The table was filled in assuming that there are four complementation groups (see part *d*).

(b) Which mutations are recessive?

All except 9 and 14.

(c) Which mutations are dominant?

9 and 14.

(d) Which mutations do you know to be in the same gene?

[1, 6, 7, 12, 13, 18] [2, 5, 11, 15, 20] [3, 10, 16] [4, 8, 17, 19]
However, there are no data in the table that exclude the possibility that the third and fourth groups listed above actually form one large complementation group, i.e. [3, 10, 16, 4, 8, 17, 19]. Thus there are either three or four complementation groups.

(e) Based on these experiments, what is the minimum number of genes required for threonine synthesis?

Four (or three; see *d* above) (if the dominant mutations 9 and 14 are in genes represented by the complementation groups). This is the number of complementation groups.

(f) What is the maximum number of genes that these 20 mutants could represent?

Six (or five; see *d* above) (if mutations 9 and 14 are in genes *not* represented by the complementation groups). This is the number of complementation groups plus the number of dominant mutations.

2. Being a well-rounded geneticist, you also maintain a colony of chickens. In a cross between a female chicken from a true-breeding strain with three toes per foot and a male from a true-breeding strain with seven toes per foot, all of the F_1 progeny have five toes per foot.

(a) Based on the information that you have at this stage, is it possible that a single gene determines the differences in toe count among the two parental strains and progeny? If so, what toe counts should appear in the F_2 generation and at what frequencies?

One gene exhibiting incomplete dominance could be responsible for the observed data.
For example: AA – seven toes
              Aa – five toes
              aa – three toes
The progeny of a mating between five-toed chickens (Aa) would yield a variant on the typical 1:2:1 phenotypic ratio: 25% seven-toed (AA), 50% five-toed (Aa), and 25% three-toed (aa).

(b) In fact, when F_1 chickens are crossed among themselves, the following F_2 progeny are produced: 46 chickens with five toes per foot, 30 chickens with four toes per foot, 28

chickens with six toes per foot, 8 chickens with seven toes per foot, and 7 chickens with three toes per foot. Propose a genetic model to account for the existence of the 5 phenotypic classes and their observed frequencies. In your answer, provide the genotypes of the parental chickens, the F_1 chickens, and each class of F_2 chickens.

The spectrum of phenotypes suggests that more than one gene is involved, and the appearance of phenotypes not seen in either the P or F_1 generations suggests incomplete dominance.
A model consistent with the data invokes two genes (e.g., A and B), both exhibiting incomplete dominance, which act additively to determine toe number. The basal number of toes, given by phenotype aabb, is three toes, and each "big" allele contributes one additional toe. For example:

P   AABB (seven toes) × aabb (three toes)
                    ↓

F_1 AaBb (five toes)
                    ↓

F_2 1/16 AABB – seven toes
              1/4   AaBB, AABb – six toes
              3/8   AaBb, AAbb, aaBB – five toes
              1/4   Aabb, aaBb – four toes
              1/16 aabb – three toes

(c) Use the Chi-square test to show that the observed frequencies fit with the expected frequencies based on your model. For your answer, give the observed and expected phenotypic ratios, the degrees of freedom, your calculated value for χ^2, and a rough estimate of the p value.

|	Observed	Expected	(Obs-Exp)	$(Obs-Exp)^2/Exp$
Three toes	7	7.44	0.44	0.026
Four toes	30	29.75	0.25	0.002
Five toes	46	44.6	1.375	0.042
Six toes	28	29.75	1.75	0.103
Seven toes	8	7.44	0.56	0.042
Total	119	119		**0.215**

$\chi^2 = 0.215$, and degrees of freedom = [(5 classes) –1] = 4.
From the chi-square table, p ˜ 0.9. Thus we are unable to reject the hypothesised model.

3. Your friends Ben and Jane are contemplating having children and seek your genetic counsel. Ben and Jane are unrelated, but both have younger brothers with the same, extremely rare genetic trait. (In fact, Ben and Jane met at a support group meeting for

children and families of children with this rare disorder.) Ben's parents and Jane's parents are all unaffected.

(a) Assume that the trait is autosomal recessive with complete penetrance. What is the probability that Ben and Jane's first child will be affected with the trait? If not affected, what is the probability that the child is a carrier?

Aa   Aa   Aa   Aa
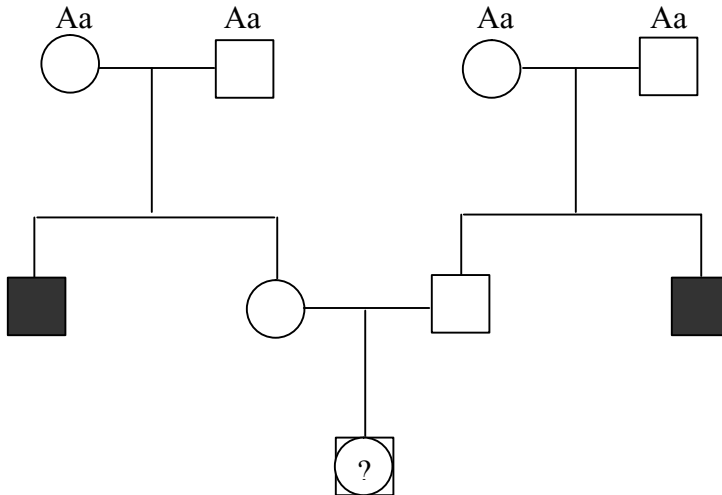
Let's call the recessive allele for the trait 'a' and its wild type counterpart 'A.'
We know that the genotype of the grandparents must be Aa because both Ben and Jane have brothers who are affected. Therefore:

p(child will be affected) = p(Jane is Aa) x p(Ben is Aa) x p (child is aa)

The probability that either parent is Aa is 2/3 because we already know that they are not affected (conditional probability). Therefore:

p(child will be affected) = $\frac{2}{3}$ x $\frac{2}{3}$ x $\frac{1}{4}$ = $\frac{1}{9}$

For the next question let's calculate the probability that the child is a carrier first.

p(child is a carrier) = p(Jane is AA)x p(Ben is Aa) x p(child is Aa) + p(Ben is AA)x p(Jane is Aa) x p(child is Aa) + p(Jane is Aa) x p(Ben is Aa) x p(child is Aa)
= 1/3 x 2/3 x 1/2 + 1/3 x 2/3 x 1/2 + 2/3 x 2/3 x 1/2 = 4/9

Since we know that the child is not affected the new total is:
1 – p(child will be affected) = 1 - 1/9 = 8/9

Therefore, given that it is not affected, p(child is a carrier) is $\frac{4/9}{8/9}$ = **1/2**

(b) Assume that the trait is autosomal recessive with 80% penetrance. What is the probability that Ben and Jane's first child will be affected?

Since the trait is now autosomal recessive with 80% penetrance, we cannot longer assume that both Ben and Jane are Aa.
Jane and Ben can be:
- AA = 1/4
- Aa = 1/2
- aa not expressing the trait (represented as aa^{not}) = 1/4 x 20% = 1/20

Since we know that both parents do not express the trait in question, the new total is the sum of probabilities of the different genotypes Jane and Ben can be:
1/4 + 1/2 + 1/20 = 16/20

Therefore the probabilities of the different genoytpes that Jane and Ben can be (given that they do not express the trait) are:

$$p(Aa) = \frac{1/2}{16/20} = 5/8$$

$$p(aa^{not}) = \frac{1/20}{16/20} = 1/16$$

Therefore:
p(child is aa) = p(Jane is Aa & Ben is Aa) x p(child is aa) + p(Jane is aa^{not} & Ben is Aa) x   p(child is aa) + p(Jane is Aa & Ben is aa^{not}) x p(child is aa) + p(Jane & Ben are aa^{not}) x p(childis aa) =
(5/8 x 5/8 x 1/4 x 80%) + 2(1/16 x 5/8 x 1/2 x 80%) + (1/16 x 1/16 x 1 x 80%) = **9/80**

(c) Assume that the trait is X-linked recessive with complete penetrance. What is the probability that Ben and Jane's first child will be affected? What is the probability that the first child will be a carrier?
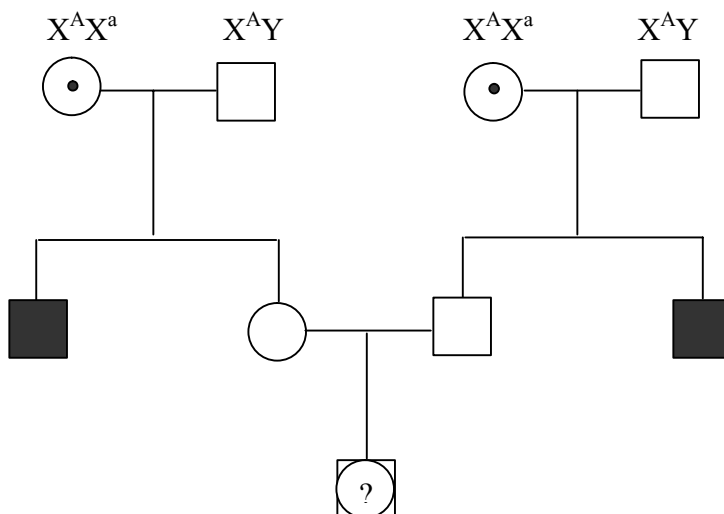
Since the trait is now X-linked recessive, males will be affected (X^aY) and females will most likely be carriers (X^AX^a). Therefore:

p(child is a carrier) = p(Jane is a carrier) x p(child is a carrier) = 1/2 x 1/4 = **1/8**

p(child is affected) = p(Jane is a carrier) x p(child is X^aY) = 1/2 x 1/4 = **1/8**

(d) (extra credit) Assume that the trait is autosomal dominant with 80% penetrance in heterozygotes and 100% penetrance in homozygotes. What is the probability that Ben and Jane's first child will be affected? If not affected, what is the probability that the child is a carrier?
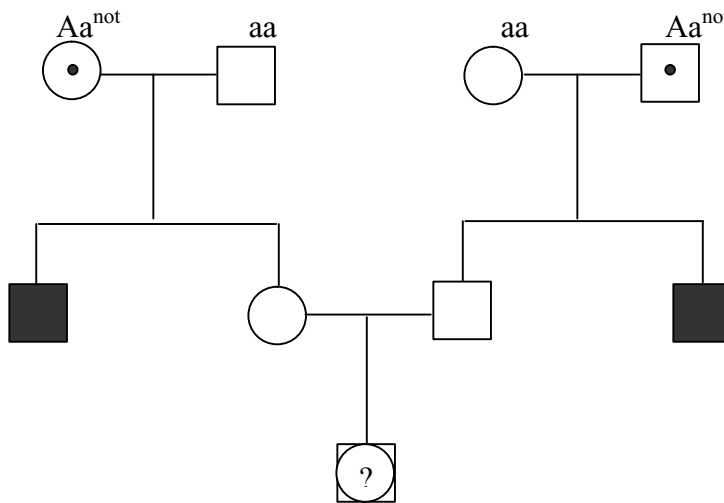
Since the trait is now autosomal dominant with 80% penetrance and is extremely rare, we assume that only one of the grandparents from each family is a carrier (heterozygote not expressing the trait) and that the other one is wild type. Therefore Ben and Jane could either be Aa^{not} (heterozygous for the trait but not affected), or wild type, denoted by aa (note that since the trait is now dominant, 'A' is being used to denote the trait in question and 'a' its wild type counterpart).

Jane and Ben can be:

p(Aa^{not}) = 1/2 x 20% = 1/10
p (aa) = 1/2

Since we know that both parents do not express the trait in question, the new total is the sum of probabilities of the different genotypes that Jane and Ben can be:

$1/10 + 1/2 = 6/10$

Therefore the probabilities of the different genotypes that Jane and Ben can be, given that they do not express the trait are:

$p(Aa^{not}) = \dfrac{1/10}{6/10} = 1/6$

$p(aa) = \dfrac{1/2}{6/10} = 5/6$

Therefore:

p(child will be affected) = p(Jane is Aa^{not} & Ben is Aa^{not}) x p(child is AA) + p(Jane is Aa^{not} & Ben is Aa^{not}) x p(child is Aa) + p(Jane is Aa^{not} & Ben is aa) x p(child is Aa) + p(Ben is Aa^{not} & Jane is aa) x p(child is Aa) = 1/6 x 1/6 x 1/4 x 100% + 1/6 x 1/6 x 1/2 x 80% + 2(1/6 x 5/6 x 1/2 x 80%) = **31/240**

For the next question let's calculate the probability that the child is a carrier first.

p(child is a carrier) = p(Jane is Aa^{not} & Ben is Aa^{not}) x p(child is Aa^{not}) + p(Jane is Aa^{not} & Ben is aa) x p(child is Aa^{not}) + p(Jane is aa & Ben is Aa^{not}) x p(child is Aa^{not}) = 1/6 x 1/6 x 1/2 x 20% + 2(1/6 x 5/6 x 1/2 x 20%) = 11/360

Since we know that the child is not affected the new total is:
1- 31/240 = 209/240

Therefore, if not affected, p(child is a carrier) is $\dfrac{11/360}{209/240}$ = **2/57**

7.03 Problem Set 2

Due before 5 PM on Thursday, September 26
Hand in answers in recitation section or in the box outside the class

1. You are interested in a new type of autosomal dominant mutation that gives purple eyes in *Drosophila*. You have a true-breeding purple-eyed line carrying a mutation that you call **pr-1**. You have isolated a second purple-eyed mutation that you call **pr-2**, which you have established in a true-breeding line that also carries a recessive marker **st**, which gives stubby legs. You cross a fly from the **pr-1** line to a fly from the **st**, **pr-2** line producing F_1 females which you then cross to males from a true breeding **st** line (normal eyes). From this cross, 1000 progeny flies can be categorized into four different phenotypic classes:

Phenotype	Number
Stubby legs, purple eyes	489
Normal legs, purple eyes	481
Stubby legs, normal eyes	3
Normal legs, normal eyes	27

(a) Normally, 3-factor crosses give eight different phenotypic classes, why does this particular cross give only four?

(b) What is the distance between **pr-1** and **pr-2** in cM?

(c) What is the phenotype of a **pr-1 pr-2** double mutant? Explain your reasoning.

(d) What is the relative order of **st**, **pr-1** and **pr-2**? Explain how you arrived at your answer.

(e) What is the distance between the **st** and **pr** genes? Again, explain how you arrived at your answer

2. In the yeast cross of ade2 met14 his3 x + + +, a total of 120 tetrads are analyzed. The following tetrad types are found in the numbers shown below each tetrad.   The markers ade2, met14, and his3 are abbreviated `a`, `m`, and `h` and the wild type allele of each gene is indicated by +.

(a)	(b)	(c)	(d)	(e)	(f)
a m h	+ m h	a m h	+ m h	a m h	+ m h
a m h	+ m h	+ + +	a + +	+ + +	a + +
+ + +	a + +	a + h	+ + h	+ + h	a + h
+ + +	a + +	+ m +	a m +	a m +	+ m +
17	1	47	3	5	2

(g)	(h)	(i)	(j)	(k)	(l)
a m h	a + h	a m h	+ m h	a + h	a m +
+ m h	+ + h	a + +	+ + +	a + h	a m +
a + +	a m +	+ + h	a + h	+ m +	+ + h
+ + +	+ m +	+ m +	a m +	+ m +	+ + h
3	4	12	9	16	1

a) Categorize each of the 12 different tetrad types by designating whether they are a parental ditype (P), tetratype (T), or nonparental ditype (N) for with respect to each of the three pairs of markers.

b) Use this categorization and the numbers of each type to determine which (if any) of the three markers are linked. For any linked markers, calculate the distance between them in cM.

To measure the distance between two linked His⁻ mutations, you cross a his2⁻ mutant to a his4⁻ mutant and then dissect 100 tetrads.   Your lab partner records which spore clones are His⁺ and which are His⁻ ; but then the fool loses the papers on which most of the data is recorded. In the end all he can remember is that there were three different tetrad types an that there was only one tetrad of the type with 2 His⁺ and 2 His⁻ spore clones.

c) What were the other two types of tetrads (how many His⁺ and His⁻ spores do they have)?

d) How many tetrads of each type do you estimate that there were? (Hint: use the assumption that crossovers are truly independent and probability of a single crossover in an interval is the square-root of the probability of a double crssover in the same interval)

e) Based on your numbers for part (**d**), calculate the distance between his2 and his4.

3. You have isolated two temperature-sensitive mutations in phage λ that you suspect may be in the same gene.   These phage mutants are called **ts-1** and **ts-2**. Each mutant will form plaques at 35˚C but not at 42˚C. You cross **ts-1** to **ts-2** phage by coinfecting *E. coli* at the permissive temperature of 35˚C.   When the resulting phage lysate is plated at 35˚C you count 10,000 plaques per ml of phage lysate, but when the same phage lysate is plated at 42˚C, there are only 80 plaques per ml.

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

(b) If the total size of the phage is 5 x 10^4 bp and the total genetic map length of the phage is 100 m.u., about how far apart are the mutations in base pairs?

You next cross a **ts-1** mutant to phage strain that carries **ts-2** in addition to a mutation called **sm**, which gives small plaques. When the resulting lysate is plated out at 42˚C and 100 plaques are examined, 25 are small and 75 are of normal size.

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

2002 Problem Set #2 Answer Key

1)

a) This three factor cross only gives us four different phenotypic classes because two of the mutations of interest give the same phenotype.

b) The distance between pr1 and pr2 can be calculated by recognizing that half of the crossovers between pr1 and pr2 will give rise to progeny with normal eyes and the other half will carry both the pr1 and pr2 alleles and thus be purple. Since a crossover event between pr1 and pr2 is the only way to get flies with normal eyes, we can state that the number of progeny that resulted from a crossover between pr1 & pr2 is twice the number of normal eyed progeny.

30 (total normal * **2** (we only observe half of the crossover = **60** (# of recombinants
eyed progeny)   classes as normal eyed progeny)   between pr1 & pr2)

Map distance   = $\underline{60 \text{ (# of recombinants between pr1 \& pr2)}}$ * 100 = **6cM**
from pr1 to pr2   **1000** (total # of progeny)

c) Pr1 and pr2 are mutant alleles that are dominant to the normal eyes allele and result in a purple-eyed phenotype. The pr1, pr2 double mutant could have a separate phenotype than the individual pr mutants, but in this case the only two phenotypes observed are normal and purple eyes. We know that the normal allele is recessive to either pr allele and that there is no normal allele present in a pr1, pr2 double mutant, so the only possible phenotype of a pr1, pr2 double mutant is **purple eyes**.

d) We know the parental classes are:

st   **+**   **pr2**   (stubby legs & purple eyes)

AND

pr   **+**   **+**   (normal legs & purple eyes)

because these are what the heterozygous mother's chromosomes looked like before any crossing over.

The least frequent class is stubby legs & normal eyes (only 3 progeny), which we assume is one of the double crossover classes. These individuals received neither mutant pr allele from their mother, but did receive her mutant st allele. Therefore the other double crossover event resulted in progeny with both mutant pr alleles and the wild-type st allele so they have normal legs & purple eyes.

From these observations we know that the double crossover classes are:

st   +   +   (stubby legs & normal eyes)

        AND

    +   **pr1**   **pr2**   (normal legs & purple eyes)

When we compare the parental classes to the double crossover classes, we see that the same st and pr1 alleles are together in both classes but with different pr2 alleles. Therefore, pr2 must be in between pr1 and st and the gene order is:

pr1_____**pr2**_____**st**

e) The distance between any one pr gene and the st gene can not be calculated directly since the two mutant pr genes give the same phenotype and one is each copy of the parental chromosome. Therefore, we must indirectly calculate this distance by taking advantage of the known distance between the two pr genes, the known number of double crossovers, and the relationship between single and double crossovers in this interval.

The probability that a double crossover with respect to the three genes of interest takes place in between pr1 and st is the product of the probability that a crossover occurs in between pr1 & pr2 *and* between pr & st.

     p(double x-over) = p(x-over between pr1 & pr2) * p(x-over between pr2 & st)

- We know the probability of a crossover between pr1 & pr2 is 60/1000 since we observe 30 progeny with normal eyes which is one of the two classes produced as a result of a pr1, pr2 crossover. (see 3b)

- We know that the probability of a double crossover is 6/1000 since we observe 3 progeny of one of the double crossover classes and assume that the other indistinguishable double crossover class has about the same number of progeny.

Therefore:

     (6/1000) = (60/1000) * p(x-over between pr2 & st)

     p(x-over between pr2 & st) = 0.1

     distance between pr2 & st = .01 * 100 = **10cM**

2.) a.) For each pair of markers, you can "cover" the column corresponding to the marker in which you are not interested with a pen and just classify each tetrad as P, N, T on the basis of the other two.

Tetrad	Number	a/m	a/h	m/h
A	17	P	P	P
B	1	N	N	P
C	47	T	P	T
D	3	T	N	T
E	5	P	T	T
F	2	N	T	T
G	3	T	T	P
H	4	T	T	N
I	12	T	T	T
J	9	T	T	T
K	16	N	P	N
L	1	P	N	N

b.) P >> NPD for two markers to be linked, if unlinked the genes would assort independently and result in a much closer number of apparent P and N type tetrads:

a/m: P = 17+5+1 = **23**
N = 1+2+16 = **19**
T = 47+3+3+4+12+9 = **78**
P = N
(a, m **unlinked**)

a/h: P = 17+47+16 = **80**
N = 1+3+1 = **5**
T = 5+2+3+4+12+9 = **35**
P > N
(a, h **linked**)

m/h: P = 17+1+3 = **21**
N = 4+16+1 = **21**
T = 47+3+5+2+12+9 = **78**
P = N
(m, h **unlinked**)

a-h map distance = 100 cM x (T + 6N)/2Σ = 100 cM x (35 + 6(5)) / 2(120) = **27 cM**

c.)

Tetrad Type	Non-Parental Ditype (N)	Tetratype (T)	Parental Ditype (P)
Phenotype	2His+:2His-	1 His+:3 His-	0 His+:4 His-
Genotypes	His2+ His4+	His2+ His4+	His2+ His4-
	His2+ His4+	His2+ His4-	His2+ His4-
	His2- His4-	His2- His4+	His2- His4+
	His2- His4-	His2- His4-	His2- His4+

d.) For every 1 N tetrad counted, we can assume there are 4 total double crossover tetrads (1 P, 1 N, 2 T), which means double crossover probability among these 100 tetrads is $4/100 = $ **0.04**. Single crossover tetrad probability must therefore be v0.04 = **0.2**, which means that there are 0.2(100) = 20 tetrads with single crossovers. Now here is where the solution encounters ambiguity and two possible solutions:

  i.   Assuming that single crossovers and double crossovers are exclusive events, this means we have 20 (single crossover) T tetrads + 2 (double crossover = 2NPD) T tetrads = **22 T**. We also know from the problem there is **1 N** tetrad so by process of elimination there are 100 – 22 – 1 = **77 P** tetrads.

  ii.   Assuming that double crossover tetrads are a subclass of single crossover tetrads that have just experienced one extra crossover, we have to subtract 4 (=4NPD) from the 20 single crossover tetrads but add back 2 (=2 NPD) double crossover (T) tetrads to get the total number of 20 – 4 +2 = **18 T** tetrads. We also know from the problem there is **1 N** tetrad so there are 100 – 18 – 1 = **81 P** tetrads.

  iii.   On a completely tangential note, this problem could have been definitively solved rather complicatedly with a Poisson distribution, but such was by no means expected. All three answers would have been accepted.

e.)

  i: his2-his4 map distance = 100cM x $(T+6N)/2\Sigma$ = 100cM x (22+6(1)/2(100)) = **14 cM**
     OR
 ii: distance = 100 cM x (18 + 6(1)/2(100)) = **12 cM**

3.) a.) Although there are only 80 observable plaque-forming units, you have to account for the equal number of reciprocal recombinants (**ts-1 ts-2** double mutants) that cannot be seen because they do not form plaques at the non-permissive temperature:

ts1-ts2 distance = 100 mu x #crossovers/#total = 100 m.u. x (**2** x 80)/10000 = **1.6 mu**

b.) $\dfrac{1.6 \text{ mu}}{100 \text{ mu}} = \dfrac{x \text{ bp}}{5 \times 10^4 \text{ bp}}$

   x = **800 bp**

c.) small, temperature-resistant plaques are the least frequent phenotypic class, hence it must be the result of double crossovers, and only the **sm-ts2-ts1** order (or its reverse, ts1-ts2-sm) can accommodate this:

```
        ----+----------------+-----------------ts1----

        ----sm---------------ts2----------------+----
```
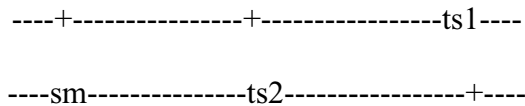
For the same reasons employed in problem 1e, looking at a singly crossed over class of known distance can give us the recombination frequency between the other two linked loci. Looking among temperature-resistant plaque forming units (wildtype at both the ts2 and ts1 loci), we count the frequency of double crossovers (to yield small plaques at 45ºC).

Sm-tm2 map distance = 100 mu x #small temp^r/#total = 100 mu x 25/100 = **25 mu**.

7.03 Problem Set 3

Due before 5 PM on Thursday, October 17
Hand in answers in recitation section or in the box outside the class

1. (a) During your study of the gene for an enzyme you have isolated an amber mutation in the middle of the gene which produces a truncated form of the enzyme. By placing a +1 frameshift mutation a short distance before the amber mutation and a –1 frameshift mutation a short distance after the amber mutation you create a triple mutant that restores expression of a functional, full-length enzyme. Explain how this could be.

(b) Working with the same gene as in part (a) you are studying intragenic suppressors of a +1 frameshift mutation near the middle of the gene. You find that a number of different –1 frameshift mutations a short distance *after* the +1 frameshift mutation will restore the expression of a functional enzyme. But to your surprise you find that –1 frameshift mutations placed a short distance *before* the +1 frameshift mutation lead to expression of a truncated enzyme. Propose a mechanism to explain these results.

(c) Organisms that live in environments exposed to light often have a relatively low proportion of A•T base pairs in their DNA (this is probably because neighboring thymine residues are particularly sensitive to reaction to UV light). The genes in such organisms tend to be biased for codons that contain G•C rather than A•T base pairs. As it turns out it is much easier to find coding sequences in organisms with a low A•T content in their DNA. To see why this is so, calculate the probability of finding an open reading frame of 50 codons by chance in DNA with equal frequencies of A•T and G•C base pairs and compare this to the frequency of finding an open reading frame of the same length in DNA that is made up of 40% A•T and 60% G•C.

(c) The three stop codons are 5'UAG3', 5'UAA3', and 5'UGA3'. You are trying to isolate tRNA mutations that will suppress and ochre (UAA) mutation. To increase the frequency of 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 gene can 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. You have isolated two *E. coli* mutants in the PyrF gene, called PyrF-1 and PyrF-2. These mutants require a pyrimidine (such as uracil) to be added to the medium in order to grow – this phenotype is referred to as Pyr⁻.

In order towork with these mutants, you wish to isolate a linked **Tn5** insertion. To do this you start with a collection of 1000 different random **Tn5** insertions in the otherwise wild type *E. coli* strain (these insertion strains are all Kan^r and Pyr$^+$). You grow **P1** phage on the entire mixed collection of **Tn5** insertion strains and then infect a PyrF-1 mutant and select for Kan^r transductants. Most of the Kan^r transductants remain Pyr⁻, but one out of 1000 is Pyr$^+$.

 (a) What do you know about the position of the **Tn5** insertion in this rare Pyr$^+$ transductant? Does this frequency with which it arose (1/1000) tell you anything more specific about its position?

Next you grow **P1** phage on the Pyr$^+$ transductant isolated above and then infect the Pyr⁻ mutant with the resulting phage. You select for Kan^r transductants and then test these transductants for the ability to grow without pyrimidine. You find that 30 out of 100 Kan^r transductants are Pyr$^+$.

(b) Give the distance between the **Tn5** insertion and PyrF-1, expressed as a cotransduction frequency.

The **Tn5** insertion shows about the same linkage to PyrF-2 as to PyrF-1. You construct a strain that has both the **Tn5** insertion and PyrF-1 and another strain that has both the **Tn5** insertion and PyrF-2. Using these strains you perform two reciprocal crosses. In the first cross, **P1** is grown on the **Tn5** PyrF-1 strain and the resulting phage are used to infect a PyrF-2 strain. In this transduction experiment, 7 out of 1000 Kan^r transductants are Pyr$^+$. In the reciprocal cross, **P1** is grown on the **Tn5** PyrF-2 strain and the resulting phage are used to infect a PyrF-1 strain. In this experiment, 50 out of 1000 Kan^r transductants are Pyr$^+$.

(c) Draw a map showing the relative order of the **Tn5** insertion, PyrF-1 and PyrF-2. Also include any relevant distances.

3. You are studying the polysaccharides displayed on the outer membrane of a bacterial species related to *E. coli*. Mutants that do not produce the normal polysaccharides can be identified because they have a dry-looking colony morphology.

(a) You use transposon mutagenesis to isolate two sets of mutants dry-looking colonies. Three mutants, numbered 1-3, were generated by random insertion of Tn5 (carrying a gene for Kan^r) into the bacterial genome , while three other mutants, numbered 4-6, were generated by random insertion of a second type of transposon Tn10 (carrying a different antibiotic resistance, Tet^r). You grow P1 phage on each of the three strains carrying a Tn5 insertion and then you use these phage for transduction of the Tn5 marker into each of the three Tn10 insertion mutants by selecting for Kan^r after infection of the recipient strain:

Donor strain	Recipient strain	% Kan^r Tet^s	% Kan^r Tet^r
1::Tn5	4::Tn10	100%	0%
1::Tn5	5::Tn10	10%	90%
1::Tn5	6::Tn10	24%	76%
2::Tn5	4::Tn10	2%	98%
2::Tn5	5::Tn10	100%	0%
2::Tn5	6::Tn10	100%	0%
3::Tn5	4::Tn10	15%	85%
3::Tn5	5::Tn10	100%	0%
3::Tn5	6::Tn10	100%	0%

How many linked groups of mutants are there? What are the distances between the linked insertions, expressed as cotransduction frequencies?

(b) As it turns out, the set of Tn5 insertions (strains 1-3) were generated in an Hfr strain. In mating experiments to a Str^r recipient, you find that although the 1::Tn5 Hfr strain transfers Kan^r late, even after a brief mating time it is possible to isolate a few Kan^r Str^r exconjugants. You choose two of these exconjugate strains for further study (for reference you call these strains a and b. The set of Tn10 insertions (strains 4-6) were generated in an F$^-$ strain. You mate strains a and b with strains 4-6 selecting for Kan^r and Tet^r.

Donor strain	Recipient strain	Kan^r Tet^r exconjugant morphology
a	4::Tn10	dry
a	5::Tn10	dry
a	6::Tn10	wildtype
b	4::Tn10	dry

| b | 5::Tn10 | wildtype |
| b | 6::Tn10 | wildtype |

Propose a mechanism to explain why the wild type colony morphology is restored in some of these matings and why the results for matings with strains a and b differ. Your answer should include diagrams of the form of F factor in strains a and b, showing any relevant genes and the position and orientation of the F factor origin of transfer.

(c) In studying one of the Tet^r Kan^r exconjugates isolated by the mating between strain b and 5::Tn10 isolated in part (b) you find a rare derivative that can transfer Tet^r (on Tn10) early and efficiently to a Str^r recipient but transfers Kan^r (on Tn5) very inefficiently. Draw a diagram showing how this unusual strain was formed.

2002 7.03 Problem Set 3 Answers

1. (a) During your study of the gene for an enzyme you have isolated an amber mutation in the middle of the gene which produces a truncated form of the enzyme. By placing a +1 frameshift mutation a short distance before the amber mutation and a –1 frameshift mutation a short distance after the amber mutation you create a triple mutant that restores expression of a functional, full-length enzyme. Explain how this could be.

The +1 frameshift puts the amber mutation out of frame so that it is no longer read as a stop. However this throws off the frame of the rest of the protein unless a -1 frameshift is inserted after the amber to restore the frame of the end of the protein. For example:

amber mutation: 5'- …ACG GCA U̲A̲G̲ AUG AUC… - 3'
                                                                                            stop

+1 frameshift mutation (nucleotide #4, A):   5'- …ACG **A**GC AUA GAU **G**AU C… - 3'

- 1 frameshift mutation (nucleotide #13, G):   5' - …ACG AGC AUA GAU AUC… - 3'

(b) Working with the same gene as in part (a) you are studying intragenic suppressors of a +1 frameshift mutation near the middle of the gene. You find that a number of different –1 frameshift mutations a short distance *after* the +1 frameshift mutation will restore the expression of a functional enzyme. But to your surprise you find that –1 frameshift mutations placed a short distance *before* the +1 frameshift mutation lead to expression of a truncated enzyme. Propose a mechanism to explain these results.

There must be a stop codon before the +1 frameshift that is normally not in frame and is thus not read, but is set into frame by a -1 frameshift. There is, however, no potential stop codon in the short distance after the +1 frameshift. For example:

wt   5' - …AAU AUA ACG AUC AAA TCG… - 3'

+1 frameshift (nucleotide 8, G)   5' - …AAU AUA **A**GC GAU CAA ATC G… - 3'

-1 frameshift after +1 (nucleotide 11, A)   5' - …AAU AUA AGC GUC AAA TCG… - 3'

-1 frameshift before +1 (nucleotide 2, A)   5' - …AUA U̲A̲A̲ GCG AUC AAA TCG… - 3'

(c) Organisms that live in environments exposed to light often have a relatively low proportion of A•T base pairs in their DNA (this is probably because neighboring thymine residues are particularly sensitive to reaction to UV light). The genes in such organisms tend to be biased for codons that contain G•C rather than A•T base pairs. As it turns out it is much easier to find coding sequences in organisms with a low A•T content in their DNA. To see why this is so, calculate the probability of finding an open reading frame of 50 codons by chance in DNA with equal frequencies of A•T and G•C base pairs and compare this to the frequency of finding an open reading frame of the same length in DNA that is made up of 40% A•T and 60% G•C.

- In an organism that is 25% A, 25% T, 25% G, and 25% C, the probability of getting any of the three possible stop codons is:

p(UAG) = 25% X 25% X 25% = 1/64

p(UAA) = 25% X 25% X 25% = 1/64

p(UGA) = 25% X 25% X 25% = 1/64

Therefore the probability of getting any non-stop codon is:

1- 3 x (1/64) = 61/64

and the probability of getting any non-stop codon 50 times in a row is:

$(61/64)^{50}$ = **9.1 %**

- In an organism that is 20% A, 20% T, 30% G, and 30% C, the probability of getting any of the three possible stop codons is:

p(UAG) = 20% X 20% X 30% = 3/250

p(UAA) = 20% X 20% X 20% = 1/125

p(UGA) = 20% X 30% X 20% = 3/250

Therefore the probability of getting any non-stop codon is:

1- [1/125 + 2 x (3/250)] = 121/125

and the probability of getting any non-stop codon 50 times in a row is:

$(121/125)^{50}$ = **19.7%**

(c) The three stop codons are 5'UAG3', 5'UAA3', and 5'UGA3'. You are trying to isolate tRNA mutations that will suppress and ochre (UAA) mutation. To increase the frequency of 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 gene can 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.

*For a mutant tRNA to recognize 5'-UAA-3', it must have a mutant anticodon 5'-UUA-3'. Therefore the mutant gene encoding the mutant tRNA must read 5'-TAA-3' on the strand that is **transcribed** (DNA uses Ts instead of Us). If the mutagen used to get this sequence causes transition mutations (a purine is changed for a purine or a pyrimidine is changed for a pyrimidine), then the original possibilities for the transcribed strand of the wildtype tRNA gene would be:*

1. 5'-**C**AA-3'
2. 5'-T**G**A-3'
3. 5'-TA**G**-3'

*(The altered base is shown in bold.)*
The first possibility would encode a wildtype tRNA that recognized the codon 5'-CAA-3', which is the tRNA for glutamine. The other two possibilities would encode wildtype tRNAs that recognized 5'-UGA-3' or 5'-UAG-3', which are both stop codons. There are no wildtype tRNAs that recognize stop codons.

Therefore:
For the wildtype tRNA:

DNA　　　　　　　　　5' -……………T T G…………..- 3'
　　　　　　　　　　　3' -**……………AA C…………..**- 5'
　　　　　　　　　　　(The transcribed strand is shown in bold)

tRNA　　　　　　3' Gln

GUU
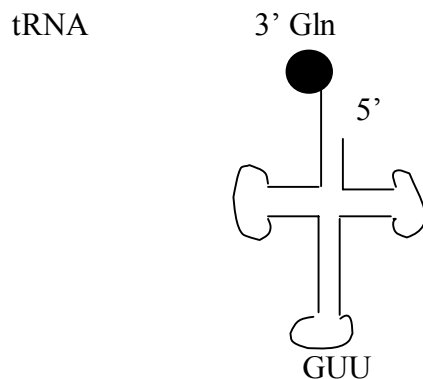
For the suppressor tRNA:

DNA　　　　　　　　　5' -……………T T A…………..- 3'
　　　　　　　　　　　3' -**……………AAT…………..**- 5'

tRNA   3' Gln

5'

AUU

2. You have isolated two *E. coli* mutants in the PyrF gene, called PyrF-1 and PyrF-2. These mutants require a pyrimidine (such as uracil) to be added to the medium in order to grow – this phenotype is referred to as Pyr⁻.

In order to work with these mutants, you wish to isolate a linked **Tn5** insertion. To do this you start with a collection of 1000 different random **Tn5** insertions in the otherwise wild type *E. coli* strain (these insertion strains are all Kan^r and Pyr$^+$). You grow **P1** phage on the entire mixed collection of **Tn5** insertion strains and then infect a PyrF-1 mutant and select for Kan^r transductants. Most of the Kan^r transductants remain Pyr⁻, but one out of 1000 is Pyr$^+$.

 (a) What do you know about the position of the **Tn5** insertion in this rare Pyr$^+$ transductant? Does this frequency with which it arose (1/1000) tell you anything more specific about its position?

This Tn5 insertion must be near to but not inside the PyrF gene and the distance between Tn5 and PyrF must be less than 10^5 bp, which is the size of the phage genome. The frequency does not tell you anything more specific than that.

Next you grow **P1** phage on the Pyr$^+$ transductant isolated above and then infect the Pyr⁻ mutant with the resulting phage. You select for Kan^r transductants and then test these transductants for the ability to grow without pyrimidine. You find that 30 out of 100 Kan^r transductants are Pyr$^+$.

(b) Give the distance between the **Tn5** insertion and PyrF-1, expressed as a cotransduction frequency.

*The Tn5 insertion was cotransduced with the wildtype copy of PyrF 30% of the time.*

The **Tn5** insertion shows about the same linkage to PyrF-2 as to PyrF-1. You construct a strain that has both the **Tn5** insertion and PyrF-1 and another strain that has both the **Tn5** insertion and PyrF-2. Using these strains you perform two reciprocal crosses. In the first cross, **P1** is grown on the **Tn5** PyrF-1 strain and the resulting phage are used to infect a PyrF-2 strain. In this transduction experiment, 7 out of 1000 Kan^r transductants are Pyr$^+$. In the reciprocal cross, **P1** is grown on the **Tn5** PyrF-2 strain and the resulting phage are used to infect a PyrF-1 strain. In this experiment, 50 out of 1000 Kan^r transductants are Pyr$^+$.

(c) Draw a map showing the relative order of the **Tn5** insertion, PyrF-1 and PyrF-2. Also include any relevant distances.
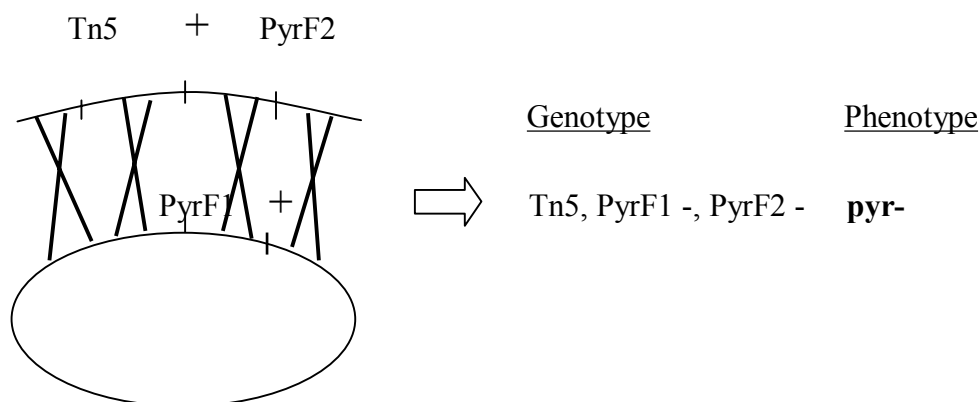
There are two possible orders: Tn5, PyrF1, PyrF2 OR Tn5, PyrF2, PyrF1. If we draw out the reciprocal crosses using both orders, we can determine what the rarest outcome would be given each order. The rarest outcome will always be the quadruple crossover, and the more common outcomes will be the double crossovers.
If the order is Tn5, PyrF1, PyrF2:
The quadruple crossover from the first cross would give:

Genotype	Phenotype
Tn5, PyrF1 +, PyrF2 +	**pyr+** (in this case 7/1000)
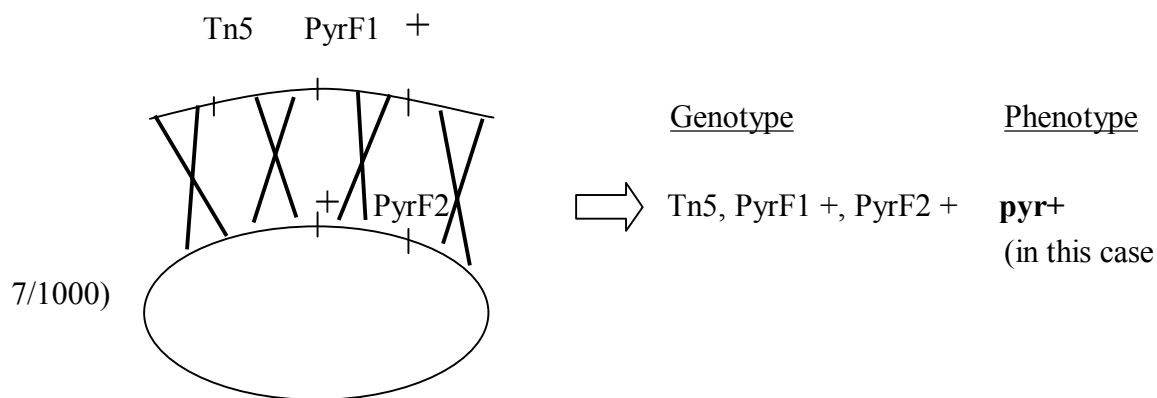
The quadruple crossover from the reciprocal cross would give:

Genotype	Phenotype
Tn5, PyrF1 -, PyrF2 -	**pyr-**

|

Both results are consistent with the data.

If the order is Tn5, PyrF2, PyrF1, *getting **pyr** + from the first cross would require a double crossover event and would occur (7/1000) times. Getting **pyr** + from the reciprocal cross would require a quadruple crossover and would occur (50/1000) times. It is not possible that a quadruple crossover event would occur more frequently than a double crossover event, and thus this order is not consistent with the data given.*

The order is therefore: **Tn5 PyrF1 PyrF2**

Once we know the order we can estimate the distance. We know that transducing Tn5 PyrF1+ PyrF2- *into a host that is* PyrF1 - PyrF2+ *gives us (50/1000) KanR pyr+ transductants. In order to get pyr+ in this cross, we must **not** get a crossover between Tn5 and PyrF1, and we must get a crossover between PyrF1 and PyrF2. The probability of getting pyr+, which is (50/1000), equals the product of the probabilities of these two events, which must happen in order to get pyr+. The probability of **not** getting a crossover between Tn5 and PyrF1 is just the cotransduction frequency, or 30%. Therefore:*
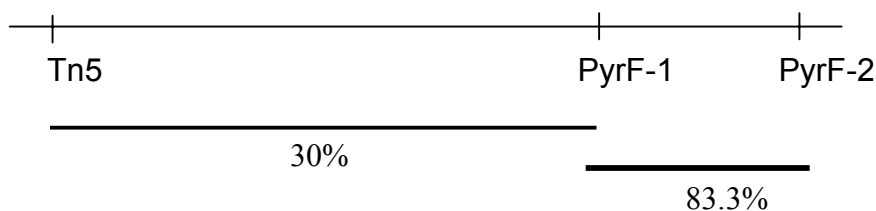
p(getting pyr+) = p(not getting crossover between Tn5 & PyrF1) x p(crossover between PyrF1 & PyrF2)

(50/1000) = (30/100) x p(crossover between PyrF1& PyrF2)

p(crossover between PyrF1 & PyrF2) = 16.7%

*Therefore PyrF1 and PyrF2 would be cotransduced **83.3%** of the time.*
Therefore:

3. You are studying the polysaccharides displayed on the outer membrane of a bacterial species related to *E. coli*. Mutants that do not produce the normal polysaccharides can be identified because they have a dry-looking colony morphology.

(a) You use transposon mutagenesis to isolate two sets of mutants dry-looking colonies. Three mutants, numbered 1-3, were generated by random insertion of Tn5 (carrying a gene for Kan^r) into the bacterial genome, while three other mutants, numbered 4-6, were generated by random insertion of a second type of transposon Tn10 (carrying a different antibiotic resistance, Tet^r). You grow P1 phage on each of the three strains carrying a Tn5 insertion and then you use these phage for transduction of the Tn5 marker into each of the three Tn10 insertion mutants by selecting for Kan^r after infection of the recipient strain:

Donor strain	Recipient strain	% Kan^r Tet^s	% Kan^r Tet^r
1::Tn5	4::Tn10	100%	0%
1::Tn5	5::Tn10	10%	90%
1::Tn5	6::Tn10	24%	76%
2::Tn5	4::Tn10	2%	98%
2::Tn5	5::Tn10	100%	0%
2::Tn5	6::Tn10	100%	0%
3::Tn5	4::Tn10	15%	85%
3::Tn5	5::Tn10	100%	0%
3::Tn5	6::Tn10	100%	0%

How many linked groups of mutants are there? What are the distances between the linked insertions, expressed as cotransduction frequencies?

If two mutations are grouped together, then they will be cotransduced 100% of the time and the resulting phenotype of the strains will be Kan^R Tet^S. If the mutations are not grouped together, then sometimes they will be cotransduced in which case the resulting strain will be $Kan^R Tet^S$ and sometimes they won't in which case the resulting strain will be $Kan^R Tet^R$.

Therefore 1::Tn5 and 4::Tn10 are in one group, and 2::Tn5, 3::Tn5, 5::Tn10, and 6::Tn10 are in another group. These insertions are grouped by position/distance only -- they cannot be considered complementation groups because we do not know the phenotypes of merodiploid bacterial strains containing pairs of insertions.

*The distance between 1::Tn5 and 5::Tn10 is **10%.***
*The distance between 1::Tn5 and 6::Tn10 is **24%.***
*The distance between 2::Tn5 and 4::Tn10 is **2%.***
*The distance between 3::Tn5 and 4::Tn10 is **15%.***

(b) As it turns out, the set of Tn5 insertions (strains 1-3) were generated in an Hfr strain. In mating experiments to a Str^r recipient, you find that although the 1::Tn5 Hfr strain transfers Kan^r late, even after a brief mating time it is possible to isolate a few Kan^r Str^r exconjugants. You choose two of these exconjugate strains for further study (for reference you call these strains a and b. The set of Tn10 insertions (strains 4-6) were generated in an F$^-$ strain. You mate strains a and b with strains 4-6 selecting for Kan^r and Tet^r.

Donor strain	Recipient strain	Kan^r Tet^r exconjugant morphology
a	4::Tn10	dry
a	5::Tn10	dry
a	6::Tn10	wildtype
b	4::Tn10	dry
b	5::Tn10	wildtype
b	6::Tn10	wildtype

Propose a mechanism to explain why the wild type colony morphology is restored in some of these matings and why the results for matings with strains a and b differ. Your answer should include diagrams of the form of F factor in strains a and b, showing any relevant genes and the position and orientation of the F factor origin of transfer.

*These rare exconjugates must have the F' form of the F factor, as they received the F factor from a strain that could transfer a marker early that is normally transferred very late. Donor **b** must have an F' that contains wildtype activity in the gene(s) in which the 5::Tn10 and 6::Tn10 insertions are inserted into, because this F' can complement the mutant phenotypes of both 5::Tn10 and 6::Tn10. The fact that the F' from donor **b** does not complement 4::Tn10 means that insertions 1 and 4 must be in the same gene. The F' from donor **a** must be wildtype for the gene in which 6::Tn10 is inserted because this F' can complement the mutant phenotype of 6::Tn10. Donor **a** does not carry a wildtype version of the gene in which 5::Tn10 is inserted or else it would complement. This is possible because from the cotransduction frequencies given above, we know that insertion 6 is closer to insertion 1 than insertion 5 is. This means that the*

F' could have recombined out of the genome and taken with it the gene in which 6 is inserted, but left the gene in which 5 is inserted behind, in the host chromosome.
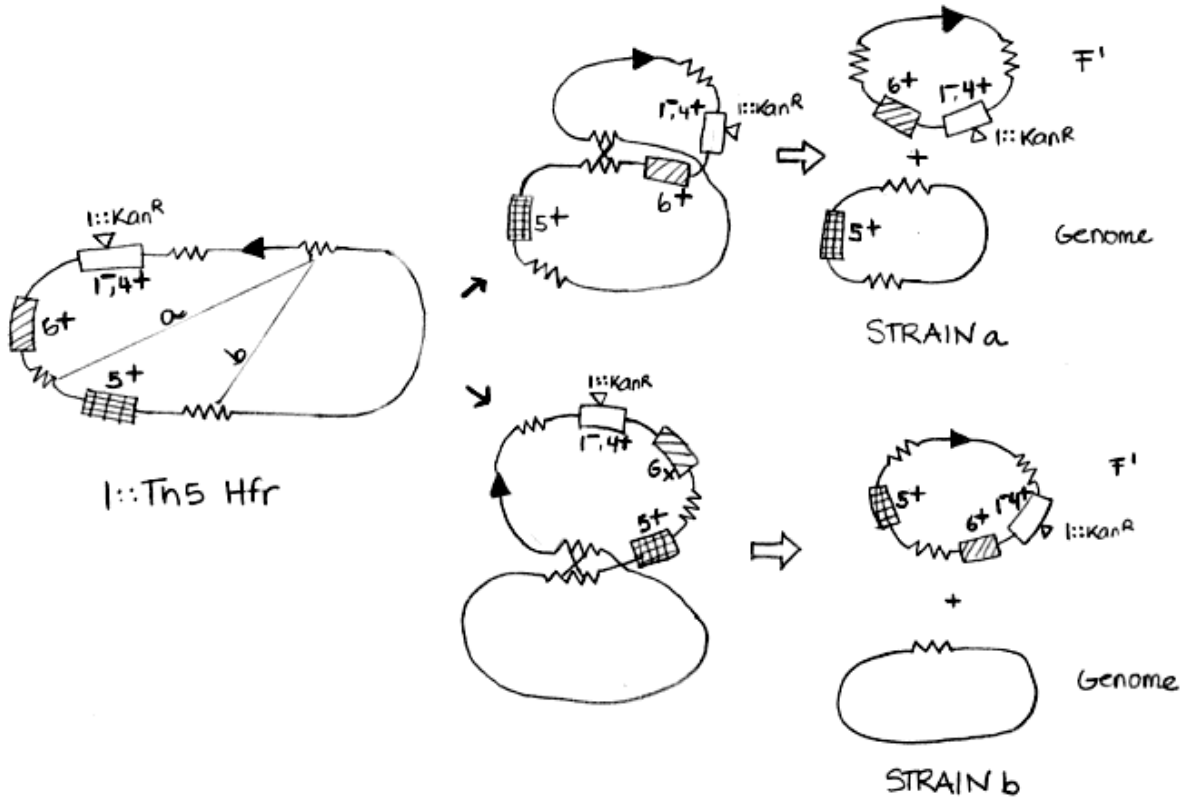
NOTE: *The IS sequences drawn are the only relevant IS sequences for this problem, although there are more IS sequences scattered throughout the genome.*

(c) In studying one of the Tet^r Kan^r exconjugates isolated by the mating between strain b and 5::Tn10 isolated in part (b) you find a rare derivative that can transfer Tet^r (on Tn10) early and efficiently to a Str^r recipient but transfers Kan^r (on Tn5) very inefficiently. Draw a diagram showing how this unusual strain was formed.

*The only way that the TetR gene could be transferred was if the F' from donor **b** integrated into the genome of a strain carrying 5::Tn10. The F' must recombine in a way so that the origin of transfer is in front of the TetR gene but behind the KanR gene, for only then would it transfer Tet resistance early and Kan resistance very late.*
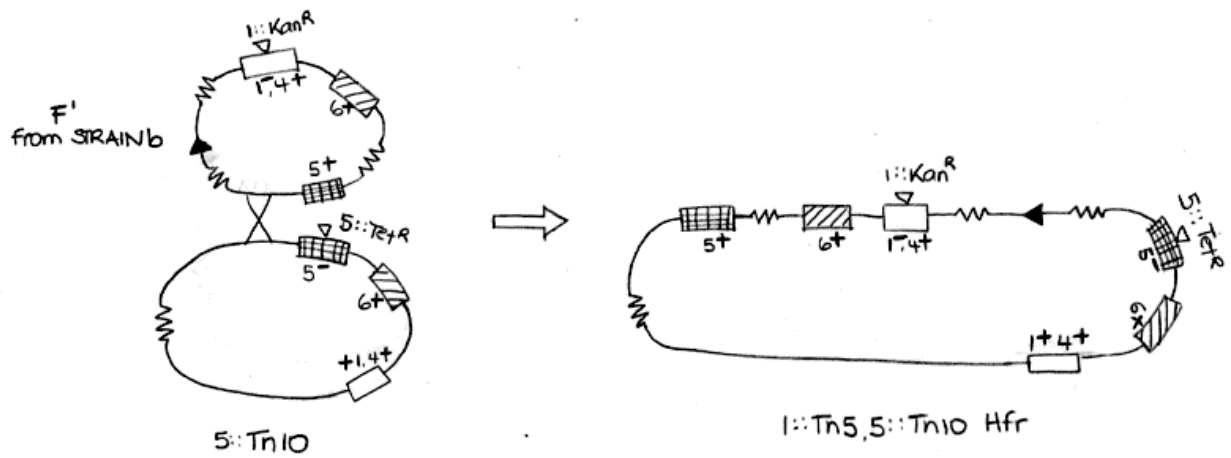
NOTE: *The F's will most likely integrate by homologous recombination between genes or intragenic regions, as these are large regions of homology between the F' and the host chromosome. The F' could integrate by homologous recombination between IS sequences on the F' and on the genome, but those sequences are very small. Therefore such an event is much rarer than recombination between genes or intragenic regions.*

7.03 Problem Set 4

Due before 5 PM on Thursday, October 24
Hand in answers in recitation section or in the box outside the class

1. You have isolated four new regulatory mutants for the Lac operon which you designate Lac1⁻, Lac2⁻, Lac3⁻, Lac4⁻, and Lac5⁻. Based on the results of the genetic tests outlined below deduce the nature of each mutation. Be as explicit as you can in your description of the type of mutation.

(a)

	ß-galactosidase activity	
	− IPTG	+ IPTG
Lac 1⁻ Lac Z⁻ / F' Lac ⁺	−	+
Lac 1⁻	+	+
Lac 1⁻ / F' LacZ⁻	+	+
Lac 1⁻ / F' Lac ⁺	+	+

(b)

	ß-galactosidase activity	
	− IPTG	+ IPTG
Lac 2⁻	−	−
Lac 2⁻ / F' Lac ⁺	−	−
Lac 2⁻ / F' LacZ⁻	−	−
Lac 2⁻ Lac Z⁻ / F' Lac ⁺	−	−

(c)

	ß-galactosidase activity	
	− IPTG	+ IPTG
Lac 3⁻	+	+
Lac 3⁻ / F' LacZ⁻	−	+
Lac 3⁻ / F' Lac ⁺	−	+
Lac 3⁻ Lac Z⁻ / F' Lac ⁺	−	+

(d)

	ß-galactosidase activity	
	− IPTG	+ IPTG
Lac 4⁻	+	+
Lac 4⁻ / F' Lac ⁺	+	+
Lac 4⁻ Lac Z⁻ / F' Lac ⁺	+	+
Lac 4⁻ / F' LacZ⁻	+	+

(e)

	ß-galactosidase activity	
	− IPTG	+ IPTG
Lac 5⁻	−	−
Lac 5⁻ / F' LacZ⁻	−	−
Lac 5⁻ Lac Z⁻ / F' Lac ⁺	−	+
Lac 5⁻ / F' Lac ⁺	−	+

2. You are studying the regulation of a new set of *E. coli* genes that are required to utilize lysine as a source of nitrogen. The first enzyme in this pathway is encoded by the gene LutA. Because this enzyme is difficult to assay you decide to study its regulation by constructing a fusion of the LutA promoter (with associated cis-acting regulatory elements) to the LacZ gene. This gene fusion (designated P_{LutA}–LacZ) is placed on a plasmid that can be transformed into *E. coli* cells of different genetic backgrounds. In wild type *E. coli* (with a deletion of the endogenous LacZ gene) the gene fusion shows the following regulation in response to lysine in the medium.

	ß-galactosidase activity	
	– lysine	+lysine
wild type (P_{LutA}–LacZ)	–	+

You have isolated two mutations that affect the regulation of P_{LutA}–LacZ, which you designate Lut1⁻ and Lut2⁻. Both mutations are located in the same region of the chromosome: you have an F' factor that carries this entire region, designated F' Lut⁺. Genetic tests with the mutants yield the following results:

	ß-galactosidase activity	
	– lysine	+lysine
Lut1⁻ (P_{LutA}–LacZ)	+	+
Lut2⁻ (P_{LutA}–LacZ)	–	–
Lut1⁻ (P_{LutA}–LacZ)/ F' Lut⁺	–	+
Lut2⁻ (P_{LutA}–LacZ)/ F' Lut⁺	–	+

(a) Classify the Lut1⁻ and Lut2⁻ mutations in terms of their basic genetic properties (constitutive/uninducible, dominant/ recessive, cis-/trans-acting). Explain the rationale behind your conclusions. Based on these properties make a proposal for the type of regulatory functions affected by the Lut1⁻ and Lut2⁻ mutations.

(b) Give two possible models for regulatory pathways for LutA that can explain the behavior of the Lut1⁻ and Lut2⁻ mutations. For each model (diagrams would be helpful), be as specific as you can about the normal function of the Lut1 and Lut2 gene products. Also be sure to include in your models a role for the inducer lysine.

(c) Next, you construct a Lut1⁻ Lut2⁻ double mutant, which gives the following behavior:

	ß-galactosidase activity	
	– lysine	+lysine
Lut1⁻ Lut2⁻ (P_{LutA}–LacZ)	+	+

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

(d) You isolate a third mutant, Lut3⁻ which has the following behavior:

	ß-galactosidase activity	
	− lysine	+lysine
Lut3⁻ (P_{LutA}–LacZ)	–	–
Lut3⁻ (P_{LutA}–LacZ)/ F' Lut⁺	–	–

As above, classify the Lut3⁻ mutation in terms of its basic genetic properties and explain how you arrived at your conclusions.

(e) P1 transduction crosses reveal that Lut3⁻ is very closely linked to Lut1⁻. Moreover the Lut1⁻ Lut3⁻ double mutant has the following behavior:

	ß-galactosidase activity	
	− lysine	+lysine
Lut1⁻ Lut3⁻ (P_{LutA}–LacZ)	+	+

On the basis of these results and the model that you developed above, propose a mechanism to explain the behavior of Lut3⁻. Your answer should include a diagram showing the entire pathway for LutA regulation indicating the function of each of the elements affected by the Lut1⁻ Lut2⁻ , and Lut3⁻ mutations and the inducer lysine.

7.03 Problem Set 4

Due before 5 PM on Thursday, October 24

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

1. You have isolated four new regulatory mutants for the Lac operon which you designate Lac1−, Lac2−, Lac3−, Lac4−, and Lac5−. Based on the results of the genetic tests outlined below deduce the nature of each mutation. Be as explicit as you can in your description of the type of mutation.

(a)

	ß-galactosidase activity		
	− IPTG	+ IPTG	
Lac 1− Lac Z− / F' Lac +	−	+	Lac1- can NOT act in trans
Lac 1−	+	+	Lac1- is constitutive
Lac 1− / F' LacZ−	+	+	Lac1- can act in cis
Lac 1− / F' Lac +	+	+	Lac1- is dominant

*Lac1- is a **dominant, constitutive, cis-acting** mutation. Based on these properties, Lac1- may be an O^C mutation, which is an operator constitutive mutation. Such a mutation makes the operator sequence unrecognizable to the LacI repressor and therefore the repressor never binds.*

(b)

	ß-galactosidase activity		
	− IPTG	+ IPTG	
Lac 2−	−	−	Lac2- is uninducible
Lac 2− / F' Lac +	−	−	Lac2- is dominant
Lac 2− / F' LacZ−	−	−	Lac2- can act in cis
Lac 2− Lac Z− / F' Lac +	−	−	Lac2- can act in trans

*Lac2- is a **dominant, uninducible, trans-acting** mutation. Based on these properties, Lac2- may be an I^S mutation, which is a superrepressor mutation. Such a mutation makes the LacI repressor insensitive to lactose and therefore the repressor is always in its DNA binding conformation and thus always represses.*

(c)

	ß-galactosidase activity		
	− IPTG	+ IPTG	
Lac 3−	+	+	Lac3- is constitutive
Lac 3− / F' LacZ−	−	+	Lac3+ can act in trans
Lac 3− / F' Lac +	−	+	Lac3- is recessive
Lac 3− Lac Z− / F' Lac +	−	+	Lac3+ can act in cis

*Lac3- is a **recessive, constitutive, trans-acting mutation**. Based on these properties, Lac3- may be an **I-** mutation, which is a mutation that causes no functional repressor to be made. Such a mutation makes no repressor and therefore repression does not occur.*

(d)

	ß-galactosidase activity		
	– IPTG	+ IPTG	
Lac 4⁻	+	+	Lac4- is constitutive
Lac 4⁻ / F' Lac ⁺	+	+	Lac4- is dominant
Lac 4⁻ Lac Z⁻ / F' Lac ⁺	+	+	Lac4- can act in trans
Lac 4⁻ / F' LacZ⁻	+	+	Lac4- can act in cis

*Lac4- is a **dominant, constitutive, trans-acting** mutation. Based on these properties, Lac4- may be an I^D mutation, which is a dominant negative mutation in the gene for the lac repressor. Such a mutation makes non-functional Lac repressor that also interferes with the function of any wildtype repressor protein that is present. Therefore the any repressor that is made cannot function to repress transcription.*

(e)

	ß-galactosidase activity		
	– IPTG	+ IPTG	
Lac 5⁻	–	–	Lac5- is uninducible
Lac 5⁻ / F' LacZ⁻	–	–	Lac5+ CAN'T act in trans
Lac 5⁻ Lac Z⁻ / F' Lac ⁺	–	+	Lac5+ can act in cis
Lac 5⁻ / F' Lac ⁺	–	+	Lac5- is recessive

*Lac5- is a **recessive, uninducible, cis-acting** mutation. Based on these properties, Lac5- may be an **P-** mutation, which is a mutation in the promoter that makes it unrecognizable to RNA polymerase. If RNA polymerase cannot recognize the promoter, transcription of the operon will not occur.*

2. You are studying the regulation of a new set of *E. coli* genes that are required to utilize lysine as a source of nitrogen. The first enzyme in this pathway is encoded by the gene LutA. Because this enzyme is difficult to assay you decide to study its regulation by constructing a fusion of the LutA promoter (with associated cis-acting regulatory elements) to the LacZ gene. This gene fusion (designated P_{LutA}–LacZ) is placed on a plasmid that can be transformed into *E. coli* cells of different genetic backgrounds. In wild type *E. coli* (with a deletion of the endogenous LacZ gene) the gene fusion shows the following regulation in response to lysine in the medium.

	ß-galactosidase activity	
	– lysine	+lysine
wild type (P_{LutA}–LacZ)	–	+

You have isolated two mutations that affect the regulation of P_{LutA}–LacZ, which you designate Lut1⁻ and Lut2⁻. Both mutations are located in the same region of the chromosome: you have an F' factor that carries this entire region, designated F' Lut⁺. Genetic tests with the mutants yield the following results:

	ß-galactosidase activity	
	– lysine	+lysine
Lut1⁻ (P_{LutA}–LacZ)	+	+
Lut2⁻ (P_{LutA}–LacZ)	–	–
Lut1⁻ (P_{LutA}–LacZ)/ F' Lut⁺	–	+
Lut2⁻ (P_{LutA}–LacZ)/ F' Lut⁺	–	+

(a) Classify the Lut1⁻ and Lut2⁻ mutations in terms of their basic genetic properties (constitutive/uninducible, dominant/ recessive, cis-/trans-acting). Explain the rationale behind your conclusions. Based on these properties make a proposal for the type of regulatory functions affected by the Lut1⁻ and Lut2⁻ mutations.

*Lut1- is **constitutive** because the presence of this mutation leads to expression of β-galactosidase activity both in the presence and absence of lysine.*
*Lut1- is **recessive** because when this mutation is present along with a wildtype copy of Lut1 on an F', inducible (wildtype) expression of β-galactosidase expression is restored.*
*Lut1- is **trans-acting** because the dominant allele of Lut1 (Lut1+) can act upon a LutA-lacZ fusion that is on a plasmid separate from the F' plasmid, which is where the Lut1+ allele is situated.*
*Based on these properties, Lut1 is a **repressor** of LutA expression that is not functional if the cell possesses a copy of the Lut1- mutation. Lut1 does not have to be a direct repressor of LutA.*

*Lut2- is **uninducible** because the presence of this mutation leads to a lack of expression of β-galactosidase activity both in the presence and absence of lysine.*
*Lut2- is **recessive** because when this mutation is present along with a wildtype copy of Lut2 on an F', inducible (wildtype) expression of β-galactosidase expression is restored.*
*Lut2- is **trans-acting** because the dominant allele of Lut2 (Lut2+) can act upon a LutA-lacZ fusion that is on a plasmid separate from the F' plasmid, which is where the Lut2+ allele is situated.*
*Based on these properties, Lut2 is an **activator** of LutA expression that is not functional if the cell possesses a copy of the Lut2- mutation. Lut2 does not have to be a direct activator of LutA.*

(b) Give two possible models for regulatory pathways for LutA that can explain the behavior of the Lut1⁻ and Lut2⁻ mutations. For each model (diagrams would be

helpful), be as specific as you can about the normal function of the Lut1 and Lut2 gene products. Also be sure to include in your models a role for the inducer lysine.

Two possible models are:
Model 1. *When lysine is present, the Lut1 repressor is not able to repress Lut2. Therefore, functional Lut2 activator is present. Lut2 then activates the expression of LutA. When lysine is absent, the Lut1 repressor is able to repress Lut2. No functional Lut2 activator is present, so LutA expression cannot be activated.*

lysine —⊣ Lut1 —⊣ Lut2 → LutA expression

Model 2. *When lysine is present, the Lut2 repressor is able to repress Lut1. Thus Lut1 cannot function as a repressor, and so LutA is actively transcribed. When lysine is absent, the Lut2 repressor ceases to be able to repress Lut1. Thus functional Lut1 repressor is present. Now that there is functional repressor, Lut1 is able to repress the expression of LutA.*

lysine → Lut2 —⊣ Lut1 —⊣ LutA expression

(Notably, a third model exists in which Lut1 and Lut2 act in parallel.)

(c) Next, you construct a Lut1⁻ Lut2⁻ double mutant, which gives the following behavior:

	ß-galactosidase activity	
	– lysine	+lysine
Lut1⁻ Lut2⁻ (P_{LutA}–LacZ)	+	+

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

From Models 1 and 2 above we can make predictions about the phenotype of a double Lut1-Lut2- mutation:
MODEL 1. No functional Lut1 repressor would be present so nothing could repress the Lut2 gene. But that doesn't really matter because functional Lut2 activator is not present either, and so LutA expression can never be activated; LutA is ALWAYS OFF.
MODEL 2. No fuctional Lut2 repressor would be present so nothing can repress Lut1. But that doesn't really matter because functional Lut1 repressor is not present either, and thus LutA expression can never be repressed. Therefore LutA is ALWAYS ON.
From these predictions, we see that the given phenotype of the double mutant is **consistent with Model 2.**

(Notably, our third model from above, in which Lut1 and Lut2 act in parallel, would lead to a double Lut1- Lut2- mutant with an uninducible phenotype. Thus this third model is not consistent with the data in part c.)

(d) You isolate a third mutant, Lut3⁻ which has the following behavior:

	ß-galactosidase activity	
	– lysine	+lysine
Lut3⁻ (P$_{LutA}$–LacZ)	–	–
Lut3⁻ (P$_{LutA}$–LacZ)/ F' Lut⁺	–	–

As above, classify the Lut3⁻ mutation in terms of its basic genetic properties and explain how you arrived at your conclusions.

__Lut3- is uninducible__ because the presence of this mutation leads to the lack of expression of β-galactosidase activity both in the presence and absence of lysine.
Lut3- is __dominant__ because, when this mutation is present along with a wildtype copy of Lut3 on an F', inducible (wildtype) expression of β-galactosidase expression is NOT restored.
Lut3- is __trans-acting__ because the dominant allele of Lut3 (Lut3-) can act upon a LutA-lacZ fusion that is on a plasmid, and therefore separate from the chromosome where the Lut3- allele is situated.
Based on these properties, Lut3 could either be a repressor of LutA (in which case Lut3- would be a super-repressor of LutA) or Lut 3 could be an activator of LutA (in which case Lut3- would be a dominant negative activator mutation).

(e) P1 transduction crosses reveal that Lut3⁻ is very closely linked to Lut1⁻. Moreover the Lut1⁻ Lut3⁻ double mutant has the following behavior:

	ß-galactosidase activity	
	– lysine	+lysine
Lut1⁻ Lut3⁻ (P$_{LutA}$–LacZ)	+	+

On the basis of these results and the model that you developed above, propose a mechanism to explain the behavior of Lut3⁻. Your answer should include a diagram showing the entire pathway for LutA regulation indicating the function of each of the elements affected by the Lut1⁻ Lut2⁻ , and Lut3⁻ mutations and the inducer lysine.

There are two possibilities for where the Lut3- mutation could be located. Lut3- could be an allele of the Lut1 gene, which is dealt with in possibility 1 below. Lut3 could also be in a separate gene, the Lut3 gene, that is closely linked to the Lut1 gene. This situation is dealt with in possibilities 2 and 3 below. If Lut3 is a separate gene, Lut3- could either be a super-repressor mutation (possibility 2) or a dominant negative activator mutation (possibility 3).

Possibility 1.
One could propose that Lut3- is another allele of the Lut1 gene. This allele is a dominant uninducible allele of the gene that produces Lut1 repressor, and will therefore produce a super-repressor that represses LutA regardless of the presence or absence of lysine. [From now on the Lut1 gene will be referred to as the Lut1(Lut3) gene.] This special Lut3- allele of the Lut1(Lut3) gene could produce Lut1(Lut3) repressor that is insensitive to repression by the Lut2 repressor.
If you combine that Lut3- mutation with the Lut1- mutation, you would get a gene that encodes this special super-repressor, but no functional Lut1(Lut3) repressor would be present anyway because of the Lut1- mutation (which disallows production of functional repressor). Thus the phenotype of the Lut1-Lut3- double mutation is constitutive expression of β-galactosidase activity due to the absence of functional Lut1(Lut3) repressor.
Revised Model 2: In the presence of lysine, the Lut2 repressor is capable of blocking the function of the Lut1(Lut3) repressor. Thus no functional Lut1(Lut3) repressor is present, and LutA cannot be repressed, so LutA expression is on. In the absence of lysine, the Lut2 repressor cannot block the function of the Lut1(Lut3) repressor protein, and this allows the Lut1(Lut3) repressor protein to function. When this repressor functions, it represses the LutA gene and thus expression of LutA is off.

lysine → Lut2 ⊣ Lut1(Lut3) ⊣ LutA expression

Possibility 2.

One could propose that Lut3- is a super-repressor allele with respect to LutA. However, the double mutant Lut3- Lut1- tells us that Lut3 acts before Lut1 in the pathway. Thus Lut3 would be an activator of Lut1, which would repress LutA. In this case, Lut3- is a net super-repressor of LutA, because it is a super-activator of Lut1, which is a repressor of LutA. This special Lut3- allele could produce Lut3 activator that is insensitive to repression by the Lut2 repressor.

Revised Model 2: In the presence of lysine, the Lut2 repressor is capable of blocking the function of the Lut3 activator. Thus no functional Lut3 activator is present, and Lut1 cannot be activated. Thus Lut1 cannot repress the transcription of LutA, so LutA expression is on. In the absence of lysine, the Lut2 repressor can no longer block the function of the Lut3 activator protein, and this allows the Lut1 repressor to be activated. When this Lut1 repressor functions, it represses the LutA gene and thus expression of LutA is off.

lysine → Lut2 ⊣ Lut3 → Lut1 ⊣ LutA expression

[Please note that the phenotype of a Lut2- Lut3- double mutant is not given and therefore Lut2 and Lut3 cannot be ordered with respect to each other. For this reason, the following model is equally appropriate:

lysine ⊣ Lut3 ⊣ Lut2 ⊣ Lut1 ⊣ LutA expression.*]*

Possibility 3.

One could propose that Lut3- is a dominant negative activator with respect to LutA. However, the double mutant Lut3- Lut1- tells us that Lut3 acts before Lut1 in the pathway. Thus Lut3 would be an repressor of Lut1, which would repress LutA. In this case, Lut3- is a net dominant negative activator with respect to LutA, because it is a dominant negative repressor of Lut1, which is a repressor of LutA. This special Lut3- allele could produce Lut3 repressor that is not only non-functional itself, but also interferes with any wildtype Lut3 repressor that is around.

Revised Model 2: In the presence of lysine, the Lut2 activator is capable of promoting the ability of the Lut3 repressor to repress. Thus no functional Lut1 repressor would be present, so LutA expression is on. In the absence of lysine, the Lut2 activator ceases to activate the Lut3 repressor protein and this allows the Lut1 repressor to function. When this Lut1 repressor functions, it represses the LutA gene and thus expression of LutA is off.

lysine → Lut2 → Lut3 ⊣ Lut1 ⊣ LutA expression

[Please note that the phenotype of a Lut2- Lut3- double mutant is not given and therefore Lut2 and Lut3 cannot be ordered with respect to each other. For this reason, the following model is equally appropriate:

lysine → Lut3 → Lut2 ⊣ Lut1 ⊣ LutA expression.*]*

7.03 Problem Set 5

Due before 5 PM on Thursday, November 14
Hand in answers in recitation section or in the box outside the class

1. You have discovered a gene in yeast that is involved in DNA repair. Mutations in this gene are more sensitive to DNA-damaging agents such as UV radiation, a property shared with many DNA repair genes, and you designate your new gene Rad66. To study the regulation of Rad66 you fuse the promoter region of the enzyme to the LacZ coding sequence and then place this hybrid gene (designated P_{rad66}–LacZ) on a yeast plasmid. As hoped, yeast cells carrying P_{rad66}–LacZ, do not express ß-galactosidase activity unless exposed to UV light, showing that the hybrid gene includes the crucial cis-acting regulatory sequences.

(a) You next identify a mutant that you call Reg1⁻, which gives expression of the P_{rad66}–LacZ reporter regardless of whether the cells have been exposed to UV light. By mating a Reg1⁻ strain to wild type you find that the resulting heterozygote only expresses the reporter in the presence of UV light. Categorize the Reg1- mutation as follows: dominant vs. recessive, constitutive vs. uninducible, and cis-acting vs. trans-acting. Propose a function for the normal Reg1⁻ gene product in the regulation of Rad66. Draw a model of Rad66 gene regulation that includes Reg1 and UV radiation.

(b) Next, you isolate a mutant that you call Reg2⁻, which will not express the P_{rad66}–LacZ reporter even after cells have been exposed to UV radiation. Mating a Reg2⁻ strain to wild type gives a heterozygote that can expresses the reporter in the presence of UV radiation. Given all of the available data on the Reg1⁻ and Reg2⁻ mutants, diagram two different linear pathways that can explain the roles of the Reg1, Reg2 gene products and UV radiation in the regulation of Rad66.

(c) You cross a strain that carries Reg1⁻ to a strain that carries Reg2⁻. You analyze only two tetrads and they both show the same pattern of expression of P_{rad66}–LacZ in response to UV light: one spore shows normal induction in response to UV radiation, two spores express the reporter even in the absence of UV radiation, and one does not express the reporter. What is the phenotype of a Reg1⁻ Reg2⁻ double mutant? Explain your reasoning and select the model from part (b) that is consistent with this result.

(d) Next you evaluate the upstream regulatory sequences necessary for expression of P_{rad66}–LacZ. The figure below shows the effect of different 50 bp deletions in the upstream regulatory region on the amount of ß-galactosidase activity expressed by the reporter gene.

```
        -300  -250  -200  -150  -100   -50      +1
          |     |     |     |     |     |       |        ß-galactosidase
                                                          −UV   +UV
wt  ———————————————————————————————————— LacZ     −      +
1   —————         ——————————————————————— LacZ     −      +
2   —————————         ———————————————————— LacZ     +      +
3   ————————————         ——————————————————LacZ     −      +
4   ———————————————————      ———————————————LacZ     −      +
5   ——————————————————————————      ————————LacZ     −      +
6   —————————————————————————————————       ——LacZ   −      −
```

Deletion 2 gives constitutive expression of P_{rad66}–LacZ whereas deletion 6 gives uninducible expression. In light of the experiments with trans-acting factors, propose functions for the two upstream regulatory segments defined by these deletions. What can you conclude from the fact that deletions 3-5 show normal regulation of P_{rad66}–LacZ?

2. You are studying an autosomal recessive human mutation that causes blindness. Through genetic linkage analysis, you map the mutation to a 0.6-Mb region of chromosome 12. Unfortunately, the genomic sequence of this region is not available, and the responsible gene has not yet been defined at a molecular level. However, you have 6 BACs (of unknown order) and 10 STSs (of unknown order) that derive from this region of chromosome 12. By PCR, you test each of the 10 BACs for the presence (+) or absence (-) of each of the 10 STSs, and you obtain the following results:

		STSs									
		1	2	3	4	5	6	7	8	9	10
	A	+	+	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	+	+	-	-	-
BACs	C	-	-	+	-	+	-	-	-	-	+
	D	+	-	-	-	-	-	-	+	-	-
	E	-	-	-	+	-	-	+	-	-	-
	F	+	-	+	-	+	-	-	+	-	-

a) Construct a physical map of this region that shows the order of the BACs, the overlaps among the BACs, the order of the STSs, and the placement of the STSs within the BACs. Are there any ambiguities within your map? If so, describe them.

b) You obtain the complete sequence of one mRNA that derives from the 0.6-Mb region. Design an experiment (without sequencing any of the BACs) that will show which of the BACs contain the 5' and 3' ends of the gene from which this mRNA is derived.

c) Several months after constructing the physical map, you obtain the full genomic sequence of the 0.6-Mb region. Using electronic tools, you identify three predicted genes, which are 2 kb, 15 kb, and 100 kb in length. The mRNA from part (b) is ~1900 bp long. Based only on the information that you have been given, can you tell which of the three predicted genes is likely (or unlikely) to be the source of the ~1900-bp mRNA? Briefly explain your answer.

d) A few weeks later, you obtain DNA samples from 10 patients who are blind because of the mutation mapping to this region. You also obtain DNA samples from 10 individuals with normal vision. How would you determine which one of the three predicted genes is responsible for the blindness that maps to this region?

3. As we will study later in the semester, there are genes in the human and mouse genomes that control cell proliferation. Mutations in such genes can result in uncontrolled proliferation and, as a result, cancer. For some such "cancer genes," a single mutant allele is sufficient to cause a diploid cell to proliferate wildly and cause cancer; such mutant alleles act **dominantly** to promote proliferation.   For other "cancer genes," mutant alleles act **recessively** to promote proliferation; in these case, both copies of the gene (both alleles) must be defective to make the cell cancerous.

We want to make a mouse model of cancer where the defect is caused by a **dominantly** acting mutation in a gene called RAS. This mutation results in uncontrolled proliferation and thus a predisposition to cancer even in the presence of wild-type alleles of the RAS gene.

a) Would it be possible to generate a suitable mouse model using pronuclear injection (transgenes)? Briefly explain why or why not.

b) Would it be possible to generate a suitable mouse model using gene-targeting techniques? Briefly explain why or why not.

c) If a suitable mouse model could be generated using either of the two techniques, which one would you choose? Briefly explain your reasoning.

d) Draw a labeled diagram of the transgene or targeting construct that you would use.

e) Once the transgene or targeting construct has been properly integrated into the mouse genome, are additional breeding experiments required before you've generated a mouse with a predisposition to cancer? Briefly explain why or why not.

You now want to make want to make a mouse model of cancer where the defect is caused by a **recessively** acting mutation in a gene called RB. This gene's normal function is to keep cells from proliferating out of control and one wild-type copy of the RB gene is sufficient to do so. However, if both RB alleles are defective, the cell will likely become cancerous as a result of uncontrolled proliferation.

f) Would it be possible to generate a suitable mouse model using pronuclear injection (transgenes)? Briefly explain why or why not.

g) Would it be possible to generate a suitable mouse model using gene-targeting techniques? Briefly explain why or why not.

h) If a suitable mouse model could be generated using either of the two techniques, which one would you choose? Briefly explain your reasoning.

i) Draw a labeled diagram of the transgene or targeting construct that you would use.

j) Once the transgene or targeting construct has been properly integrated into the mouse genome, are additional breeding experiments required before you've generated a mouse with a predisposition to cancer? Briefly explain why or why not.

Problem Set #5 Answer Key

1. You have discovered a gene in yeast that is involved in DNA repair. Mutations in this gene are more sensitive to DNA-damaging agents such as UV radiation, a property shared with many DNA repair genes, and you designate your new gene Rad66. To study the regulation of Rad66 you fuse the promoter region of the enzyme to the LacZ coding sequence and then place this hybrid gene (designated P_{rad66}–LacZ) on a yeast plasmid. As hoped, yeast cells carrying P_{rad66}–LacZ, do not express ß-galactosidase activity unless exposed to UV light, showing that the hybrid gene includes the crucial cis-acting regulatory sequences.

(a) You next identify a mutant that you call Reg1⁻, which gives expression of the P_{rad66}–LacZ reporter regardless of whether the cells have been exposed to UV light. By mating a Reg1⁻ strain to wild type you find that the resulting heterozygote only expresses the reporter in the presence of UV light. Categorize the Reg1- mutation as follows: dominant vs. recessive, constitutive vs. uninducible, and cis-acting vs. trans-acting. Propose a function for the normal Reg1⁻ gene product in the regulation of Rad66. Draw a model of Rad66 gene regulation that includes Reg1 and UV radiation.

Reg1⁻ is a **constitutive** (because it gives expression of the Prad-LacZ reporter regardless of whether the cells have been exposed to UV light), **recessive** (because in a diploid containing a wild-type copy of Reg1⁻ there is wild-type expression) and **trans-acting** (because the only Prad-LacZ fusion is on an exogenous piece of DNA tat is separate from the yeast chromosomes, where the Reg1⁻ mutation is located) mutation. Therefore, it must be a **net repressor of transcription of Rad66 transcription when UV light is not present**.

UV -----| Reg1 -----| Rad66

(b) Next, you isolate a mutant that you call Reg2⁻, which will not express the P_{rad66}–LacZ reporter even after cells have been exposed to UV radiation. Mating a Reg2⁻ strain to wild type gives a heterozygote that can expresses the reporter in the presence of UV radiation. Given all of the available data on the Reg1⁻ and Reg2⁻ mutants, diagram two different linear pathways that can explain the roles of the Reg1, Reg2 gene products and UV radiation in the regulation of Rad66.

Reg2⁻ is an uninducible (because there is no expression of the Prad-LacZ reporter regardless of whether the cells have been exposed to UV light), recessive and trans-acting (for the same reasons as in part a). Therefore it must be a net activator of Rad66 transcription when UV light is present. Since we also

know Reg1⁻ is a net repressor of transcription from part a, the two possible *linear* pathways are:

UV -----| Reg1 -----| Reg2 ---→ Rad66

AND

UV ---→ Reg2 -----| Reg1 -----| Rad66

(c) You cross a strain that carries Reg1⁻ to a strain that carries Reg2⁻. You analyze only two tetrads and they both show the same pattern of expression of P_{rad66}–LacZ in response to UV light: one spore shows normal induction in response to UV radiation, two spores express the reporter even in the absence of UV radiation, and one does not express the reporter. What is the phenotype of a Reg1⁻ Reg2⁻ double mutant? Explain your reasoning and select the model from part (b) that is consistent with this result.

A Reg1⁻ ,Reg2⁺ strain was mated to a Reg1⁺ ,Reg2⁻ strain and the resulting diploid was doubly heterozygous at the two loci (Reg1⁺ / Reg1⁻ ,Reg2⁻ / Reg2⁺). Since the tetrads observed have spores with three different phenotypes we know that we are looking at *tetratype tetrads* from the Reg1⁻ & Reg2⁻ cross. We know this because either parental ditype or non-parental ditype tetrads would give only two different phenotypes. From the data we know that two of our spores have constitutive expression (they express the reporter even in the absence of UV radiation), one has uninducible expression, and one has wildtype expression. The genotypes of the four tetratype spores are:

1. Reg1⁻ ,Reg2⁺ which we know results in constitutive expression (expression even in the absence of UV radiation)
2. Reg1⁺ ,Reg2⁻ which we know results in uninducible expression (no expression of the reporter)
3. Reg1⁺ ,Reg2⁺ which gives wild-type expression (normal induction in response to UV radiation)
4. Reg1⁻ ,Reg2⁻ - we don't know the phenotype of the double mutant from previous experiments, however since there are **two spores in the tetratype tetrad that express the reporter even in the absence of UV radiation (constitutive) and only one has been accounted for in the single mutants**, the phenotype of the double mutant must be **constitutive**.

Since the double mutant gives constitutive expression like the Reg1⁻ single mutant, we know that Reg1⁻ is epistatic to Reg2⁻. Therefore Reg1⁻ must be acting downstream of Reg2⁻ and the correct model is:

UV ---→ Reg2 -----| Reg1 -----| Rad66

(d) Next you evaluate the upstream regulatory sequences necessary for expression of P_{rad66}–LacZ. The figure below shows the effect of different 50 bp deletions in the upstream regulatory region on the amount of ß-galactosidase activity expressed by the reporter gene.

-300	-250	-200	-150	-100	-50	+1		ß-galactosidase
								_–UV_ _+UV_

		–UV	+UV
wt	———————————————————————— LacZ	–	+
1	——— ——————————————————— LacZ	–	+
2	————— ————————————————— LacZ	+	+
3	————————— —————————————LacZ	–	+
4	——————————— ———————————LacZ	–	+
5	————————————— ———————LacZ	–	+
6	—————————————————— —LacZ	–	–

Deletion 2 gives constitutive expression of P_{rad66}–LacZ whereas deletion 6 gives uninducible expression. In light of the experiments with trans-acting factors, propose functions for the two upstream regulatory segments defined by these deletions. What can you conclude from the fact that deletions 3-5 show normal regulation of P_{rad66}–LacZ?

Region 6 is required for B-gal expression under the rad66 promoter when UV light is present since a deletion of this sequence gives uninducible expression. It could either be the binding site for an activator or the TATA box required for RNA polymerase recruitment. Since the region is so close to the transcriptional start and we have no evidence of an activator working directly on the Rad66 promoter, this **region 6 is likely the location of the TATA box**. Region 2 is required for repression of B-gal expression under the rad66 promoter when UV light is absent since a deletion of this sequence gives constitutive expression. This **region 2 must be the binding site for a repressor (URS) and is likely where the Reg1 repressor binds** to the Rad66 promoter to repress transcription from the promoter in the absence of UV light. Since deletions 3,4,and 5 show normal regulation of the B-gal gene under the Rad66 promoter, **these DNA segments must not serve a vital role in the transcriptional initiation or regulation of the Rad66 promoter**. Also, deletions in this area demonstrate that the distance between the Reg1⁻ repressor and the start of transcription can be altered without affecting the regulation of this promoter. Therefore, the repressor's interaction with the transcriptional machinery to cause repression in the absence of UV light is **distance independent** to a certain degree.

2. You are studying an autosomal recessive human mutation that causes blindness. Through genetic linkage analysis, you map the mutation to a 0.6-Mb region of chromosome 12. Unfortunately, the genomic sequence of this region is not available, and the responsible gene has not yet been defined at a molecular level. However, you have 6 BACs (of unknown order) and 10 STSs (of unknown order) that derive from this region of chromosome 12. By PCR, you test each of the 10 BACs for the presence (+) or absence (-) of each of the 10 STSs, and you obtain the following results:

		STSs									
		1	2	3	4	5	6	7	8	9	10
	A	+	+	-	-	-	-	-	-	-	-
	B	-	-	-	-	-	+	+	-	-	-
BACs	C	-	-	+	-	+	-	-	-	-	+
	D	+	-	-	-	-	-	-	+	-	-
	E	-	-	-	+	-	-	+	-	-	-
	F	+	-	+	-	+	-	-	+	-	-

(a) Construct a physical map of this region that shows the order of the BACs, the overlaps among the BACs, the order of the STSs, and the placement of the STSs within the BACs. Are there any ambiguities within your map?   If so, describe them.

To build a physical map of a region using the STS content mapping technique you should use the following strategy:

1. Pick any BAC.
2. Note which STSs it contains.
3. Look for overlapping BACs, i.e. BACs that contain the same STSs. They may extend your map "to the right" or "to the left", that is they could lie on either side of your original BAC.
4. If you can't find a BAC that contains the most "left" or the most "right" STS, there is probably a gap in your map. In this case, just pick another BAC (from the ones still unordered) and start working from the new BAC to order the rest of the BACs and STSs.

The physical map of the genomic region can be as follows:

STSs:
2   1   8   [3,5]   10   gap   6   7   4

A   B
              D   F   E
                                    C
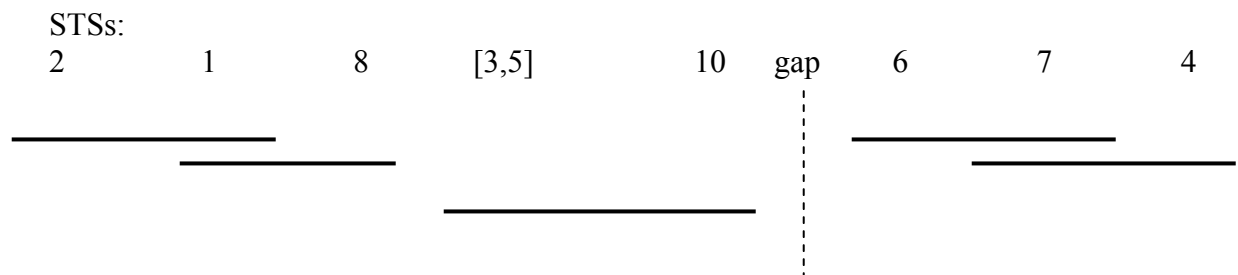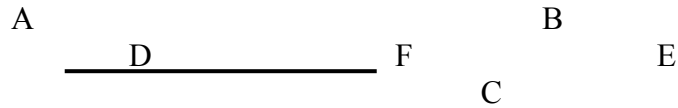
There are several ambiguities within this map. First of all you have no information
to determine the order of STSs 3 &5 relative to 8 and 10, so the order of these
four markers can be 8, 3, 5, 10 or 8, 5, 3, 10.

You also have a gap in your map, so it is impossible from this data to determine
which BACs and STSs lie "to the left" of the gap, and which are "to the right".
Therefore, there are four possible orders:

            1) 2,…..,10...gap…6,7,4
            2)2,…..,10…gap…4,7,6
            3) 10,…,2…gap…6,7,4
            4) 10,…,2…gap…4,7,6

Finally, STS 9 is not present in any of the BACs, so clearly you can not
determine its location on your map. It could be in the gap or on either end of the
entire region.

(b) You obtain the complete sequence of one mRNA that derives from the 0.6-Mb
region. Design an experiment (without sequencing any of the BACs) that will show
which of the BACs contain the 5' and 3' ends of the gene from which this mRNA is
derived.

Design two pairs of PCR primers, one pair corresponding to the 5' end, and one
pair corresponding to the 3' end of the gene (this is possible since you know the
mRNA sequence). Probe each BAC by PCR with each of these sets of primers.
The BAC(s) that give you a PCR product with primers for the 5' end of the gene
contain this region. The same is true for the 3' end.

(c) Several months after constructing the physical map, you obtain the full genomic
sequence of the 0.6-Mb region. Using electronic tools, you identify three predicted
genes, which are 2 kb, 15 kb, and 100 kb in length. The mRNA from part (b) is ~1900
bp long. Based only on the information that you have been given, can you tell which of
the three predicted genes is likely (or unlikely) to be the source of the ~1900-bp mRNA?
Briefly explain your answer.

Any of the three predicted genes could be your gene of interest, since each one
is longer than the mRNA sequence, and mammalian genes include introns which
can vary greatly in size. The 2kb gene is least likely to be your gene of interest,

since it is only 200bp longer than the mRNA, and introns **usually** span more than 10% of the gene.

(d) A few weeks later, you obtain DNA samples from 10 patients who are blind because of the mutation mapping to this region. You also obtain DNA samples from 10 individuals with normal vision. How would you determine which one of the three predicted genes is responsible for the blindness that maps to this region?

a) You can sequence the three genes from the DNA of the affected and normal individuals. The gene responsible for blindness should have a sequence variation in every one of the affected individuals that does not occur in any of the normal individuals. These mutations are likely to be in the exons of the gene.

3. As we will study later in the semester, there are genes in the human and mouse genomes that control cell proliferation. Mutations in such genes can result in uncontrolled proliferation and, as a result, cancer. For some such "cancer genes," a single mutant allele is sufficient to cause a diploid cell to proliferate wildly and cause cancer; such mutant alleles act **dominantly** to promote proliferation. For other "cancer genes," mutant alleles act **recessively** to promote proliferation; in these case, both copies of the gene (both alleles) must be defective to make the cell cancerous.

We want to make a mouse model of cancer where the defect is caused by a **dominantly** acting mutation in a gene called RAS. This mutation results in uncontrolled proliferation and thus a predisposition to cancer even in the presence of wild-type alleles of the RAS gene.

(a) Would it be possible to generate a suitable mouse model using pronuclear injection (transgenes)? Briefly explain why or why not.

Yes, if the dominant allele of the RAS gene is randomly inserted anywhere into the genome as occurs in pronuclear injection you would observe the uncontrolled proliferation phenotype **since it is dominant to the wild-type RAS allele**.

(b) Would it be possible to generate a suitable mouse model using gene-targeting techniques? Briefly explain why or why not.

Yes, the dominant allele of the RAS gene could be targeted into the genome in a specific location **using the gene targeting "knock-in" technique** to give a mouse with an uncontrolled proliferation phenotype since it is dominant to the wild-type RAS allele. The difference between a mouse made using this

technique or pronuclear injection would be that the location where the dominant allele was inserted would be known in the gene targeting case. Also, the pronuclear injection mouse would have three alleles of RAS while the targeted mouse would have only two.

(c) If a suitable mouse model could be generated using either of the two techniques, which one would you choose? Briefly explain your reasoning.

You could use either technique, but **pronuclear injection (transgenes)** would be the more practical choice since it involves a much simpler procedure than gene-targeting but gives effectively the same final result.

(d) Draw a labeled diagram of the transgene or targeting construct that you would use.

[Dominant Allele of the RAS gene], there is no need for a selectable marker in this case because random integration of the entire linear piece of DNA as a result of pronuclear injection is effective half of the time. You could verify integration by performing a PCR of an area unique to the dominant allele of the RAS gene.

(e) Once the transgene or targeting construct has been properly integrated into the mouse genome, are additional breeding experiments required before you've generated a mouse with a predisposition to cancer? Briefly explain why or why not.

No, one copy of the dominant allele of RAS is enough to give the desired phenotype and since the mutation is dominant to wild-type the endogenous copies of the wild-type RAS gene do not need to be deleted.

You now want to make want to make a mouse model of cancer where the defect is caused by a **recessively** acting mutation in a gene called RB. This gene's normal function is to keep cells from proliferating out of control and one wild-type copy of the RB gene is sufficient to do so. However, if both RB alleles are defective, the cell will likely become cancerous as a result of uncontrolled proliferation.

(f) Would it be possible to generate a suitable mouse model using pronuclear injection (transgenes)? Briefly explain why or why not.

No, random integration of a defective recessive RB gene would not give the desired phenotype since there would also be two **dominant wild-type alleles present in the mouse which would mask the effect of the recessive mutation**.

(g) Would it be possible to generate a suitable mouse model using gene-targeting techniques? Briefly explain why or why not.

Yes, if you target the endogenous wild-type RB alleles by using gene targeted "knock-out" technology you can generate a mouse with recessive defective copies of the RB gene by **targeted knock-outs of the wild-type dominant alleles** or by targeted knock-ins that replace wildtype versions of RB with non-functional versions of RB.

(h) If a suitable mouse model could be generated using either of the two techniques, which one would you choose? Briefly explain your reasoning.

N/A, only one of the above techniques (gene targeting) could be used to make this mouse.

(i) Draw a labeled diagram of the transgene or targeting construct that you would use.

5' region of RB sequence   3' region of RB sequence
--**[Neo resistance]**------------------------------------

where Neo is a mamalian selectable marker flanked on both sides by DNA sequence homologous to the region of the genome where you want to insert it. In this case, the flanking regions should be homologous to the wildtype RB gene. (This would be an example of the targeted "knock-out" technology described in class). You could have also drawn out a construct where Neo *and* a defective allele of RB were both flanked by regions surrounding the RB sequence to effectively replace the wildtype allele of RB with a defective copy and the selectable marker. (This would be an example of the "knock-in" technology described in class).

(j) Once the transgene or targeting construct has been properly integrated into the mouse genome, are additional breeding experiments required before you've generated a mouse with a predisposition to cancer? Briefly explain why or why not.

Yes, in order for this mouse to display the desired phenotype of uncontrolled proliferation, both copies of the wildtype RB gene must be deleted since the mutation is recessive to wildtype. In the original targeting experiment, the construct only integrates into one chromosome causing one wildtype copy of RB to be deleted. The resulting mouse is heterozygous at the RB locus with one wildtype and one defective copy of RB and is a chimera. By mating with wildtype, you would first create a non-chimeric heterozygous mouse. Then, you could generate a homozygous RB defective mouse by breeding two non-chimeric

heterogygous RB defective mice from the same knock-out experiment together and look for their homozygous RB defective progeny by PCR.

7.03 Problem Set 6

Due before 5 PM on Thursday, November 21, 2002
Hand in answers in recitation section or in the box outside the class

1. Consider the ABO "blood type" gene. This single gene has three alleles called A, B, and O. There are four resultant "blood types" (phenotypic classes), as follows:

Blood type	Genotype(s)
A	AA or AO
B	BB or BO
A B	A B
O	O O

Assume Hardy-Weinberg equilibrium.

In Norway, the frequencies of the A and B alleles are 0.26 and 0.07, respectively. Estimate the following (showing your calculations):

(a) The frequencies of the six genotypes (AA, AO, BB, BO, AB, and OO) in Norway.

(b) The frequencies of the four blood types in Norway.

(c) You now examine the ABO blood types in a new population. You observe that the A, B, and O blood types all have equal frequencies in this population. What are the frequencies of the A and O alleles in this population? (Show your calculation or otherwise explain your answer.)

(d) You now examine the ABO blood types in yet another population. You observe that the AB blood type has a frequency of 0.50 in this population. What are the frequencies of the A and O alleles in this population? (Show your calculation or otherwise explain your answer.)

2. 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.

First, consider an autosomal recessive disease that is usually lethal in childhood, and that has an incidence among newborns of 1/3000.

(a) What mutation rate would be required to maintain this frequency in the population?

(b) If your answer to part (a) seems too high, perhaps the explanation for the incidence of the disease is heterozygote advantage. How large would this heterozygote advantage have to be (assuming the mutation rate is negligible)?

Colorblindness is an X-linked recessive trait that is found in about 5 percent of males.

(c) What fraction of females are heterozygous carriers of the colorblindness allele?

(d) What fraction of females are homozygous for the colorblindness allele?

Hemophilia is an X-linked recessive disease that affects 1/10,000 newborns males. Assume that the fitness of these males is zero.

(e) Assuming equal mutation rates in the two sexes, what fraction of females are heterozygous carriers of the hemophilia allele? Why is the method used in answering parts (c) and (d) not suitable in answering this question?

3. In answering the various parts of this question, show your calculations, and state any additional simplifying assumptions that you employ.

In lecture 26 we learned that the frequency of PKU among newborns is approximately 1/10,000 (when the parents are unrelated).

(a) What is the risk of a child with PKU if the parents are siblings?

(b) What is the risk of a child with PKU if the parents are uncle and niece?

(c) What is the risk of a child with PKU if the parents are second cousins?

P, Q, and R are three different inbred, true-breeding strains of mice. A hybrid between P and Q is mated to R, and a hybrid between Q and R is mated to P. A male from the first mating is mated to a female from the second mating.

(d) What is the inbreeding coefficient of their progeny?

7.03 Problem Set 6
SOLUTIONS (2002)

1. Consider the ABO "blood type" gene. This single gene has three alleles called A, B, and O. There are four resultant "blood types" (phenotypic classes), as follows:

Blood type	Genotype(s)
A	AA or AO
B	BB or BO
AB	AB
O	OO

Assume Hardy-Weinberg equilibrium.

In Norway, the frequencies of the A and B alleles are 0.26 and 0.07, respectively. Estimate the following (showing your calculations):

(a) The frequencies of the six genotypes (AA, AO, BB, BO, AB, and OO) in Norway.

Frequency of the A allele $= p = 0.26$
Frequency of the B allele $= q = 0.07$
Because the total frequency of all the alleles must equal 1 ($p + q + r = 1$):
Frequency of the O allele $= r = 0.67$
Note that the genotype frequencies should add up to 1.
$f(AA) = p^2 = (0.26*0.26) = \textbf{0.0676}$
$f(A0) = pr + rp = 2pr = 2(0.26*0.67) = \textbf{0.3484}$
$f(BB) = q^2 = (0.07*0.07) = \textbf{0.0049}$
$f(BO) = qr + rq = 2qr = 2(0.07*0.67) = \textbf{0.0938}$
$f(AB) = pq + qp = 2pq = 2(0.26*0.07) = \textbf{0.0364}$
$f(OO) = r^2 = (0.67*0.67) = \textbf{0.4489}$

(b) The frequencies of the four blood types in Norway.

The frequency of each blood type is equal to the sum of the corresponding genotype frequencies:

$f(\text{blood type A}) = f(AA) + f(AO) = 0.0676 + 0.3484 = \textbf{0.416}$
$f(\text{blood type B}) = f(BB) + f(BO) = 0.0049 + 0.0938 = \textbf{0.0987}$
$f(\text{blood type AB}) = f(AB) = \textbf{0.0364}$
$f(\text{blood type O}) = f(OO) = \textbf{0.4489}$

(c) You now examine the ABO blood types in a new population. You observe that the A, B, and O blood types all have equal frequencies in this population. What are the frequencies of the A and O alleles in this population? (Show your calculation or otherwise explain your answer.)

$$f(\text{blood type A}) = f(\text{blood type B}) = f(\text{blood type O})$$
$$f(AA) + f(A0) = f(BB) + f(BO) = f(OO)$$
$$(p^2 + 2pr) = (q^2 + 2qr) = (r^2)$$

From this series equation, we can get the following equations:
$$p^2 + 2pr = r^2$$
$$q^2 + 2qr = r^2$$

By inspection, p and q must be equal for both of the above equations to be true
$$p = q$$

We also know that the sum of all three allele frequencies must be 1
$$p + q + r = 1$$

Since we determined that $p = q$, we can rewrite r in terms of just p or q
$$r = 1 - 2p.$$

(We chose to express r in terms of p, using q would be mathematically equivalent.)
Substituting this expression for r back into $p^2 + 2pr = r^2$ gives the following equation
$$p^2 + 2p(1 - 2p) = (1 - 2p)^2$$

Which simplifies to
$$7p^2 - 6p + 1 = 0$$

Using the quadratic formula,
$$p = [6 +\!\!- v(36\text{-}28)] / 14 = 0.631 \text{ or } 0.227$$

If $p = 0.631$, then $(p + q + r)$ would definitely be greater than 1, so we will reject this answer and conclude that $p = 0.2265$
We know that $[p = q]$ and $[r = (1 - p - q)]$ so the calculated allele frequencies are as follows:

p = Frequency of allele A = 0.2265
r = Frequency of allele O = 0.5470
[q = Frequency of allele B = 0.2265]

(d) You now examine the ABO blood types in yet another population. You observe that the AB blood type has a frequency of 0.50 in this population. What are the frequencies of the A and O alleles in this population? (Show your calculation or otherwise explain your answer.)

As shown by the graph on page 6 of Lecture 25, the maximal frequency of a heterozygote genotype is 0.5. This occurs when the frequencies of the two alleles making up the heterozygote are each 0.5. Therefore, since we know that $f(AB)=0.5$, it must be true that $f(A)=0.5$, $f(B) = 0.5$, and $f(O)= 0$ (because the frequencies of the three alleles must add up to 1).
This can also be solved algebraically:
Since $f(\text{blood type AB}) = f(AB) = 2pq = 0.50$, we can get: $q = 0.25/p$
Now we need to use $[p + q + r = 1]$ and $[q = 0.25/p]$ to derive the following relation:

$$p + (0.25/p) + r = 1$$

Which simplifies into the following quadratic equation:

$$p^2 + (r - 1)p + 0.25 = 0$$

To solve for p, we can use the quadratic formula:

$$p = \{-(r - 1) +- \sqrt{[(r - 1)^2 - (4*1*0.25)]}\} / 2$$
$$p = \{(1 - r) +- \sqrt{(r^2 - 2r)}\} / 2$$

Since p needs to be a real number between 0 and 1, we need the term under the square root to be greater than or equal to zero:

$$(r^2 - 2r) = 0$$
$$r(r - 2) = 0$$

This inequality is satisfied if [r < 0], [r = 2], or if [r = 0].

Of these three conditions on r, only [r = 0] fits the genetic condition that allele frequencies (p, q, and r) range from 0 to 1

Therefore the only possible answer is [r = 0]. By substituting this value of r into the quadratic equation, we get [p = 0.5], and since [q = 0.25/p], we get [q = 0.5] as well.

Frequency of the A allele = p = 0.5

Frequency of the O allele = r = 0

[Frequency of the B allele = q = 0.5]

We can verify this answer by seeing that the three allele frequencies add up to one and that the frequency of blood type AB (given the allele frequencies above) is indeed 0.50.

2. 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.

First, consider an autosomal recessive disease that is usually lethal in childhood, and that has an incidence among newborns of 1/3000.

(a) What mutation rate would be required to maintain this frequency in the population?

NOTE: For all parts of this problem, migration is assumed not to occur.

Assumption: since we are told that the recessive disease is usually lethal in childhood, we assume that the affected individuals never have offspring, meaning the selective disadvantage S = 1.

If we call the recessive disease allele "a" and define "q" as the frequency of allele a, then

$$f(a/a) = q^2 = 1/3000 = 3.33*10^{-4}$$
$$q = \sqrt{(3.33*10^{-4})} = 1.83*10^{-2}$$

We know that for recessive diseases, the following relationship exists between S, q, and mutation rate μ:

$$q = \sqrt{(\mu/S)}$$

By substituting S = 1, we get

$$\mu = q^2 = \textbf{1/3000} = \textbf{3.33*10}^{-4}$$

(b) If your answer to part (a) seems too high, perhaps the explanation for the incidence of the disease is heterozygote advantage. How large would this heterozygote advantage have to be (assuming the mutation rate is negligible)?

If mutation rate is negligible, we know the following relationship between h (the heterozygote advantage), S, and q:

$$q = h/S$$

Using S=1 and our know value of q:

$$h = v(3.33*10^{-4}) = \mathbf{1.83*10^{-2} = 1.83\%}$$

This means: On average, heterozygotes have 1.83% more offspring than the average individual.

Colorblindness is an X-linked recessive trait that is found in about 5 percent of males.

(c) What fraction of females are heterozygous carriers of the colorblindness allele?

Assumption: Colorblindness is not associated with any selective disadvantage (S = 0), and no new mutation is occurring.
5% of males are colorblind means that $f(X^cY) = 0.05$, where X^c indicates an X chromosome that carries the allele responsible for colorblindness (X denotes X chromosomes carrying all other non-colorblindness alleles).
Since each male only has one X chromosome, the frequency of the colorblindness allele must also be q = 0.05, and that of non-colorblindness alleles is p = 1 - 0.05 = 0.95.
f(female heterozygote carriers) = $f(X^c/X) = 2qp = 2(0.05*0.95) = \mathbf{0.095 = 9.5\%}$

(d) What fraction of females are homozygous for the colorblindness allele?

Assumption: Colorblindness is not associated with any selective disadvantage (S = 0), and no new mutation is occurring.
f(homozygous females) = $f(X^c/X^c) = q^2 = (0.05*0.05) = \mathbf{0.0025 = 0.25\%}$

Hemophilia is an X-linked recessive disease that affects 1/10,000 newborns males. Assume that the fitness of these males is zero.

(e) Assuming equal mutation rates in the two sexes, what fraction of females are heterozygous carriers of the hemophilia allele? Why is the method used in answering parts (c) and (d) not suitable in answering this question?

The effect of mutation was negligible in (c) and (d) but is not negligible in this problem.

1/10,000 males are affected by hemophilia, and therefore are the genotype Xa/Y.
If we assume that S=1 for hemophilia, then 1/10,000 males die of hemophilia.
Roughly 0/10,000 females die of hemophilia.

Given that males have one X and females have two X chromosomes, this means that of those 20,000 people above, who have 30,000 X chromosomes between them (or 30,000 alleles of the gene responsible for hemophilia), 1/30,000 alleles are removed due to death.

If this population is at an equilibrium between mutation and selection, then the number of alleles removed by selection (Δqsel) should equal the number of alleles added by mutation (Δqmut), so 1/30,000 alleles are added by mutation.

When a male is created, his chance of getting an Xa = 1/10,000. His Xa must come from his mother, and must either be the result of a new mutation, or must be inherited from a carrier mother of genotype XA/Xa (because Xa/Xa females will not produce offspring).

The chance of a male getting an Xa allele from a new mutation = 1/30,000 as we have stated above.

Therefore the chance of a male getting an Xa allele from a pre-existing mutation that existed in a carrier mother = (1/10,000) - (1/30,000) = (2/30,000).

The probability that a male gets an Xa from his mother is 1/2 * the probability that his mother has an Xa allele. Therefore the chance that his mom had an Xa allele = 2 * (2/30,000).

If a mother has an Xa allele, then she must be a carrier, because Xa/Xa females do not produce offspring. Therefore the chance of having a female carrier =

$2 * (2/30,000) = 1.3 * 10^{-4}$.

3. In answering the various parts of this question, show your calculations, and state any additional simplifying assumptions that you employ.

In lecture 26 we learned that the frequency of PKU among newborns is approximately 1/10,000 (when the parents are unrelated).

(a) What is the risk of a child with PKU if the parents are siblings?

NOTE: In all parts of this problem, we assume that S=0, and that no new mutation and or migration are occurring.

If the frequency of PKU (an autosomal recessive disease) among newborns is 1/10,000 when the parents are unrelated, then the allele frequency (q) of the disease allele (a) is

$$q = v(1/10,000) = 0.01$$

To estimate the risk of getting a PKU affected child from a brother-sister mating, we need to calculate the inbreeding coefficient for such a mating:
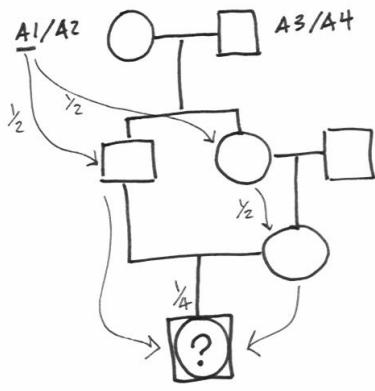
[From lecture 27] F = P(homozygous by descent) = 4 (1/2 * 1/2 * 1/4) = 1/4

The risk, then, is just equal to (F * q) = (1/4 * 0.01) = **$0.0025 = 2.5*10^{-3}$**

(b) What is the risk of a child with PKU if the parents are uncle and niece?

Again, we need to calculate the inbreeding coefficient F.

The pedigree looks like the following (the arrows follow the heredity of the A1 allele only):

$F = P(\text{homozygous by descent}) = 4\ (1/2 * 1/2 * 1/2 * 1/4) = 1/8$

$\text{Risk} = (F * q) = 0.00125 = 1.25*10^{-3}$

(c) What is the risk of a child with PKU if the parents are second cousins?

Second cousins share the same great-grandparents, so the pedigree is as follows (the arrow follows the heredity of the A1 allele only):
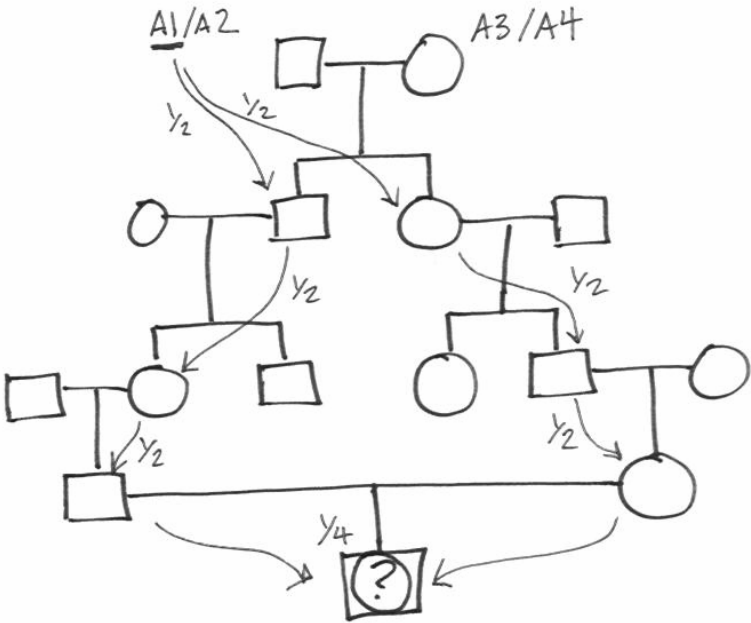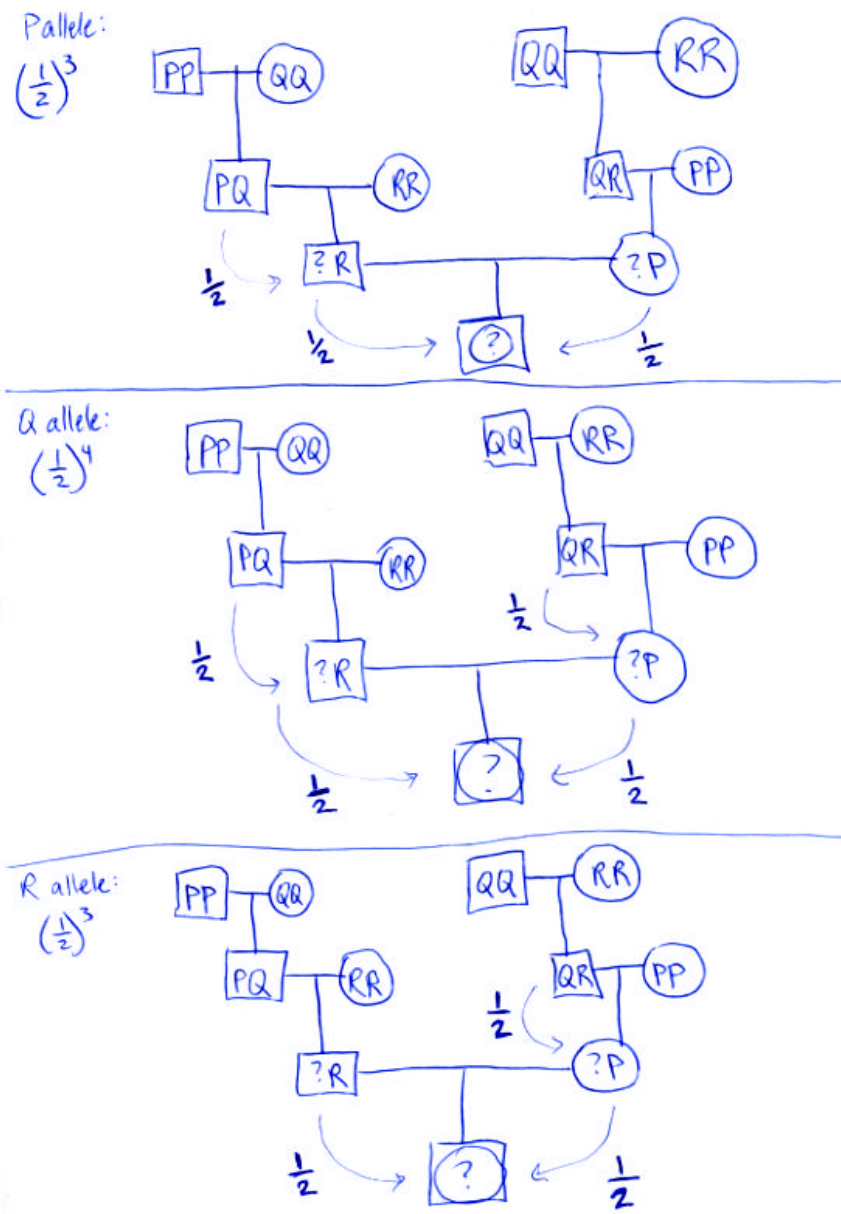

$F = P(\text{homozygous by descent}) = 4\ (1/2 * 1/2 * 1/2 * 1/2 * 1/2 * 1/2 * 1/4) = 1/64$

$\text{Risk} = (F * q) = 1.5625*10^{-4}$

P, Q, and R are three different inbred, true-breeding strains of mice. A hybrid between P and Q is mated to R, and a hybrid between Q and R is mated to P. A male from the first mating is mated to a female from the second mating.

(d) What is the inbreeding coefficient of their progeny?

Since P, Q, and R are different true-breeding strains, they are homozygous for different alleles at all genetic loci.

The pedigree looks like the following:

F = probability of (homozygous by descent) = prob(PP) + prob (QQ) + prob (RR)
= (1/2 * 1/2 * 1/2) + (1/2 * 1/2 * 1/2* 1/2) + (1/2 * 1/2* 1/2) = 1/8 + 1/16 + 1/8 = **5/16**

7.03 Problem Set 7

Due before **4 PM** on **Friday**, December 6, 2002. Hand in answers to
problems 1 and 2 in recitation section or in the box outside the class.

1. You are conducting genetic linkage studies of an autosomal recessive disease whose chromosomal location has not been firmly established. You are focused on two SSR markers that may be linked to each other and to the disease. Here are two families in which some individuals are affected:
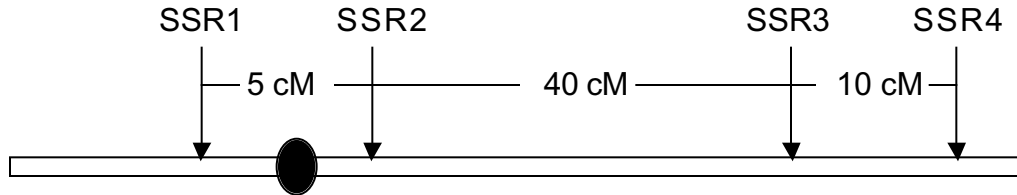
Family 1

SSR42
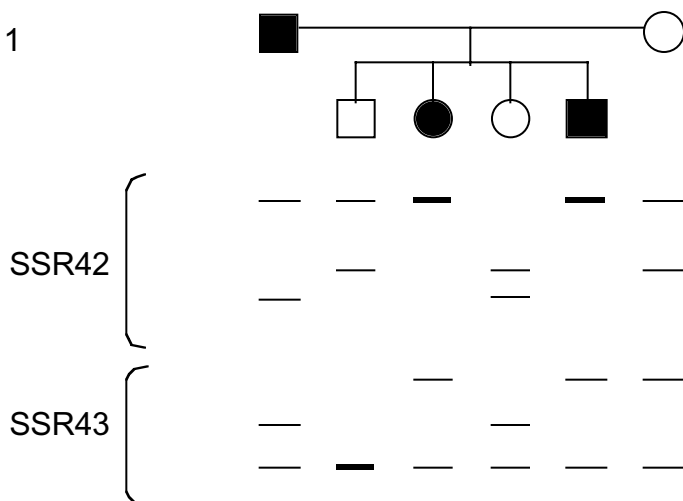
SSR43

Family 2
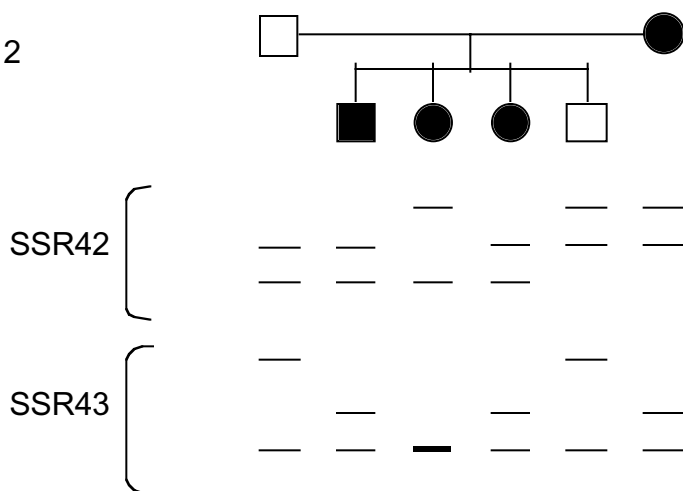
SSR42

SSR43

1. (continued).

Calculate LOD scores for linkage at . = 0.02 between:

(a) The disease and SSR42 in Family 1.

(b) The disease and SSR43 in Family 1.

(c) The disease and SSR42 in Family 2.

(d) The disease and SSR43 in Family 2.

(e) SSR42 and SSR43 in Family 1.

(f) SSR42 and SSR43 in Family 2.

Is it appropriate to add together the LOD scores calculated in:

(g) Parts (a) and (b) of this question? Why or why not?

(h) Parts (b) and (d) of this question? Why or why not?

(i) Parts (e) and (f) of this question? Why or why not?

(j) What conclusion (with respect to genetic linkage) can you publish based on these findings?

2. Trisomy 18 is one of the most common trisomies observed in human populations. You prepare DNA samples from two unrelated infants, both with trisomy 18, and from their parents. You then type the infants and their parents for four SSRs distributed along chromosome 18:
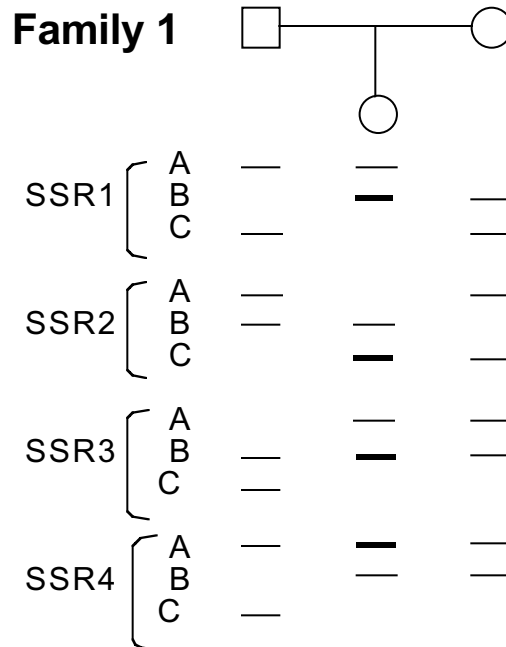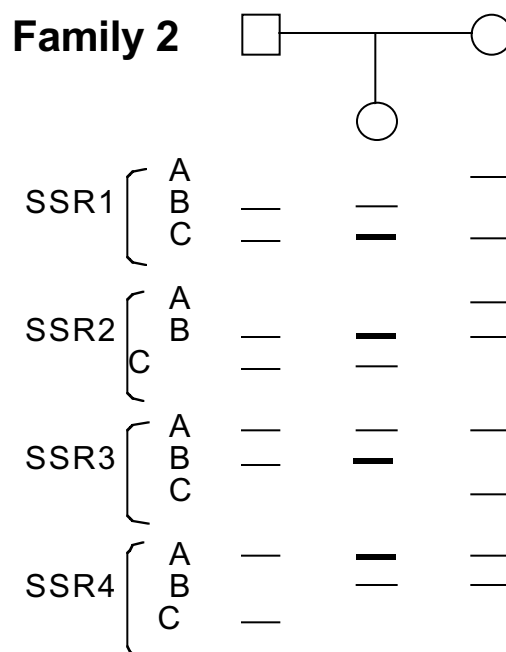

(a) In which parent did nondisjunction occur in Family 1?

(b) In which division of meiosis did nondisjunction occur in Family 1?

(c) Sketch the meiotic event in which nondisjunction occurred in Family 1. Your drawing should include the SSRs present along chromosome 13.

2. (continued).

Family 2

SSR1 [A B C

SSR2 [A B C

SSR3 [A B C

SSR4 [A B C

(d) In which parent did nondisjunction occur in Family 2?

(e) In which division of meiosis did nondisjunction occur in Family 2?

(f) Sketch the meiotic event in which nondisjunction occurred. Your drawing should include the SSRs present along chromosome 13.

Problem Set 7 Solutions

1. Note that in this problem, the disease is autosomal RECESSIVE
   Also, we will name the SSR42 alleles B, C, D (top to bottom) and the SSR43 alleles P, Q, R (top to bottom)

(a) In family 1, the affected father is homozygous for the recessive disease allele (RD/RD) and the mother must be a carrier (RD/+) since some of their offspring are affected.

   With respect to the disease, the segregation of the father's SSR42 alleles is uninformative because his two alleles of SSR42 (B and D) are both in phase with the same RD disease allele (i.e. the two possible phases in this case are identical, so we cannot determine whether recombination occurred between the disease and SSR42 loci). Therefore, the father is uninformative and we use only the carrier mother to generate a LOD score.

   The carrier mother is RD/+ and B/C (for SSR42). Since we do not know the phase relationship, we have to assume two phases exist, one where RD is in phase with B and the other where RD is in phase with C. Here are these two possible phase relationships in the mother:

   Phase 1:
   $$\frac{\text{RD} \quad \text{B}}{+ \quad \text{C}}$$

   Phase 2:
   $$\frac{\text{RD} \quad \text{C}}{+ \quad \text{B}}$$

   $\theta = 0.02$ means that the crossover frequency (between the two genetic loci) is 0.02, so 0.98 is the frequency of the parental classes. Therefore, for each phase relationship, we observe each crossover gamete at a frequency of 0.01 (0.5*0.02) and each parental gamete at a frequency of 0.49 (0.5*0.98)

   Assuming phase 1, all four children correspond to a parental gamete $[(0.49^4)(0.01^0)]$. Assuming phase 2, all four children correspond to crossover gametes $[(0.49^0)(0.01^4)]$.

   LOD $(\theta = 0.02)$ = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
         = 0.87

(b) The same logic applied to part (a) can be used to establish the linkage relationship between SSR43 and the disease locus.

   In this case, the two phases in the mother look like the following:

Phase 1:

$$\dfrac{\text{RD} \quad \text{P}}{+ \quad \text{R}}$$

Phase 2:

$$\dfrac{\text{RD} \quad \text{R}}{+ \quad \text{P}}$$

Again, assuming phase 1, all four children correspond to a parental gamete $[(0.49^4)(0.01^0)]$ and assuming phase 2, all four children correspond to crossover gametes $[(0.49^0)(0.01^4)]$.

So the LOD score equation is identical:

LOD ($\theta = 0.02$) = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
            = 0.87

(c) In family two, the affected mother is uninformative, while the carrier father is, for the same reasons as discussed in part (a)

Again, phase is unknown in the informative parent. Again, all four children represent parental classes assuming one phase and crossover classes assuming the other phase.

Phase 1:

$$\dfrac{\text{RD} \quad \text{D}}{+ \quad \text{C}}$$

Phase 2:

$$\dfrac{\text{RD} \quad \text{C}}{+ \quad \text{D}}$$

So the LOD equation is identical:

LOD ($\theta = 0.02$) = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
            = 0.87

(d) Once again, all the numbers are identical. The possible phases in the father are:

Phase 1:

$$\dfrac{\text{RD} \quad \text{R}}{+ \quad \text{P}}$$

Phase 2:

$$\dfrac{\text{RD} \quad \text{P}}{+ \quad \text{R}}$$

LOD (θ= 0.02) = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
       = 0.87

(e) None of the parents are homozygous for SSR42 or SSR43 (neither of them have two copies of the same allele) so both are informative with respect to linkage between the two SSRs. The key thing to realize for this part of the question (and the following part (f)), is that SSR inheritance from each parent is independent of the other. That is, the SSR alleles that the mother gives do not affect which SSR alleles the father gives. So the total LOD score for linkage between SSR42 and SSR43 in family 1 is the sum of the LOD scores from both sides of the family.

Again, we do not know phase.

The possible phases in the dad are:

Phase 1:
$$\frac{B \qquad Q}{D \qquad R}$$

Phase 2:
$$\frac{B \qquad R}{D \qquad Q}$$

And the corresponding LOD score is:
LOD (θ = 0.02) = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
       = 0.87

The possible phases in the mom are:

Phase 1:
$$\frac{B \qquad P}{C \qquad R}$$

Phase 2:
$$\frac{B \qquad R}{C \qquad P}$$

And the corresponding LOD score is:
LOD (θ = 0.02) = **log [(0.5(0.49^4)(0.01^0) + 0.5(0.49^0)(0.01^4)) / (0.25^4)]**
       = 0.87

Since these two scores were derived independently, the total LOD score is
0.87 + 0.87 = 1.74

(f) The work is identical in this part. Only the phase diagrams are different:

Dad's phases are:

Phase 1:
$$\frac{D \quad\quad P}{C \quad\quad R}$$

Phase 2:
$$\frac{D \quad\quad R}{C \quad\quad P}$$

Mom's phases are:

Phase 1:
$$\frac{C \quad\quad Q}{B \quad\quad R}$$

Phase 2:
$$\frac{C \quad\quad R}{B \quad\quad Q}$$

Conveniently, the numbers are identical as in part (e) so the total score, once again, is **0.87 + 0.87 = 1.74**

To add a set of LOD scores together, two criteria must be met:
1. Each score in the set must have been independent of all the others
2. All the scores must refer to linkage between the same two genetic locations.

(g) **No**, we cannot add these two LOD scores because the two pairs of genetic locations are different (one LOD score describes linkage between **SSR42** and the disease while the other describes linkage between **SSR43** and the disease).

(h) **Yes**, the two families are independent of each other, and both scores describe linkage between the same two genetic locations (SSR43 and the disease).

(i) **Yes**, both criteria are satisfied.

(j) If we total our LOD scores for the disease and SSR42 (for families 1 and 2), we get 0.87 + 0.87 = **1.74**. This falls short of our target LOD score of 3, so we cannot publish any conclusions about linkage at $\theta = 0.02$ between the disease and SSR42.

Similarly, for the disease and SSR 43, the total score is also **1.74**, a non-publishable result.

However, the total LOD score for linkage between SSR42 and SSR43 is 1.74 + 1.74 = **3.48**. Therefore, we can publish the conclusion that SSR42 and SSR43 are very likely to be linked at $\theta = 0.02$.

2. Note that for this problem, we included the SSR number with the letter to make the notation clearer. In other words, the SSR1 alleles are called 1A, 1B, and 1C, and the SSR2 alleles are called 2A, 2B, 2C, and so on.

(a) In family 1, **mom** gave to the daughter two copies of 1B, two copies of 2C, and one each of 3A, 3B, 4A, 4B.

(b) Meiosis II. Looking at the centromere-linked SSR1 and SSR2 only, we see that the mother passed on two copies of the same allele at each of these loci.

(c)
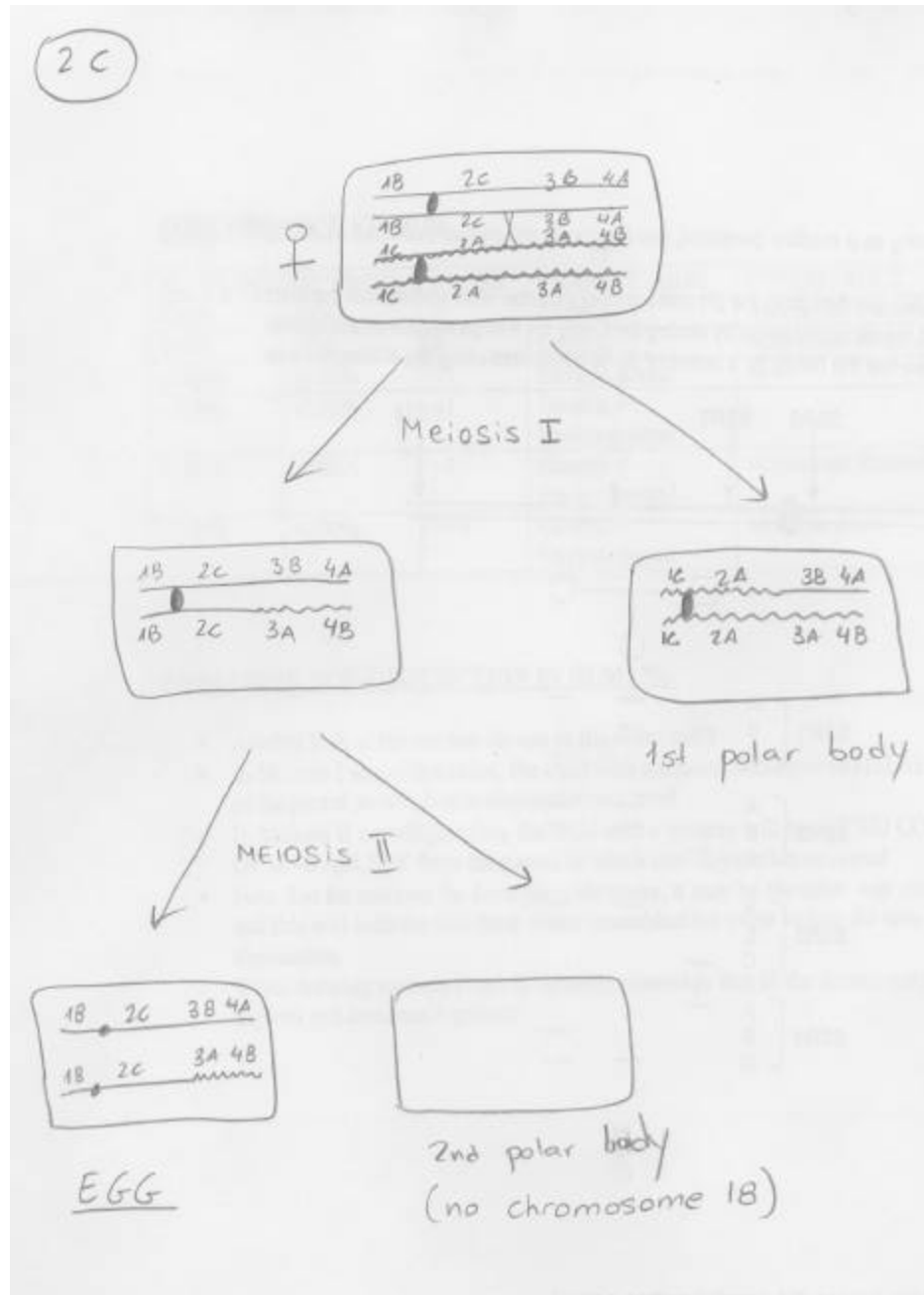
(d) In family 2, **dad** gave to the daughter one each of 1B, 1C, 2B, 2C, and two copies each of 3B and 4A.

(e) Meiosis I. Looking at the centromere-linked SSR1 and SSR2 only, we see that the father passed on one copy of each of his alleles.

(f)

this gamete
gave rise to
the trisomic
daughter