

Problem Sets

Fall 2000

7.03 Problem Set 1

Due before 5 PM on Thursday, September 21

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

1. Wild type yeast form tan colonies. You have isolated twelve mutants that make white colonies — six mutants are in mating type a and six are in mating type α . To determine the gene relationships between mutants, you perform all of the possible pairwise matings shown in the table below. A “T” at the intersection of the two parental strains indicates that the diploid makes a tan colony whereas a “W” indicates that the diploid makes a white colony.

Mutant:	1	2	3	4	5	6
7	T	W	W	T	T	T
8	T	T	W	T	T	T
9	W	T	W	W	W	T
10	W	T	W	W	W	T
11	W	W	W	W	W	W
12	T	T	W	T	T	T

Give as much information as you can about your new mutations (assume that each strain carries only one mutation). **(a)** Indicate which mutations are dominant and which are recessive, **(b)** state how many genes are represented and which mutations lie the same gene, **(c)** state any remaining ambiguities and suggest some general ways that the ambiguities might be resolved.

2. Consider the following mouse breeding experiment involving two different rare traits. A male mouse with both traits is crossed to a normal female and all of the offspring appear normal. A female offspring from this cross is mated multiple times to a normal male to produce several litters of offspring. A total of 32 offspring are scored as having the following characteristics:

16 normal females

6 normal males

2 males with trait 1

1 male with trait 2

7 males with both traits

(a) What is the mode of inheritance of each of the two traits? Explain your reasoning.

(b) Use the chi-square test to determine whether the two traits appear to be linked. For this test, you are trying to determine whether or not an expectation that the two traits are unlinked differs significantly from the observed data. Note that there are a number of different ways to set up this test, but there is one best way to test for linkage. Show your work and use the table below which gives p values as a function of chi square values and degrees of freedom. For your final answer use a p value < 0.05 as the cut-off for significant deviation from expectation.

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

(c) Based on the data give your best estimate of the distance between the genes for trait 1 and trait 2.

3. The producers of a soap opera have hired you as a consultant. The story line includes two families, each with individuals that have a rare trait. The families are diagramed below — individuals are numbered and those expressing the trait are represented by the filled symbols. The scriptwriters are contemplating a number of different couplings between individuals in the two families. Because they are concerned with the genetic accuracy of the story they want you to figure out what the offspring from each possible mating might be like.



(a) Assume that the rare trait is **recessive**. Consider the possible matings given below. For each, calculate the probability that the child will have the rare trait.

Female 2 and Male 5

Female 6 and Male 4

Female 7 and Male 4

Female 3 and Male 8

(b) Now assume that the rare trait is **dominant**. Again for each of the possible matings given below, calculate the probability that the child will have the rare trait.

Female 2 and Male 5

Female 6 and Male 4

Female 7 and Male 4

Female 3 and Male 8

(c) Finally, assume that the rare trait is **X-linked**. For each of the possible matings given below, calculate the probability that the child will have the rare trait. Explicitly give each probability in the cases where the probabilities for a boy or a girl having the trait differ.

Female 2 and Male 5

Female 6 and Male 1

Female 7 and Male 4

Female 3 and Male 8

Problem Set 1 Solutions.

1a. Those mutants that make tan colonies when crossed to another mutant can be assumed to carry recessive mutations, because the white phenotype is not present in a heterozygote. Those mutants that never make tan colonies as a diploid likely carry dominant mutations, although we can not completely rule out the possibility that these mutants carry mutations in multiple genes (discussed in part C). Thus, mutations **1, 2, 4, 5, 6, 7, 8, 9, 10, and 12 are recessive. Mutations 3 and 11 are most likely dominant.**

1b. Based on non-complementation of the recessive mutations, we can conclude that 1, 4, 5, 9, and 10 form one complementation group (Group A) and are mutations in the same gene. Likewise, 2 and 7 fail to complement and are members of a second complementation group (Group B) representing mutations in a second gene. Mutant 6 complements members of both Group A and Group B and therefore represents a third gene. Mutants 8 and 12 both complement Group A and Group B mutants as well as mutant 6 and therefore represent at least one more gene and possibly two genes. We have no data regarding the phenotype of the 8 x 12 diploid and can't determine whether they are mutations in the same or different genes. Taken together, we can say that the recessive mutations represent at least 4 genes.

We are unable to make any conclusions regarding the number of new genes represented by the dominant mutants, 3 and 8. It is possible that 3 and 8 are mutations in the same gene. It is also possible that one or both of these mutations is in one of the genes that we recovered a recessive mutation in above.

It is also possible that the white phenotype in one or more of the mutants is caused by mutations in more than one gene. Thus, the best answer for the **number of mutants represented is at least 4.**

1c. The most obvious ambiguity is whether 8 and 12 are mutations in the same gene. This could be resolved by generating the 8 x 12 diploid. Note that to do this you would have to sporulate, for example the 8 x wild type diploid, and select for a haploid spore of the appropriate mating type that shows the white phenotype.

A second ambiguity is whether 3 and 11 represent unique genes. One way to approach this question would be to generate the 3 x 11 diploid, sporulate, and look at the segregation pattern of the white phenotype in the resulting haploids. If any of the haploids form tan colonies then 3 and 11 are likely to be mutations in different genes.

A third ambiguity is whether or not the white phenotype in each case is due to a mutation in a single gene. To determine this, you would cross each mutant to wild type, sporulate the resulting diploids, and look for 2 tan : 2 white segregation in the haploids. Any other segregation pattern is inconsistent with the white phenotype being caused by

mutation of a single gene. An extreme example of this possibility would be if a single strain carried recessive mutations in genes representing all of the complementation groups identified. Such a mutant would appear “dominant” by the type of crosses performed in the table, because it would fail to complement every other mutant. To avoid this possibility, you would cross each mutant to a wild type haploid and look at the diploid phenotype.

2a. **Both traits are X-linked recessive.** They must be recessive because all F1 offspring appear normal. We can deduce that both traits are X-linked because, of the 32 progeny shown, all females are normal and only males are affected.

2b. To use the Chi Square test, we first need to formulate a hypothesis that can be tested. In this case an appropriate hypothesis would be: The genes for trait 1 and trait 2 are unlinked and segregate randomly.

Now we must choose our classes for the Chi Square test. First, note that since all F2 females are normal, we can tell nothing about their genotype and must disregard them. The best way to test for linkage is to group X chromosomes of the F2 males as either parental or recombinant based on the arrangement of alleles for traits 1 and 2. For example, the F1 female is a carrier for both traits and both traits are present on the same X chromosome, so her genotype can be represented as $X^{1,2}X$. Thus, there are two parental chromosome types: X and $X^{1,2}$. The other two possible chromosome types (X^1 and X^2) must be due to a crossover between the genes for traits 1 and 2 and are recombinant.

The Chi Square table would look as follows:

Chromosome Type	Observed	Expected
Parental	13	8
Recombinant	3	8

The observed number of parental chromosomes is the sum of the number of normal males and the number of males expressing both traits. The observed number of recombinant chromosomes is the sum of the number of males expressing only trait 1 and the number of males expressing only trait 2. Based on our hypothesis, if both traits are segregating randomly, then parental and recombinant chromosomes would be expected to arise with equal frequency.

$$\text{Thus, } \chi^2 = \sum_{df=1} \frac{(O-E)^2}{E} = \frac{5^2}{8} + \frac{5^2}{8} = 6.25$$

Using the Chi Square table with $df=1$, $0.025 > p > 0.01$. Therefore, an expectation that the two traits are unlinked differs significantly from the observed data.

An alternative method for solving this problem would be to consider each chromosome or genotype type as a different class. While this method is acceptable, it is less robust than grouping by parental and recombinant. This is because as the number of classes increases, the class sizes decrease and the degrees of freedom increase. With the Chi Square test it is important to have as large a data set for each group as possible. As shown below, this method results in not being able to reject the hypothesis.

Male X Genotype	Observed	Expected
X	6	4
X ¹	2	4
X ²	1	4
X ¹²	7	4

$$\chi^2 = \sum \frac{(O-E)^2}{E} = \frac{2^2}{4} + \frac{2^2}{4} + \frac{3^2}{4} + \frac{3^2}{4} = 6.5$$

$$df = 3$$

In this case, using the Chi Square table with $df=3$, $0.1 > p > 0.05$. Therefore, an expectation that the two traits are unlinked does not differ significantly from the observed data.

2c. Number of recombinants = number of non-parental chromosomes = 3.

$$m.u. = 100 \left(\frac{3}{16} \right) = 18.75$$

The map distance between the genes for trait 1 and 2 is approximately 18.75 cM.

3a. Recessive. A=wt a=mutant.

Person	Genotype
1	aa
2	Aa
3	aa
4	Aa
5	aa
6	Aa
7	As
8	aa

Cross	P(aa)
2x5	1/2
6x4	1/4
7x4	1/4
3x8	1

3b. Dominant. A = wt A^{dom} = dominant mutant

Person	Genotype
1	$A^{dom}A$
2	AA
3	$A^{dom}A$
4	AA
5	$A^{dom}A$
6	AA
7	AA
8	$A^{dom}A$

Cross	P($A^{dom}A$ or $A^{dom}A^{dom}$)
2x5	1/2
6x4	0
7x4	0
3x8	3/4

3c. X-linked. X^A = wild type X^a = mutant.

Note that the X-linked trait must be recessive. If the trait were X-linked dominant then all progeny of individual #5 would show the trait.

Person	Genotype
1	X^aY
2	X^aX^A
3	X^aX^a
4	X^AY
5	X^aY
6	X^aX^A
7	X^aX^A
8	X^aY

Cross	Male	Female
	P(X^aY)	P(X^aX^a)
2x5	1/2	1/2
6x4	1/2	1/2
7x4	1/2	0
3x8	1	1

7.03 Problem Set 2

Due before 5 PM on Thursday, September 28

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

1. You have been studying eye color mutations in *Drosophila*, which normally have red eyes. White (**w**) is a recessive mutation on the X chromosome that gives flies with white eyes. You have isolated a new recessive mutation that causes apricot colored eyes, designated **ap**.

(a) A female from a true breeding **ap** strain is crossed to a male from a true breeding **w** strain. The females in the F₁ progeny have pale-apricot colored eyes. Explain what this result tells you about the relationship between the **w** and **ap** mutations and why.

(b) What would you expect the phenotype of the male F₁ progeny to be and why?

(c) F₁ females are crossed to wild type males and the male progeny are examined. Most of these males have either white eyes or apricot eyes, however, 0.1% of the males have red eyes. What is the origin of these red eyed males?

(d) What is the distance between the **w** and **ap** mutations in cM?

(e) You suspect that male flies that carry both the **w** and **ap** mutations will have white eyes. You decide to search for such a double mutant from among 500 white-eyed male flies produced in the cross described above. Describe a genetic test that will allow you to distinguish double mutants from **w** single mutant flies and describe how **w** single mutants and **w, ap** double mutants will behave in this test. (Hint; you will need to do a lot of crosses and score a lot of progeny to find the double mutant you are looking for).

(f) Crossveinless (**cv**) is a recessive mutation that maps about 10 cM away from **w**. A female from a true breeding **cv, w** strain is crossed to a male from a true breeding **ap** strain. The females from this cross are then crossed to wild type males and the resulting male progeny are examined. Eight males with red eyes are found and seven of these have normal wings and one has crossveinless wings. Draw a map showing the relative positions of the **cv, w, and ap** mutations and the approximate distances between them.

2. You are studying serine biosynthesis in yeast and you know that three different genes that are required (Ser1, Ser2, and Ser3) in the sense that a strain with a mutation in any one of these genes will not grow unless serine is provided in the medium. You have isolated a collection of new Ser⁻ mutants and all but one can be placed in one of the three Ser genes by complementation tests. This last mutation, designated SerX, is dominant and can't be analyzed by complementation testing. Therefore you decide to cross SerX to a representative recessive mutation in each of the three Ser genes. Three types of tetrads can be produced from these crosses.

<u>Type 1</u>	<u>Type 2</u>	<u>Type 3</u>
2 Ser ⁻ : 2 Ser ⁺	3 Ser ⁻ : 1 Ser ⁺	4 Ser ⁻

(a) In the cross of SerX to Ser1, 4 tetrads are of Type 1, 15 are of Type 2, and 6 are of Type 3. What does this result tell you about the relationship between Ser1 and SerX?

(b) In the cross of SerX to Ser2, 25 tetrads are examined and all are of Type 3. What is the upper limit of the distance between the SerX and Ser2 mutations? Given that the average yeast gene is about 2 kbp in length and the recombination rate in yeast is about 2 kbp/cM is SerX likely to be an allele of Ser2? Why or why not?

(c) In the cross of SerX to Ser3, 20 tetrads are of Type 3 and 5 tetrads are of Type 2. What is the distance between the SerX and Ser3 mutations?

(d) You would like to isolate a SerX, Ser3 double mutant. Describe the genetic test(s) that you would perform on the three Ser⁻ spore clones from one of the Type 2 tetrads described in **(c)** above that you would use to identify the double mutant. Be as specific as possible and describe the expected results of the tests.

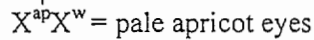
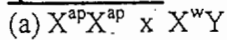
3. Consider a phage gene that encodes the enzyme lysozyme. The lysozyme protein has a molecular mass of 60 kDa (kilodaltons). You have isolated a small collection of mutants in the lysozyme gene that fail to produce a functional lysozyme enzyme. Taking advantage of the fact that you can detect the lysozyme protein from an extract of phage infected cells, you determine the size of the lysozyme proteins expressed by each mutant.

Wild type	60 kDa
Mutant 1	60 kDa
Mutant 2	15 kDa
Mutant 3	20 kDa
Mutant 4	50 kDa

- (a)** In genetic crosses, you find that Mutant 1 and Mutant 2 lie very close to one another. List the types of nucleotide sequence changes that could have produced Mutant 1, and the types of changes that could have produced Mutant 2.
- (b)** When Mutant 2 is crossed to Mutant 3, only 2 out of 1000 of the progeny produce wild type lysozyme. What frequency of wild type progeny would you expect from a cross of Mutant 3 with Mutant 4?
- (c)** Frameshift mutations can be classified as follows: addition of one base = +1, deletion of one base = -1, addition of two bases = +2, deletion of two bases = -2 etc. Within the lysozyme gene you find a +1 frameshift mutation that can be combined with a nearby -1 frameshift to produce a double mutant that makes functional lysozyme. But when a +2 frameshift and a -2 frameshift at the same positions as the +1 and -1 frameshifts are combined, you find that functional lysozyme is not produced — explain.
- (d)** Assume that Mutant 2 is a +1 frameshift and that Mutant 3 is a -1 frameshift. Without knowing anything else about the sequence of the lysozyme gene, calculate the probability that full-length lysozyme will be produced when Mutant 2 and Mutant 3 are combined as a double mutant. (For your calculation assume the average mass of an amino acid to be 110 daltons).

Problem Set 2 -- SOLUTIONS

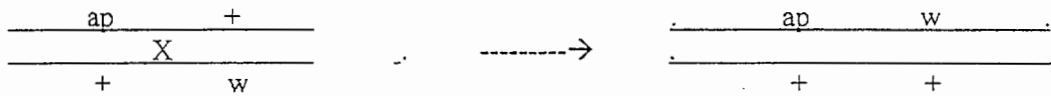
Question #1:



The w and ap mutations do not complement. This means they must be alleles of the **same gene**. The pale apricot phenotype is probably the result of the ap mutation being a partial loss of function and the w mutation being a null mutation.

(b) The phenotype of the male F_1 progeny will be **apricot eyes**; all males will be $X^{ap}Y$.

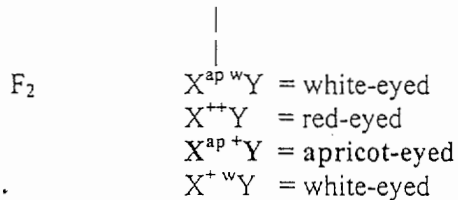
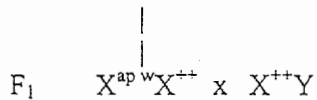
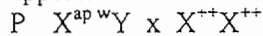
(c) The red-eyed males are a **recombinant class**; they are the result of a cross-over between the ap and w mutations in the F_1 female (even though the mutations are in the same gene, you can still get a crossover between the two mutant sites).



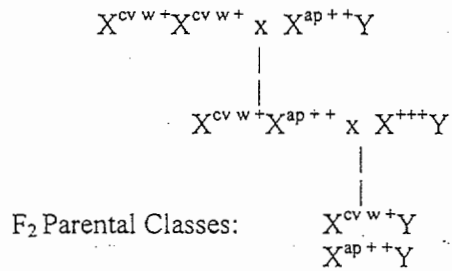
The red-eyed males could also be the result of non-disjunction in the meiosis stage of the F_1 female. This would result in male progeny of genotype X^{++} (X chromosome inherited from the WT father). This occurs $1/1700 = 0.03\%$ of the time.

(d) Distance between w and ap mutations = $\frac{\text{total \# of recombinant gametes}}{\text{Total \# of gametes}} \times 100\%$
 = % of recombinants
 = $2 \times (0.1\% - 0.03\%)$
 = $0.14\% = 0.14\text{cM}$

- (e) -Cross males (progeny to be tested for double mutants) to wild type females.
 -Cross F_1 females to wild type males.
 -Examine F_2 male progeny -- look for **apricot-eyed F_2 male**. Male parent was a double mutant.
 -If the parental male was a single w mutant, only red and white-eyed F_2 males would appear.



(f) Parents are cv, w true-breeding and ap true-breeding. Therefore, parental classes in F1 will be cross-veinless and white-eyed, or normal winged and apricot eyed:



- The data we are given is for red-eyed F2 males -- these must be recombinant classes. The rare crossveinless, red-eyed male ($X^{cv^{++}}Y$) is probably the result of a double cross-over. The 7 red-eyed, normal winged flies ($X^{+++}Y$) are probably the result of a single cross-over between w and ap.

Possible Gene Orders	w	cv	ap	cv	w	ap	cv	ap	w
Parental Chromosomes	w	cv	+	cv	w	+	cv	+	w
	+	+	ap	+	+	ap	+	ap	+
Genotypes after single cross-over	w	cv and +	ap	cv	w and +	ap	cv	+	+
	+	+	+	+	+	+	+	ap	w
Genotypes after double cross-over	w	+	+	cv	+	+	cv	ap	w
	+	cv	ap	+	w	ap	+	+	+
Same as given recombinant classes		NO		YES			NO		

-Therefore, final genetic map:



Question #2:

(a) $serX SER1^+ \times SERX^+ ser1^-$

PD: $serX SER1^+$
 $serX ser1^-$
 $SERX^+ ser1^-$
 $SERX^+ SER1^+$

Ser⁻:Ser⁺ = 4:0 \implies Type 3

TT: $serX SER1^+$
 $serX ser1^-$
 $SERX^+ SER1^+$
 $SERX^+ ser1^-$

Ser⁻:Ser⁺ = 3:1 \implies Type 2

NPD: $serX ser1^-$
 $serX ser1^-$
 $SERX^+ SER1^+$
 $SERX^+ SER1^+$

Ser⁻:Ser⁺ = 2:2 \implies Type 1

-Given: 4 NPD : 15 TT : 6 PD \approx 1 : 4 : 1

Therefore, *SERX* and *SER1* are **unlinked**.

(b) If the genes were 1cM apart, you would predict that out of 50 tetrads = 200 spores, an average of 2 spores would be recombinant (1%). Thus, it is likely that the SerX and Ser2 mutations are less than 1cM apart. Since 1cM (2kbp/cM) = 2kbp, it is likely that SerX and Ser2 are less than 2kbp apart and are, therefore, mutations in the same gene.

(c) 20 PD : 5 TT, therefore, PD \gggg NPD \rightarrow the two genes are **linked**.

$$\text{Distance (cM)} = \frac{100(T + 6NPD)}{2\Sigma} = \frac{5}{50}(100) = 10\text{cM}$$

(d) The spore genotypes from a type 2 tetrad are:

$serX SER3^+$	Ser- spores	$serX SER3^+$
$serX ser3^-$	=====>	$serX ser3^-$
$SERX^+ SER3^+$		
$SERX^+ ser3^-$		$SERX^+ ser3^-$

Test: Cross the Ser⁻ spore clones to a wild type haploid of opposite mating type. Sporulate and dissect tetrads. If the Ser⁻ spore clone was a single mutant in either *serX* or *ser3* then the Ser⁻ phenotype will always aggregate 2:2 in the tetrads. For the double mutant, one out of every five tetrads should segregate 3Ser⁻:1Ser⁺

7.03 Problem Set 3

Due before 5 PM on Thursday, October 19

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

1. Consider a phage gene that encodes the enzyme lysozyme. The lysozyme protein has a molecular mass of 60 kDa (kilodaltons). You have isolated a small collection of mutants in the lysozyme gene that fail to produce a functional lysozyme enzyme. Taking advantage of the fact that you can detect the lysozyme protein from an extract of phage infected cells, you determine the size of the lysozyme proteins expressed by each mutant.

Wild type	60 kDa
Mutant 1	60 kDa
Mutant 2	15 kDa
Mutant 3	20 kDa
Mutant 4	50 kDa

(a) In genetic crosses, you find that Mutant 1 and Mutant 2 lie very close to one another. List the types of nucleotide sequence changes that could have produced Mutant 1, and the types of changes that could have produced Mutant 2.

(b) When Mutant 2 is crossed to Mutant 3, only 2 out of 1000 of the progeny phage make wild type lysozyme. What frequency of wild type progeny would you expect from a cross of Mutant 3 with Mutant 4?

(c) Frameshift mutations can be classified as follows: addition of one base = +1, deletion of one base = -1, addition of two bases = +2, deletion of two bases = -2 etc. Within the lysozyme gene you find a +1 frameshift mutation that can be combined with a nearby -1 frameshift to produce a double mutant that makes functional lysozyme. But when a +2 frameshift and a -2 frameshift at the same positions as the +1 and -1 frameshifts are combined, you find that functional lysozyme is not produced — explain.

(d) Assume that Mutant 2 is a +1 frameshift and that Mutant 3 is a -1 frameshift. Without knowing anything else about the sequence of the lysozyme gene, calculate the probability that full-length lysozyme will be produced when Mutant 2 and Mutant 3 are combined as a double mutant. (For your calculation assume the average mass of an amino acid to be 110 daltons).

2. (a) Write out the RNA sequence of the anticodon segment of tRNA^{trp}. Be sure to note the 5' and 3' ends. Also remember that in RNA, U (uracil) takes the place of T (thymine).

(b) Write out the DNA base pairs that encode the anticodon segment of tRNA^{trp}. Show both DNA strands with the 5' and 3' ends labeled.

(c) Hydroxylamine will deaminate cytosine in DNA to produce uracil. On replication, uracil can base-pair with adenine producing a net change in the DNA base sequence. Which, if any, of the three nonsense codons could in principle be suppressed by a single hydroxylamine-generated mutation in tRNA^{trp}. For your answer, write out the RNA sequence of the anticodon segment of each suppressor mutant. Also write out the corresponding double stranded DNA sequence of the tRNA^{trp} suppressor genes

3. You have used EMS mutagenesis to isolate four different *E. coli* mutants that will not grow unless histidine is provided in the growth medium. You label these mutants **his1⁻**, **his2⁻**, **his3⁻**, and **his4⁻**.

(a) In order to test for linkage between **his1⁻** and the other **his⁻** mutants, you set out to isolate a **Tn5** insertion linked to the **his1⁻** mutant. 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 kanamycin resistant (Kan^r) and **his⁺**). You grow **P1** phage on a mixture of the entire collection of **Tn5** insertion strains and then infect the **his1⁻** mutant and select for Kan^r transductants. Most of the Kan^r transductants are **his⁻**, but one out of 1000 is **his⁺**.

Explain how this **his⁺** transductant arose.

(b) Next you grow **P1** phage on the **his⁺** transductant isolated above and infect the original **his1⁻** mutant with the resulting phage. After selecting for Kan^r transductants you test these transductants for their ability to grow in the absence of histidine. You find that among 100 Kan^r transductants 20 are **his⁻** and 80 are **his⁺**.

Give the distance between the **Tn5** insertion and **his1**, expressed as a cotransduction frequency.

(c) The same **P1** phage preparation generated in part **(b)** above is used to infect either a **his2⁻** or a **his3⁻** mutant and Kan^r transductants are isolated. For the infections of the **his2⁻** and the **his3⁻** mutants, none of the Kan^r transductants are **his⁺** (you examine hundreds of

Kan^r transductants from each transduction experiment). What does this tell you about the relationship between the **his1⁻** mutation and the **his2⁻** and **his3⁻** mutations, and why?

(d) How would you determine whether the **his2⁻** and **his3⁻** mutations are likely to be alleles of the same gene?

(e) The **P1** phage preparation generated in part **(b)** is used to infect a **his4⁻** mutant and Kan^r transductants are isolated. Among 100 Kan^r transductants examined, 19 are **his⁻** and 81 are **his⁺**. What does this tell you about the relationship between the **his1⁻** and the **his4⁻** mutations, and why?

(f) Using the procedure outlined above, you construct a strain that has both the **Tn5** insertion and **his1⁻** and another strain that has both the **Tn5** insertion and **his4⁻**. Using these strains you perform two reciprocal crosses. In the first cross, **P1** is grown on the **Tn5 his1⁻** strain and the resulting phage are used to infect a **his4⁻** strain. In this transduction experiment, 10 out of 500 Kan^r transductants are **his⁺**. In the reciprocal cross, **P1** is grown on the **Tn5 his4⁻** strain and the resulting phage are used to infect a **his1⁻** strain. In this experiment, 1 out of 500 Kan^r transductants are **his⁺**.

Draw a map showing the relative order of the **Tn5** insertion, **his1⁻** and **his4⁻**.

(g) Clearly the **his1⁻** and **his4⁻** mutations are close together. This could mean that these mutations are different alleles of the same gene or it could mean that they are alleles of two different genes that are close to one another. Describe in general terms how you might distinguish these two possibilities.

Solutions for Problem Set #3

Problem 1

- a. Mutant 1 must be a missense mutation because a full-length protein is produced. Mutant 2 could be either a nonsense mutation or a frameshift mutation because a truncated protein is produced.
- b. The size difference between mutants 2 and 3 is 5 kDa. The size difference between mutants 3 and 4 is 30 kDa. This means that, in the DNA, the distance between mutants 3 and 4 is about 6-fold greater than the distance between mutants 2 and 3. Thus, the frequency of wild type lysozyme in a cross of 3 with 4 will be 6-fold greater than in a cross of 2 and 3. The answer is $12/1000 = 0.012$.
- c. There are two possibilities here. First, you must recognize that the reading frame caused by a +1 frameshift will be different than the reading frame caused by a +2 frameshift.

Case #1. There is a stop codon between the two frameshift mutations in the +2 reading frame but not in the +1 reading frame.

Case #2. The amino acid changes between the +1 and -1 frameshifts can be tolerated to yield a functional protein. The amino acid changes between the +2 and -2 frameshifts lead to production of a full length, but non-functional protein product.

- d. Since both mutations are frameshifts, we know that mutant #2 must occur before codon specifying the amino acid present at the 15 kDa position of the protein. Likewise, mutant #3 must occur before the 20 kDa position. We also know that there is a stop codon at the 15 kDa position in the +1 reading frame and there is a stop codon present at the 20 kDa position in the -1 reading frame. In order for the 2, 3 double mutant to make a full length protein it must be the case that mutant 3 also occurs before the 15 kDa position and restores the proper reading frame (otherwise the protein would terminate at the stop codon present in the +1 reading frame at 15 kDa). So this problem really boils down to calculating the probability that mutant 3 could occur before the 15 kDa position and still produce a 20 kDa product. In other

words, what is the probability that over a 5 kDa length of protein no stop codon will occur at random in the DNA coding for that 5 kDa region?

To answer this question, we must first calculate how many codons correspond to 5 kDa of protein. Given that each amino acid is 0.110 kDa \Rightarrow

$$\# \text{ of codons} = \frac{5}{0.110} = 45.5$$

Therefore, the probability that no stop codon will occur in 46 random codons is

$$p(\text{no stop}) = \left(\frac{61}{64}\right)^{46} = 0.11$$

Problem 2

a. 5'-CCA-3'

b. 5'-CCA-3'
3'-GGT-5'

c. Deamination of cytosine to uracil will result in a net change of C \rightarrow T in the DNA and C \rightarrow U in the RNA. Since there are two C's in the tRNA^{trp} anticodon, we must consider each case:

5'-UCA-3' This will suppress a UGA nonsense codon. The DNA sequence is:

5'-TCA-3'
3'-AGT-5'

5'-CUA-3' This will suppress a UAG nonsense codon. The DNA sequence is:

5'-CTA-3'
3'-GAT-5'

Problem 3

- a. Some members of the Tn5 insertion collection have transposons that are closely linked to the wild type *HIS1* gene. Occasionally one of the phage will accidentally package a portion of the bacterial chromosome that contains both the Tn5 insertion and the wild type *HIS1* gene. When this phage then infects a *his1*⁻ mutant, His⁺ bacteria result from the following recombination event:

$$\begin{array}{l} \text{Transductant DNA} \\ \text{E.coli Chromosome} \end{array} \quad \frac{\begin{array}{cc} \text{HIS1}^+ & \text{Tn5} \\ \hline \text{X} & \text{X} \end{array}}{\text{his1}^-}$$

- b. The Kan^r His⁺ transductants are the result of the cotransduction of *HIS1* with Tn5. Therefore, the cotransduction frequency of Tn5 and *HIS1* is:

$$\begin{aligned} \text{cotransduction frequency} &= \frac{\# \text{ His}^+ \text{ transductants}}{\text{Total \# transductants}} \\ &= \frac{80}{100} \\ &= \mathbf{0.8} \end{aligned}$$

- c. The absence of His⁺ transductants means that *HIS2*⁺ and *HIS3*⁺ were not cotransduced with Tn5 and *HIS1*. Therefore, *HIS2* and *HIS3* must be more than 10⁵ bp away from Tn5 and *HIS1*.
- d. There are several possible ways to approach this problem. You could clone *his2* by generating a genomic library and isolating a complementing clone. You could then test this clone for the ability to complement a *his3* mutation and sequence the clone to determine which genes are present. An alternative method would be to isolate a strain carrying a Tn linked to the *his2* gene. You would then grow P1 on this strain and infect each single mutant. If *his2* and *his3* are allelic then the cotransduction frequencies should be identical (or very similar).

e. All we can really say is that *HIS1* and *HIS4* are within 10^5 bp of Tn5. If both genes are on the same side of Tn5 then they are very close together because the cotransduction frequency of *HIS4* and Tn5 is almost exactly the same as that for *HIS1* and Tn5. It is also possible that Tn5 is located between *HIS1* and *HIS4* and is approximately the same distance from each gene.

f.

Cross	Order A	Order B	Order C
#1: Donor: <i>his1⁻ HIS4⁺ Tn5</i> Recipient: <i>HIS1⁺ his4⁻</i>	$\frac{1^- \quad 4^+ \quad \text{Tn5}}{\text{X} \quad \quad \quad \text{X}}$ 1 ⁺ 4 ⁻ His ⁺ with 2 cross-overs	$\frac{4^+ \quad 1^- \quad \text{Tn5}}{\text{X} \quad \text{X} \quad \text{X} \quad \text{X}}$ 4 ⁻ 1 ⁺ His ⁺ with 4 cross-overs	$\frac{1^- \quad \text{Tn5} \quad 4^+}{\text{X} \quad \quad \quad \text{X}}$ 1 ⁺ 4 ⁻ His ⁺ with 2 cross-overs
#2: Donor: <i>HIS1⁺ his4⁻ Tn5</i> Recipient: <i>his1⁻ HIS4⁺</i>	$\frac{1^+ \quad 4^- \quad \text{Tn5}}{\text{X} \quad \text{X} \quad \text{X} \quad \text{X}}$ 1 ⁻ 4 ⁺ His ⁺ with 4 cross-overs	$\frac{4^- \quad 1^+ \quad \text{Tn5}}{\text{X} \quad \quad \quad \text{X}}$ 4 ⁺ 1 ⁻ His ⁺ with 2 cross-overs	$\frac{1^+ \quad \text{Tn5} \quad 4^-}{\text{X} \quad \quad \quad \text{X}}$ 1 ⁻ 4 ⁺ His ⁺ with 2 cross-overs

We are given in the problem that the number of Kan^r His⁺ transductants are greater from cross #1 than cross #2. Therefore, His⁺ must be the result of 2 cross-overs in cross#1 and 4 cross-overs in cross#2.

Therefore, the correct order is Order A:

HIS1----*HIS4*-----Tn5

g. As in part d, you could clone *his1* using a genomic library. You could then determine whether the clone also complements *his4*. If not, the mutations are in different genes. If the clone complements both *his1* and *his4* you would have to sequence the insert to determine how many genes are present on the plasmid. If only a single gene is present then they are allelic. If multiple genes are present, you could subclone each gene individually and test for the ability to complement each mutant.

7.03 Problem Set 4

Due before 5 PM on Thursday, October 26

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

1. A colleague gives you an *E. coli* strain that she thinks has a Tn5 insertion linked to the Lac operon. The strain she gives you is Lac⁺ and Kan^r.

(a) You decide to test for linkage to the Lac operon by growing P1 on her strain and then using this phage lysate to infect a LacI^S mutant strain that you have in your strain collection. You select Kan^r transductants and then note their Lac phenotypes. Among 100 transductants, 70 have normal regulated Lac gene expression and 30 have uninducible expression. What is the distance between the Tn5 insertion and the LacI^S mutation, expressed as a cotransduction frequency?

(b) Knowing the distance between the Tn5 insertion and the Lac operon, you now want to determine on which side of the Lac operon the insertion resides. To do this you set up a three-factor transduction cross. You grow P1 on one of the uninducible Kan^r transductants isolated in part **a**. First, the resulting phage are used to infect a wild type strain and out of 1000 Kan^r transductants 700 are uninducible and 300 are regulated like wild type. Next, you use the same phage lysate to infect a LacO^C mutant that you have in your collection. Out of 1000 Kan^r transductants from this cross, 648 are uninducible, 350 are constitutive, and 2 are regulated like wild type. What is the most likely relative relationship between the Tn5 insertion and the LacO^C and LacI^S mutations?

(c) Suppose that you had set up the three factor cross differently such that P1 was grown on a strain that carried both the Tn5 insertion and the LacO^C mutation, and that this phage was used to infect a LacI^S mutant strain. From 1000 Kan^r transductants recovered from this cross how many transductants of each type would be expected?

(d) Let's say that you isolate a new mutation that gives uninducible Lac expression. You use P1 phage grown on the Lac⁺ and Kan^r strain that your colleague gave you to infect the new Lac⁻ mutant. All of the 100 Kan^r transductants that you examine are uninducible. What does this result tell you about the new mutation?

(e) Present two different molecular mechanisms to describe the new Lac⁻ mutant. Also state in general terms the type of experiment that would allow you to distinguish between these models.

2. Say that you are studying the ability of a bacterial strain to use urea as a nitrogen source. You have identified the structural gene for urease which you designate UreA. You find that normally urease is not expressed, but that urease is induced when urea is present in the growth medium.

(a) You isolate a mutant that gives constitutive expression of urease that you designate ure1. Through the use of cotransduction experiments with a transposon linked to UreA you find that ure1 is closely linked to UreA. Propose two different molecular models to explain the ure1 mutation.

(b) You construct a plasmid that contains the wild type UreA gene and surrounding chromosomal sequences (assume the chromosomal segment on the plasmid includes the wild type version of the region where the ure1 mutation resides). When the plasmid is introduced into a ure1 mutant the resulting merodiploid expresses urease constitutively. Propose two models for the mechanism of the ure1 mutation that are consistent with this new finding.

(c) Next you construct a double mutant that contains both a ure1 mutation and a ureA⁻ mutation (this strain does not express urease). When the plasmid described above is introduced into the double mutant strain you find that the resulting merodiploid only expresses urease when urea is present in the medium. Which of your models for the ure1 mutation is consistent with this observation.

(d) Using transposon mutagenesis you isolate a second mutation that is constitutive for urease expression. The site of transposon insertion is unlinked to the UreA gene. Bearing in mind that transposon insertions usually inactivate their target gene, propose a molecular mechanism to explain the behavior of the ure2 mutation.

(e) You isolate a new mutation that gives uninducible urease expression, which you call ure3. The ure3 mutation is unlinked to UreA. You construct a ure1 ure3 double mutant and a ure2 ure3 double mutant and find that both strains express urease constitutively. Assume that the ure2 and ure3 mutations are unlinked. Propose an explicit model for UreA regulation that takes into account all of the properties of the ure1, ure2 and ure3 mutations. Your model should include a role for urea in controlling urease regulation.

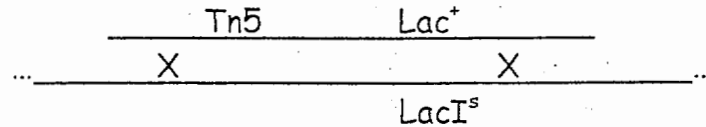
(f) Now assume that the ure3 mutation is very tightly linked to the ure2 transposon insertion. Propose a new model to explain the behavior of the ure1, ure2 and ure3 mutations that is different from the model in part e. This model should also include a role for urea in controlling urease regulation.

7.03 Problem Set 4 Solutions

Fall 2000

Problem 1:

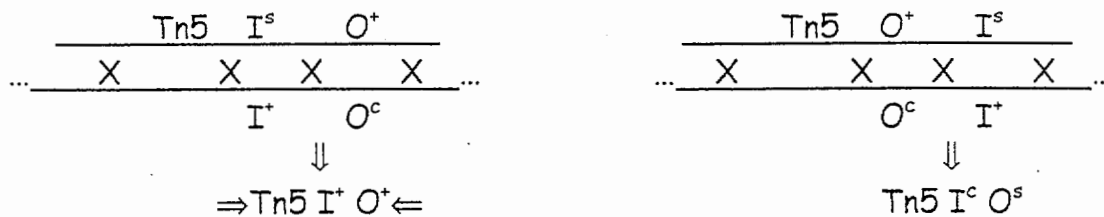
- a) The transductants that have normal regulated Lac gene expression are result of Tn5 and Lac⁺ cotransduction. Therefore, the c.f. = 70/100 = 70%.



- b) Here are four possible genotypes with their phenotypes:

A	Tn5 I ^s O ⁺	uninducible	648
B	Tn5 I ^s O ^c	constitutive	52
C	Tn5 I ⁺ O ^c	constitutive	298
D	Tn5 I ⁺ O ⁺	inducible (wildtype)	2

We are given that A = 648, B + C = 350, and D = 2. We know that the c.f. between Tn5 and I is 70%, so we expect A + B = 700. From this we can do some algebra and say that B = 52 and C = 298. Therefore the rarest class is D, which is the result of four crossovers, giving us that LacI is between LacO and Tn5.



- c) In this experiment, the donor/host genotype is different, however the gene order/distance is the same, so we expect the same frequency for each type of transduction.

A'	Tn5 I ⁺ O ^s	constitutive	648
B'	Tn5 I ⁺ O ⁺	inducible (wildtype)	52
C'	Tn5 I ^s O ⁺	uninducible	298
D'	Tn5 I ^s O ^c	constitutive	2

Therefore, we expect 298 uninducible, 650 constitutive and 52 inducible transductants.

- d) Since Lac^+ never cotransduced with Tn5, we can say that the distance between the mutation and Tn5 is more than one phage head length (10kb). So the new mutation is unlinked and likely to be trans-acting.
- e) A possible reason for the uninducible expression is over-expression of LacI. If there are more repressors present than inducers, there will always be some repressor that can bind to the operator to prevent transcription.

Here are two possible mechanisms:

- 1) LacI is positively regulated (as in Mal operon), and activator^s mutation is causing constitutive (over-expression) of LacI
- 2) LacI is negatively regulated (as in Lac operon), and repressor^r mutation is causing constitutive (over-expression) of LacI.

To test between these two mechanisms, we can transform a plasmid containing a wildtype version of Lac^- . Since activator^s is dominant, and repressor^r is recessive, if the resulting merodiploid is uninducible we take Model 1, otherwise Model 2.

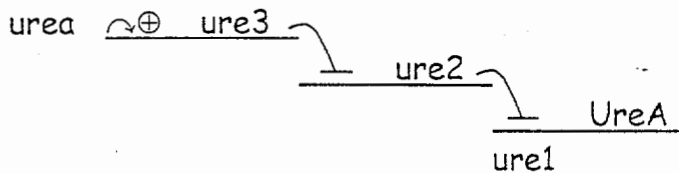
Problem 2:

- a) $ure1$ could be an operator mutation which prevents repressor binding (O^c). It could also be a repressor mutation (R^- or R^{-d}), or a super activator (A^s)
- b) This shows that $ure1$ is a dominant mutation. It could still be all of the above with the exception of R^- .
- c) This shows that $ure1$ needs to be cis to show the constitutively active phenotype. Therefore it is likely to be a mutation in the Operator/Promoter (O^c).
- d) $ure2$ is a recessive and constitutive mutation, so it is most likely a mutation in a repressor of UreA transcription (R^-). Inactivation is usually associated with recessive mutations.

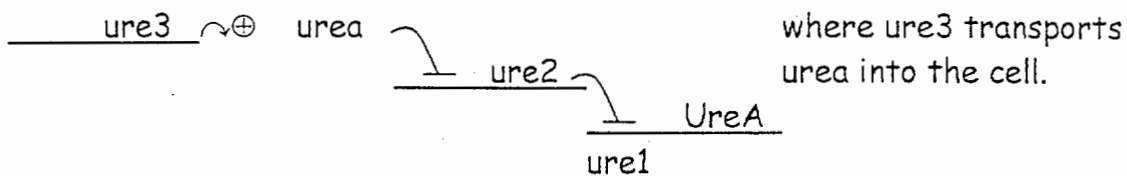
e) ure3 is uninducible, therefore wildtype ure3 activates transcription of UreA. The two possible mechanisms are:

urea → ure3 —| ure2 —| ure1/UreA ure2⁻/ure3⁻=constitutive
 urea → ure2 —| ure3 → ure1/UreA ure2⁻/ure3⁻=uninducible

The top model is consistent with the epistasis test, where ure2 is epistatic to ure3.

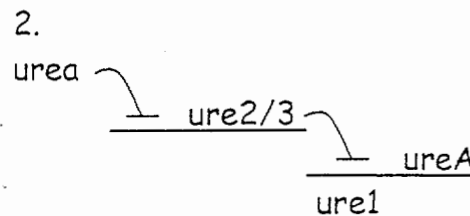
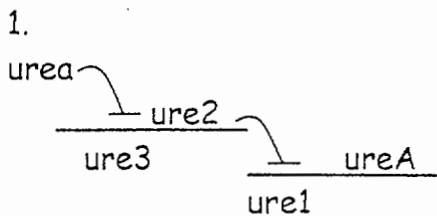


Also:



f) ure2 is tightly linked to ure3, so they are likely in the same gene. There are two possible molecular mechanisms suggested by this information:

- 1- ure3 is a mutation in the ure2 operator which destroys the urea response element, therefore rendering it constitutively active, always making the ure2 transcript.
- 2- ure3 is a mutation which makes a superrepressor (R^s), while ure2 kills the protein (R⁻), which makes them different alleles of the same gene



7.03 Problem Set 5

Due before 5 PM on Thursday, November 16, 2000

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

1. You are studying the regulation of an enzyme in yeast and to begin your analysis of regulation you first fuse the promoter region of the enzyme to the LacZ coding sequence and then place this hybrid gene on an appropriate yeast plasmid. You are relieved to find that cells carrying the hybrid gene do not express β -galactosidase activity unless the known inducer for the enzyme synthesis is present, meaning that the promoter region you have selected contains all of the necessary cis-acting sequences for normal regulation. You next identify two different mutants that show abnormal regulation of your reporter. Mut1⁻ gives constitutive expression whereas Mut2⁻ shows uninducible expression.

(a) Are Mut1⁻ or Mut2⁻ cis or trans-acting? Explain.

Next you cross a strain that carries Mut1⁻ to a strain that carries Mut2⁻. Three different tetrad types are obtained:

<u>Type 1</u>	<u>Type 2</u>	<u>Type 3</u>
regulated (wt)	regulated (wt) constitutive	
regulated (wt)	constitutive	constitutive
constitutive	constitutive	uninducible
constitutive	uninducible	uninducible

Tetrad Type 2 is the most abundant and Type 1 and Type 3 occur at roughly equal frequency.

(b) Are the Mut1⁻ and Mut2⁻ mutations linked? What is the phenotype of a Mut1⁻ Mut2⁻ double mutant? Produce a model to explain the regulation of the enzyme that is consistent with all of the data you have.

In the examples of gene expression that we have covered in class the genes were regulated in some way or another. Many eukaryotic genes are expressed constitutively regardless of environmental conditions; nevertheless transcription of these constitutive genes depends on the same kind of transcriptional activator proteins that are employed in regulated transcription. The difference being that for regulated genes at least one activator (or repressor) must itself be regulated, whereas for constitutive genes the activator(s) are always active. In the next parts of this problem we will analyze the expression of a hypothetical constitutive gene in yeast.

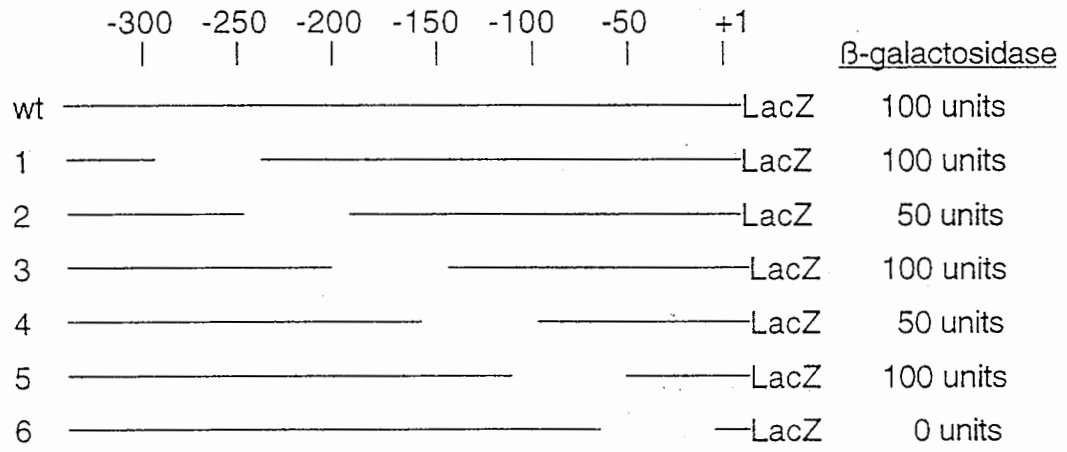
As before, the first step in your analysis is to fuse the promoter region of the gene to the LacZ coding sequence and to place the hybrid gene on a yeast plasmid. Say that cells carrying the hybrid gene express 100 units of β -galactosidase activity under all conditions that you test. You next identify two different mutations that show decreased β -galactosidase activity: either Mut1⁻ or Mut2⁻ express about 50 units of β -galactosidase activity.

When you cross a strain that carries Mut1⁻ to a strain that carries Mut2⁻ you find three different tetrad types that are distinguishable by the amount of β -galactosidase activity that each spore clone expresses. Tetrad Type 2 is the most abundant and Type 1 and Type 3 occur at roughly equal frequency.

<u>Type 1</u>	<u>Type 2</u>	<u>Type3</u>
100 units	100 units	50 units
100 units	50 units	50 units
none	50 units	50 units
none	none	50 units

(c) Are the Mut1⁻ and Mut2⁻ mutations linked? What is the phenotype of a Mut1⁻ Mut2⁻ double mutant? Produce a model to explain the involvement of Mut1⁻ and Mut2⁻ in the expression of the gene that is consistent with all of the data you have.

Next you evaluate the promoter sequences necessary for expression of the gene. The figure below shows the effect of different 50 bp deletions in the promoter region on the amount of β -galactosidase activity expressed by the reporter gene.



When deletion 2 is placed in the Mut1⁻ strain no β -galactosidase activity is expressed, whereas deletion 2 in the Mut2⁻ strain expresses 50 units of β -galactosidase activity.

Conversely, deletion 4 in the Mut1⁻ strain expresses 50 units of β -galactosidase activity, whereas deletion 4 in the Mut2⁻ strain expresses no β -galactosidase activity.

(e) Based on this new information, refine your model to account for how the Mut1 and Mut2 gene products interact with the promoter sequences.

2. Human genes vary greatly in size, ranging from less than 1 kb to more than 2 Mb (as measured from the 5' end of the first exon to the 3' end of the last exon; the promoter is, of course, an essential additional component). Assuming an average gene size of 30 kb, and an average mRNA size of 1.5 kb, calculate:

(a) The percentage of the human genome that is transcribed.

(b) The percentage of the human genome that is represented in spliced mRNAs.

(c) The percentage of the human genome that is accounted for by introns.

In human genes, exons are typically about 200 bp each. You obtain the complete nucleotide sequence of a 4 kb mRNA from a newly discovered human gene: gene XYZ. You strongly suspect that gene XYZ is unusually long, spanning about 500 kb. Your colleague has given you a collection of 10 human genomic BACs that are likely, as a set, to encompass the entirety of the XYZ gene, but unfortunately there is as yet no nucleotide sequence available for any of these BACs, and you do not know their physical order and the overlaps among them.

(d) Describe how you could use the mRNA sequence and the BACs to construct a physical map of the region containing the XYZ gene (without doing any sequencing of the BACs themselves). Provide a drawing of the resulting physical map, including the XYZ mRNA and the locations of the 5' end, 3' end, and promoter of the XYZ gene:

(e) Why is it riskier or trickier to select PCR primers (to define STSs useful in physical mapping) using mRNA rather than genomic DNA as the source of sequence information?

3. In mammals, including humans and mice, growth hormone (a protein) is speculated to play a prominent role in determining adult size. You decide to test this hypothesis in mice using transgenic methods. Growth hormone is encoded by a single gene (the GH gene) in humans and in mice; the DNA sequences of the human and mouse GH genes are very similar but not identical. You have available genomic DNA clones for both the human and mouse GH genes.

You first decide to test the specific hypothesis that additional copies of the mouse GH gene would yield mice larger than wildtype (which, of course, have two copies of the GH gene).

- (a) What modification to the mouse genome would allow you to generate a mouse with three copies of the GH gene?
- (b) What additional step would yield mice with four copies of the GH gene?
- (c) What additional modification would yield mice with five copies of the GH gene?
- (d) What additional step would yield mice with six copies of the GH gene?

You then decide to test the specific hypothesis that mice with zero or one copy of the mouse GH gene would be smaller than wildtype.

- (e) What modification to the mouse genome would allow you to generate a mouse with only one copy of the GH gene? Draw the DNA construct that you would use to modify the mouse genome, and explain how your construct would integrate into the mouse genome.
- (f) What additional step would yield mice with zero copies of the GH gene?
- (g) Finally, you decide to test the hypothesis that the mouse and human GH genes are functionally interchangeable. Outline a series of modifications to the mouse genome that would allow you to test this hypothesis.

7.03 Problem Set 6

Due before 5 PM on Wednesday, November 22, 2000

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

1. In the 1950's, a screening of Europeans revealed that 30% were unable to taste the chemical compound phenylthiocarbamide (PTC). This is an autosomal recessive trait; the ability to taste PTC is dominant. Assume that, in the 1950's, the European population was in Hardy-Weinberg equilibrium with respect to the gene that is critical to PTC tasting.

(a) In Europe in the 1950's, what was the frequency of the allele (call it allele NT) associated with the inability to taste PTC?

(b) What was the frequency of heterozygotes in the European population?

(c) In Europe in the 1950's, what fraction of NT alleles were found in individuals who could not taste PTC?

What was the probability that a child born in Europe in the 1950's would be a PTC taster if:

(d) Both parents were PTC tasters.

(e) One parent was a PTC taster, but the other parent was not.

Also in the 1950's, a screening of West Africans revealed that 1% were unable to taste PTC. Assume that, in the 1950's, the West African population was in Hardy-Weinberg equilibrium with respect to the gene that is critical to PTC tasting

(f) In West Africa in the 1950's, what was the frequency of the NT allele?

(g) What was the frequency of heterozygotes in the West African population?

(h) In West Africa in the 1950's, what fraction of NT alleles were found in individuals who could not taste PTC?

(i) In the 1950's, what was the probability that a child would be an NT/NT homozygote if one parent was a European PTC taster and the other parent was a West African PTC taster?

2. In practice, it can be very difficult to detect subtle selection for or against the heterozygote for an allele that appears to be recessive. Consider a homozygous-lethal allele that has a steady-state frequency of 0.0004 when completely recessive (in which case there would be no selection for or against the heterozygote).
- (a) Calculate the mutation rate for this gene in this population.
- (b) Now change one assumption: Assume that heterozygous carriers have a fitness of 0.99. (Assume no change in the mutation rate.) What would be the steady-state frequency of the homozygous-lethal allele under these conditions?
- (c) Now reverse the assumption: Assume that heterozygotes experience an advantage of $h = 0.01$. (Again, assume no change in the mutation rate.) What would be the steady-state frequency of the homozygous-lethal allele under these conditions?
3. In class, we discussed the deleterious impact of inbreeding on the frequency and appearance of autosomal recessive diseases in human populations. But inbred strains are highly desirable in laboratory animals such as mice.
- (a) Consider two large but completely isolated human populations (populations M and N). A particular autosomal recessive disease affects 1 in 2500 people in both populations. Population M is characterized by random marriage. However, 10% of marriages in population N are between first cousins. (Assume that all other marriages in population N are random.) What is the frequency of the disease-associated allele in population M? In population N? Show your calculations.
- (b) Consider two highly inbred, true-breeding strains of mice, BL6 and DBA. BL6 and DBA differ genetically at many loci on all 20 chromosome pairs. You are studying a new mutation, with a recessive disease phenotype, that has arisen in your otherwise true-breeding BL6 mouse colony. The new mutation is recombinationally inseparable from an SSR at which BL6 and DBA differ. Having identified the new mutation in BL6 mice, you now wish to study the same mutation in DBA mice. You decide to "move" the mutation to DBA by careful breeding (rather than by transgenic manipulations). Outline a breeding plan, requiring as few generations as possible, that would yield mice that 1) are homozygous for the new mutant allele but 2) are homozygous for DBA alleles at >99% of all genes in the genome.

7.03 Problem Set 6 Solutions
Fall 2000

Problem 1:

a) $f(\text{NT}/\text{NT}) = q^2 = 0.3$
 $f(\text{NT}) = q = 0.55 = 55\%$

b) $p = 1 - q$
 $p = 1 - 0.55 = 0.45$
 $f(\text{NT}/\text{T}) = 2pq = 2(0.55)(0.45) = 0.50 = 50\%$

c) $(0.3)/(0.3 + 1/2(0.5)) = 0.545 = 55\%$

d) $p(\text{a taster is heterozygous}) = (2pq)/(p^2 + 2pq)$
 $= (0.5)/(0.2 + 0.5) = (0.5/0.7)$

$p(\text{taster offspring}) = 1 - p(\text{non-taster offspring})$

$p(\text{both parents were heterozygous and both give NT allele}) =$

$(0.5/0.7)(0.5/0.7)(1/4) = 0.1276$

$p(\text{PTC taster}) = 1 - 0.1276 = 0.87 = 87\%$

e) as above, only this time one parent is certain to give the NT allele, as he or she is a non-taster (NT/NT): $1 - (0.5/0.7)(1/2)(1) = 0.64 = 64\%$

Alternatively, this could be figured out by looking at the ways of making a taster:

$p(\text{taster parent is T/T and gives T allele}) + p(\text{taster parent is NT/T and gives T allele}) =$
 $(0.2/0.7)(1) + (0.5/0.7)(1/2) = 0.64 = 64\%$

f) $f(\text{NT}/\text{NT}) = q^2 = 0.01$
 $q = 0.1$

g) $p = 1 - q = 0.9$
 $f(\text{NT}/\text{T}) = 2pq = 2(0.1)(0.9) = 0.18 = 18\%$

h) $(0.01)/(0.01 + 1/2(0.18)) = 0.1 = 10\%$

i) $p(\text{each parent was a heterozygote and gave the NT allele}):$
 $(5/7)(18/99)(1/4) = 0.032 = 3\%$

Problem 2:

a) $S=1; q = 0.0004$
 $\mu = Sq^2$
 $\mu = 0.00000016 = 1.6 \times 10^{-7}$

b) $\Delta q_{\text{mut}} + \Delta q_{\text{sel}} = 0$
 $\mu - \frac{1}{2}(2S_{\text{het}}pq) - Sq^2 = 0$

$$\begin{aligned} \mu &= S_{\text{het}}pq + Sq^2 \\ Sq^2 + S_{\text{het}}pq - \mu &= 0 \\ q^2 + 0.01(q)(1-q) - 0.00000016 &= 0 \\ 0.99q^2 + 0.01q - 0.00000016 &= 0 \text{ (a quadratic equation)} \\ q &= 1.57 \times 10^{-5} \cong 1.6 \times 10^{-5} \end{aligned}$$

What if we make the approximation, $1-q$ ($\cong p$) is about equal to 1?

$$\text{Then we get: } q^2 + 0.01q - \mu = 0$$

$$\text{Solving for } q, \text{ we get: } q = 1.597 \times 10^{-5} \cong 1.6 \times 10^{-5}$$

Thus, in this case, the approximation is valid.

c) Again, $\Delta q_{\text{mut}} + \Delta q_{\text{sel}} = 0$

Using the above approximation (that $p \cong 1$):

$$Sq^2 - hq - \mu = 0 \text{ (see lecture notes for derivation of } \Delta q_{\text{sel}})$$

$$q^2 - 0.01q - 0.00000016 = 0$$

$$q = 1.0 \times 10^{-2}$$

Problem 3:

a) for M: $q^2 = 0.0004$; $q = 0.02$
 for N: $(F_{1\text{st cousins}})(q)(f_{1\text{st cousin marriages}}) + q^2 = f(aa)$
 $(1/16)q(0.1) + (9/10)q^2 = 0.0004$
 $(0.9)q^2 + (1/160)q - 0.0004 = 0$
 $q = 1.8 \times 10^{-2}$

b) Cross BL6 mouse carrying the mutation to a DBA mouse. The resulting progeny will carry genetic material that is 50% BL6 and 50% DBA in origin. These progeny can be screened for the presence of the mutation of interest by looking for the SSR which was inseparable from that mutation. Such mice, which carry the mutation, can be again bred to DBA mice, resulting in progeny which carry genetic material that is 75% DBA and 25% BL6. As with their parent before them, mice which carry the original mutation can be identified due to the presence of the linked SSR. Continuing this logic, one can determine how many generations of such matings are necessary to obtain a mouse which is homozygous for DBA alleles at greater than 99% of all genes in the genome:

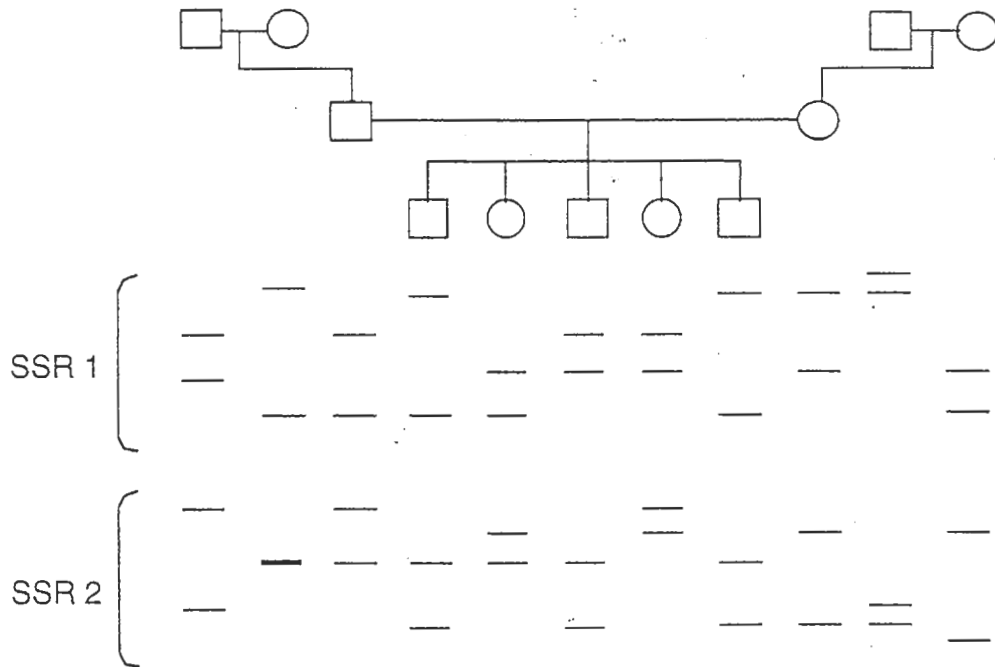
number of breedings (n) required: $1 - (1/2)^n = \text{proportion of genes from DBA}$
 thus, for $n=7$, $1 - (1/2)^7 = 0.9922$. So 7 breedings will be required to get a mouse with 99% of its loci DBA in origin. This mouse must then be bred with its siblings (the "infamous" brother - sister matings) to produce a homozygous mouse. (This entire process will likely take more than 2.5 years of your life).

7.03 Problem Set 7

Due before 5 PM on Wednesday, December 13, 2000

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

1. As we have discussed in class, SSR-based genetic linkage studies in human families can be used to chromosomally localize heritable traits, including diseases. Such studies can also be used to build genetic maps among the SSRs themselves, and indeed this is how detailed genetic maps of the human genome were constructed in the 1990's. Shown here are results of typing the members of a family for two SSRs.



- (a) Calculate a LOD score for linkage at $\theta = 0.1$ between SSR1 and SSR2 in this family.
- (b) Identify a value of θ at which this family will yield a higher LOD score for linkage between SSR1 and SSR2. Calculate the LOD score for linkage between SSR1 and SSR2 at that new θ value.
- (c) Estimate (roughly) the physical distance between SSR1 and SSR2. (Assume that additional family studies confirm that SSR1 and SSR2 are located on the same chromosome and are genetically linked.)

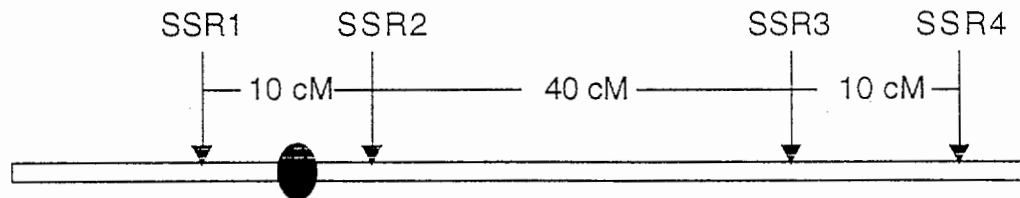
2. What concordance rates (approximate answers will suffice) might you expect in MZ twins, DZ twins, and first cousins for each of the following diseases? Briefly justify each of your responses.

(a) Chicken pox, a very common and contagious viral disorder.

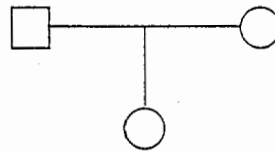
(b) Tay-Sachs disease, a rare autosomal recessive disorder in which environmental effects are negligible.

(c) A disease in which both environment and a single gene are important determinants of risk.

3. Trisomy X (that is, XXX) is one of the most common trisomies observed in human populations. XXX women are usually fertile and phenotypically unremarkable. You prepare DNA samples from two unrelated girls, both with trisomy X, and from their parents. You then type the girls and their parents for four SSRs distributed along the X chromosome:



Family 1

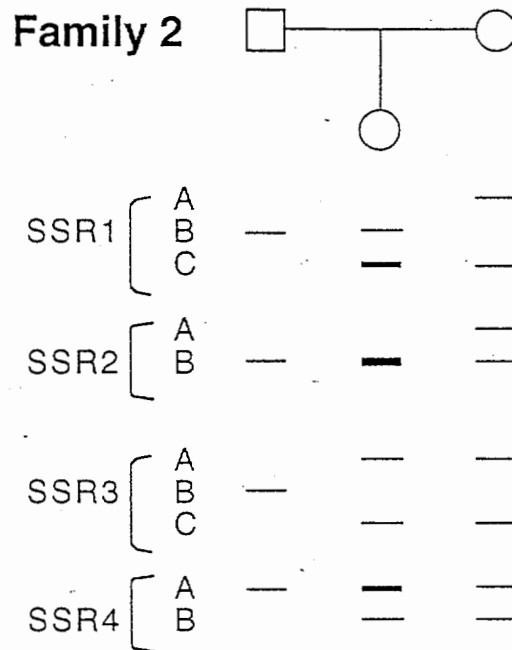


SSR1	{	A	—	—	—
		B	—	—	—
		C	—	—	—
SSR2	{	A	—	—	—
		B	—	—	—
		C	—	—	—
SSR3	{	A	—	—	—
		B	—	—	—
SSR4	{	A	—	—	—
		B	—	—	—

(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 the X chromosome.



(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 the X chromosome.

4. The tammar wallaby is one of many marsupial mammalian species that live in Australia. Like other marsupials, tammar wallabies are "born" as primitive embryos that develop more fully in their mother's pouch. (Males do not have pouches. The embryonic structure that gives rise to the pouch in the developing female gives rise to the scrotum [the eventual location of the testes] in males.) Like placental mammals (which include mice and humans), tammar wallabies and other marsupials have an XX/XY sex determination system, with XX females and XY males. Occasionally, scientists identify a tammar wallaby with XO or XXY sex chromosomes. XO wallabies are observed to have ovaries but no pouch; instead, they have a scrotum-like structure. XXY wallabies are observed to have testes and a pouch, but no scrotum.
- (a) What does this information suggest about the role(s) of the X and Y chromosomes in sex determination in tammar wallabies?
- (b) Briefly compare and contrast the roles of the sex chromosomes in sex determination in humans, *Drosophila melanogaster*, and tammar wallabies.