Name: KEY

1. (**a** 6 pts.) You are initiating a study of the Lac operon in *E. coli*, and you wish to isolate a Tn5 insertion (which carries kanamycin resistance) linked to the Lac genes. You already have a collection of 10^4 random Tn5 insertions in wild type *E. coli* and you grow P1 phage on a mixture of the entire collection. You use the resulting lysate to infect a LacZ$^-$ strain and select for Kan^r clones. You then plate 2,000 Kan^r clones on X-gal plates; among a vast majority of clones which form white colonies on X-gal you identify 3 clones that are both Kan^r and that form blue colonies on X-gal. For these three clones what can you conclude about the distance between the Tn5 insertion and the LacZ gene? Be as complete and specific as you can.

6 pts: The Tn5 insertion and LacZ are within 1 P1 head-full ~50 E. coli genome distance 10^5 Bp

(**b** 6 pts.) You grow P1 phage on one of the three Kan^r clones that is blue on X-gal. You use the resulting lysate to infect your LacZ$^-$ strain and select for Kan^r clones. Plating 100 of these clones on X-gal plates you find that 25 form white colonies and 75 form blue colonies. What is the distance (measured as a cotransduction frequency) between the Tn5 insertion and LacZ$^-$?

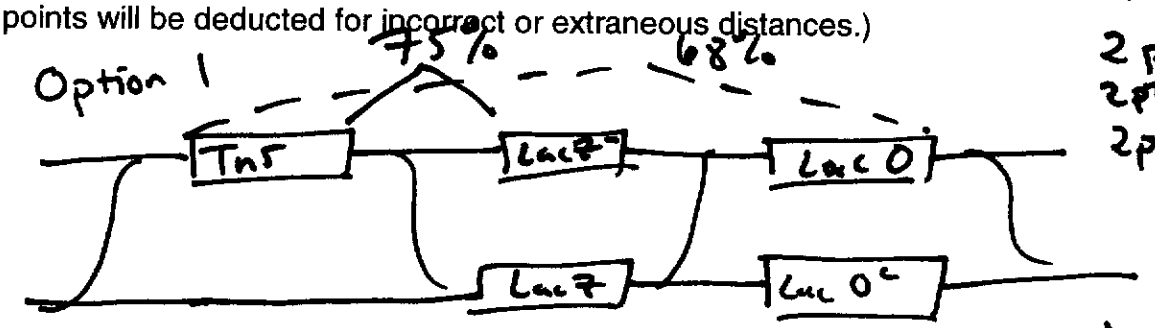
6 pts 75% cotransduction

Name: _____

(c 8 pts.) To map the Tn5 mutation relative to both LacZ$^-$ and LacO^c mutations you carry out two crosses. In the first cross you grow P1 phage on a strain with the Tn5 insertion and LacZ$^-$ and then use the lysate to infect a LacO^c strain selecting for Kan^r clones. Among 100 Kan^r clones, 25 express Lac constitutively, 1 has normal Lac expression, and 74 have uninducible Lac expression. In the second cross you grow P1 phage on a strain with the Tn5 insertion and LacO^c and then use this lysate to infect a LacZ$^-$ strain selecting for Kan^r clones. Among 100 Kan^r clones, 68 express Lac constitutively, 7 have normal Lac expression, and 25 have uninducible Lac expression. Draw a map showing the relationship between the Tn5 insertion and the LacZ$^-$ and LacO^c mutations. Show any relevant distances as cotransduction frequencies. (Note: assume the Tn5 insertion is not between LacO^c and LacZ$^-$, also note points will be deducted for incorrect or extraneous distances.)

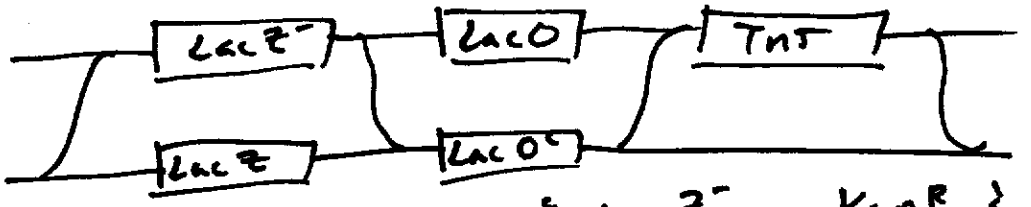
2 pts set up
2 pts rare
2 pts each distance

Option 1



rare = Tn5 : LacZ : LacO → KanR ; WT

Option 2



rare : Tn5 : LacO^c : LacZ$^-$ KanR ; uninducible

Name: _____

2. A short segment from the middle of a bacterial gene is shown below. Also shown are the amino acid sequence translations in each of the three reading frames reading left to right. The first of the three frames corresponds to the open reading frame for the gene. The arrow marks a point in the sequence between a C and an A residue. Stop codons (UAG, UAA, and UGA) are translated as: XXX.

```
                    ↓
5'CTTCGGGCGATAGAGATACGCAAGCGAGATATCTCACTGACTGAGATC 3'
  LeuArgAlaIleGlnIleArgLysArgAspIleSerAlaArgAspLeu
  PheGlyArgXXXArgTyrAlaSerGluIleSerGlnXXXLeuArg
    SerGlyAspArgAspThrGlnAlaArgTyrLeuThrAspXXXAsp
```

(a 8 pts.) Consider a mutation in the gene caused by insertion of a T at the position of the arrow. Write out the last three amino acids of the protein product of the mutated gene.

insertion causes stop TAA

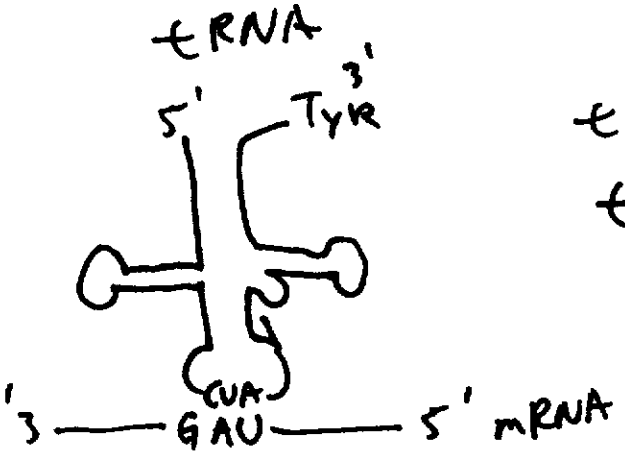
Gln, Ile, Arg

(b 8 pts.) Next consider a mutation in the gene caused by insertion of two T residues (i.e. sequence TT) at the position of the arrow. Write out the last three amino acids of the protein product of the mutated gene..

insertion causes frame shift to 2nd frame

Ile, Ser, Gln

(c 8 pts.) Write out the sequence of the **template** strand of the portion of an amber (UAG) suppressing allele of a tRNA gene that codes for the anti-codon portion of the tRNA. Be sure to label the 5' and 3' ends of the DNA sequence.



tRNA 5'—CUA—3'

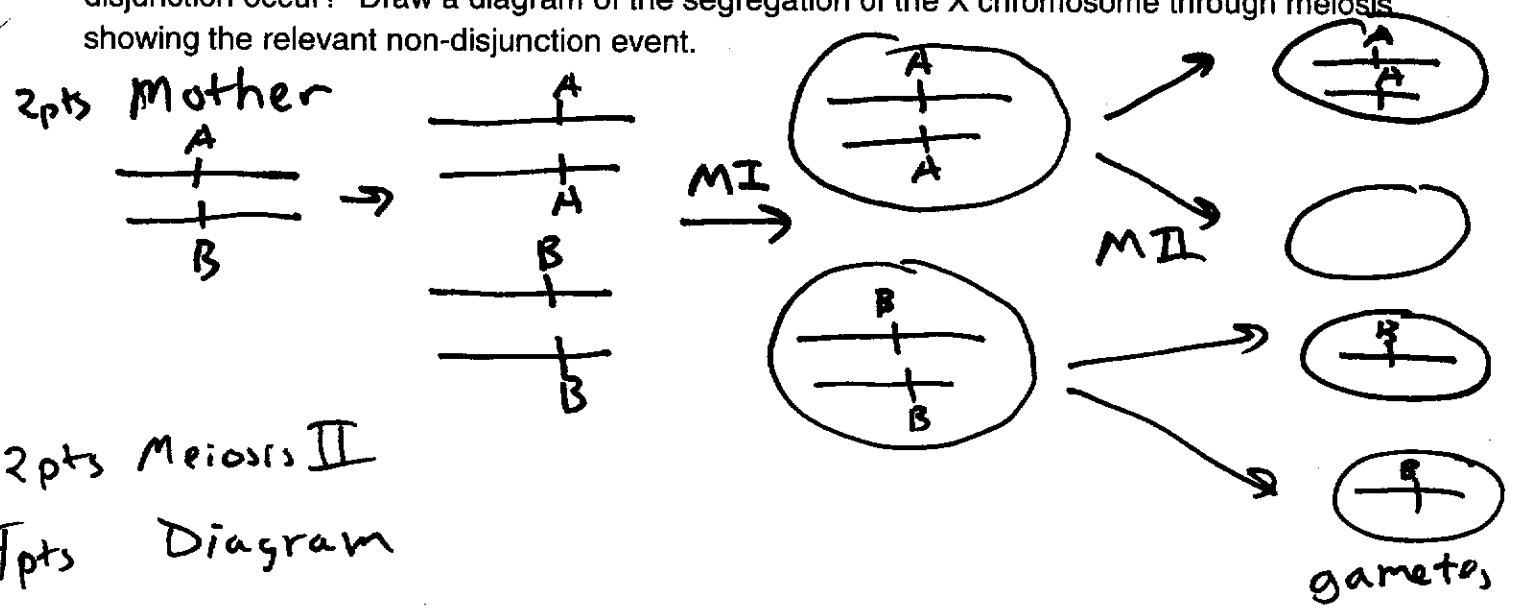
tRNA template 3'—GAT—5'

5'—TAG—3'

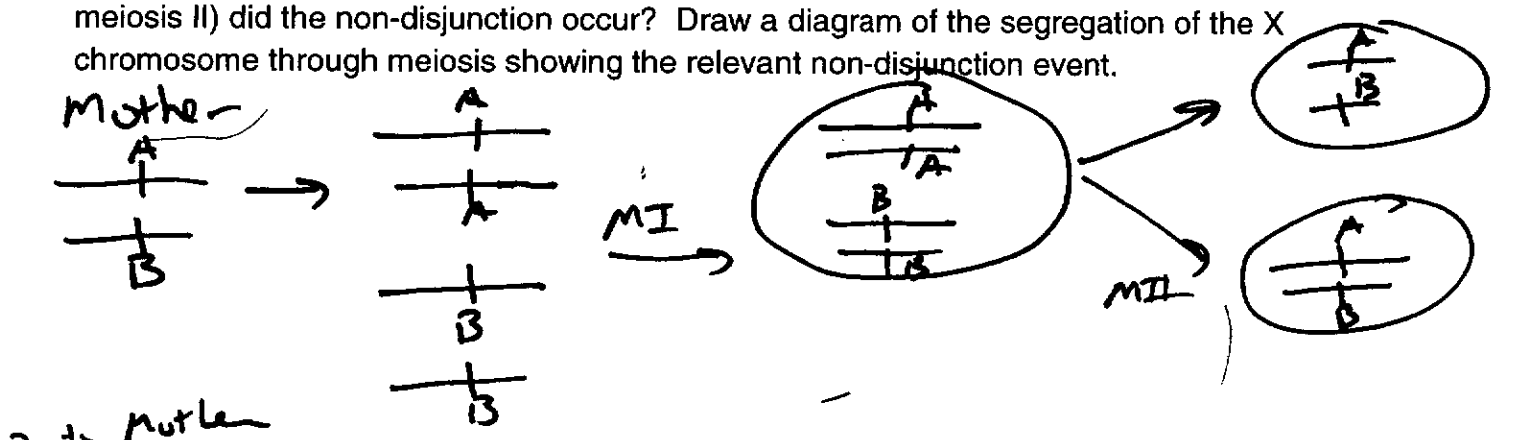
Name: _____

3. Kleinfelter's syndrome is an abnormality caused by a non-disjunction event during meiosis leading to an extra X chromosome. Individuals with Kleinfelter's syndrome have sex chromosome karyotype XXY (normal females have karyotype XX and normal males have karyotype XY). You have identified an SSR (simple sequence repeat) DNA marker that lies close to the centromere of the X chromosome. Analysis of the SSR alleles in the parents of a Kleinfelter's child reveals that the mother has SSR alleles A and B, and the father has allele C.

(a 8 pts.) Say that the Kleinfelter's child has two copies of SSR allele A. Which parent did the non-disjunction occur in and at what stage of meiosis (meiosis I or meiosis II) did the non-disjunction occur? Draw a diagram of the segregation of the X chromosome through meiosis showing the relevant non-disjunction event.



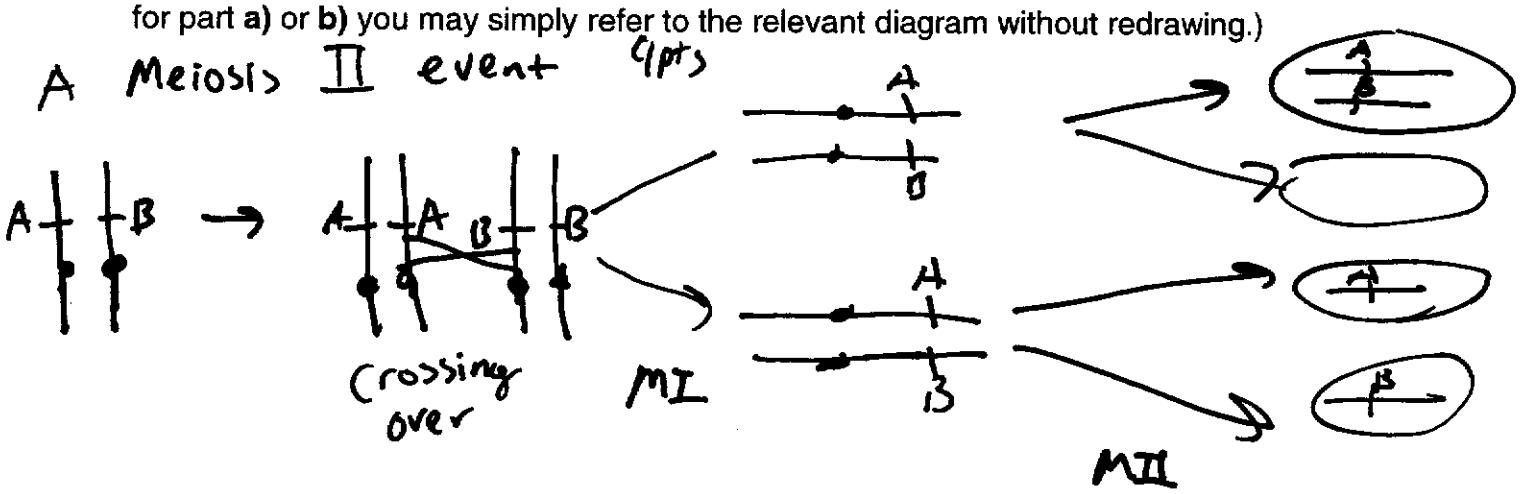
(b 8 pts.) Say that the Kleinfelter's child has one copy each of SSR allele A and allele B. Which parent did the non-disjunction occur in and at what stage of meiosis (meiosis I or meiosis II) did the non-disjunction occur? Draw a diagram of the segregation of the X chromosome through meiosis showing the relevant non-disjunction event.



Name: _____

(c 8 pts.) An SSR marker that is unlinked to the centromere on the X chromosome can produce ambiguous results in such tests. Assume that the SSR marker is distant from the centromere on the X chromosome. Diagram two different nondisjunction events that could produce a Kleinfelter's child with allele A and allele B. (If one of the events is the same as that for part **a)** or **b)** you may simply refer to the relevant diagram without redrawing.)

A Meiosis II event 4pts



A Meiosis I event same as part B 4pts

4. You are studying regulation of a pathway that controls growth of yeast cells. To begin your analysis you isolate two mutants called Gro1$^-$ and Gro2$^-$ that produce cells that are <u>smaller</u> than normal.

(a 4 pts.) You cross each of the haploid mutants Gro1$^-$ and Gro2$^-$ to wild type and find that the resulting diploids have normal cell size. Furthermore diploids formed by crossing Gro1$^-$ to Gro2$^-$ are also normal in size. What do these results tell you about the relationship of the two Gro mutations? Explain.

Name:

(c 8 pts.) An SSR marker that is unlinked to the centromere on the X chromosome can produce ambiguous results in such tests. Assume that the SSR marker is distant from the centromere on the X chromosome. Diagram two different nondisjunction events that could produce a Kleinfelter's child with allele A and allele B. (If one of the events is the same as that for part **a)** or **b)** you may simply refer to the relevant diagram without redrawing.)

4. You are studying regulation of a pathway that controls growth of yeast cells. To begin your analysis you isolate two mutants called Gro1$^-$ and Gro2$^-$ that produce cells that are <u>smaller</u> than normal.

(a 4 pts.) You cross each of the haploid mutants Gro1$^-$ and Gro2$^-$ to wild type and find that the resulting diploids have normal cell size. Furthermore diploids formed by crossing Gro1$^-$ to Gro2$^-$ are also normal in size. What do these results tell you about the relationship of the two Gro mutations? Explain.

+1/ Gro1$^-$ x WT → 1$^-$/+ Normal → Recessive

+1/ Gro2$^-$ x WT → 2$^-$/+ normal → Recessive

+2/Gro1$^-$ x Gro2$^-$ → 1$^-$/2$^-$ Normal

Complement: diff. genes

Haploid Mutants: Gro1$^-$: small; Gro2$^-$: small

Name: _____

(b 6 pts.) Assume that the Gro1⁻ and Gro2⁻ mutations affect act in a pathway that controls cell growth (you can think of cell growth as the output of the pathway). What normal regulatory function can you ascribe to the wild type function corresponding to each of these two mutations. Diagram the possible linear pathways are consistent with the data you have so far.

Activators

2 possibilities! (+3)

(1) Gro2 (+) → Gro1 (+) → Growth

(2) Gro1 (+) → Gro2 (+) → Growth (+3)

(c 6 pts.) To move the analysis forward you decide to isolate a suppressor of the Gro1⁻ mutation. You mutagenize a haploid Gro1⁻ mutant and isolate a revertant that makes large cells. Careful examination reveals that the revertant produces cells that are even larger than wild type. You mate the revertant strain to a wild type haploid and note that the resulting diploid has cells that are larger than normal. Sporulation of this diploid produces 20 tetrads of the following types: 11 tetrads have 2 with large cells and 2 clones with normal cells; 8 tetrads have 2 clones with large cells, 1 clone with normal cells and 1 clone with small cells; and 1 tetrad has 2 clones with large cells and 2 clones with small cells. What do these results tell you about the nature of the reversion mutation? Provide as much information and be as specific as you can. (+3)

→ downstream rev. and Gro1 (+1)

Gro1⁻ revertant = Large

Gro1^rev x WT

[gro1⁻, rev] ↓ [+, +]

Larger (+1)

Dominant

Sporulate.

1⁻ rev / + +

PD	NPD	TT
1⁻ rev L	1⁻ + S	1⁻ + S
1⁻ rev L	1⁻ + S	+ rev L
+ + N	+ rev L	1⁻ rev L
+ + N	+ rev L	+ + N
11	1	8

11 PD : 8 TT : 1 NPD

cM = 100 × (TT + 6 NPD) / 22

= 100 × (8 + 6) / 40

= 35 cM (+1)

(+1) trans acting diff. gene

(d 4 pts.) Describe how you would pick out a yeast strain with just the reversion mutation on its own (i.e. separated from the original Gro1⁻ mutation) from the tetrads in part **c)**.

The large cells from 1NPD ⇒ (rev, +)

(+4)

Name: _____

(e 6 pts.) Next you mate the reversion mutation on its own isolated in part **d)** to a Gro2$^-$ mutant. Sporulation of the resulting diploid produces 20 tetrads all of which have 2 clones with large cells and 2 clones with small cells (i.e. none of the clones have normal cells. What does this result tell you about the relationship between the revertant mutation and the Gro2$^-$ mutation? Be as specific as you can.

rev, + (NPD) × +, Gro 2$^-$

↓

rev, +
+ 2$^-$

(PD)

rev+ L
rev+ L
+ 2$^-$ S
+ 2$^-$ S

20

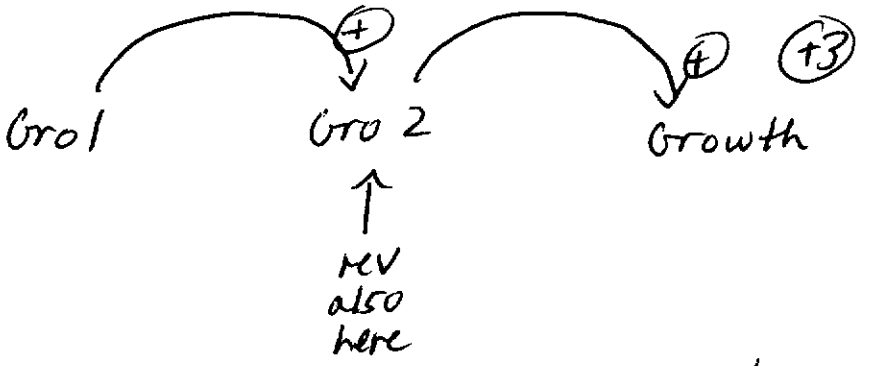
All PD (+4)
so completely (+2)
LINKED
(rev and Gro 2)

(f 6 pts.) Given all of this information construct a model for regulation of growth by the wild type regulatory factors affected by the Gro1$^-$ and Gro2$^-$ mutations. Assume that the factors for this pathway work together in a single linear pathway. Also propose a molecular explanation for the revertant mutation in the context of your model.

(1) Gro 1$^-$ and Gro^{rev} is 35cM (far enough apart, trans)

(2) Gro 2$^-$ and rev are completely LINKED

(3) Gro 1$^-$ and Gro 2$^-$ are therefore trans



(+3)
→ Gro Rev$^-$ activator^S mutation, causing activation of Gro 2 and Growth

→ const. active

Name: _____

5. You are trying to recapitulate in the mouse the human autosomal recessive disease cystic fibrosis. To do this, you isolate the human CFTR gene carrying the ΔF508 mutation, which is the most common cystic fibrosis causing mutation. You inject DNA for this gene into the pronuclei of a number of fertilized mouse eggs and then implant the injected eggs into pseudopregnant female mice. To screen for mice carrying the ΔF508 transgene you design a pair of PCR primers (called primer pair 1) that will amplify a 1.2 kb segment spanning the human CFTR gene but that don't match the endogenous mouse CFTR gene. Using these primers you test by PCR a tail sample from each of the mice produced from the pronuclear injection procedure and find a single mouse that yields a 1.2 kb band with Primer pair 1. You designate this mouse: Mouse 1.

(a 4 pts.) Next, you cross Mouse 1 to a wild type mouse and test each of the nine progeny from this cross with Primer pair 1. Among the nine progeny, four produce a 1.2 kb PCR product whereas five do not produce any PCR product. What does this result tell you about Mouse 1. Be as specific and complete as possible.

Mouse 1 has successfully integrated the human CFTR gene. However, since you cross Mouse 1 to WT, ~~it is~~ the progeny are hetreozygous, Not all gametes have the human CFTR gene

(+4)

(b 4 pts.) All nine of the progeny from part **a)** appear to be normal (i.e. they show no sign of a cystic fibrosis like disease) regardless of whether they test positive for PCR product with Primer pair 1. Is this surprising? Explain why or why not.

No, cystic fibrosis - autosomal recessive

Presence of endogenous mice gene that encode fo CFR are sufficient to ensure that mice do not show trait

(+4)

Name: _____

In a parallel experiment you make a construct that replaces the mouse CFTR gene with the marker Neo^r. This construct is introduced into mouse ES cells and a gene replacement by homologous replacement is selected. ES cells carrying an appropriate knockout construct are introduced into blastocysts, which are implanted into pseudopregnant females. The resulting chimeric progeny are then backcrossed to produce mice that carry the knockout construct in the germline. To test the genotypes of these mice you design a second pair of PCR primers (called Primer pair 2) that will amplify a 2.0 kb segment spanning the intact mouse CFTR gene or a 2.5 kb segment if the mouse CFTR gene has been disrupted with the Neo^r marker.

(c 4 pts.) You test tail samples from the backcrossed mice using Primer pair 2 and you identify several mice that each yield both 2.5 kb and 2.0 kb PCR products. Crosses among these mice produce many progeny that exhibit the 2.5 kb and 2.0 kb PCR products as well as many progeny that produce only the 2.0 kb PCR product, but there are **no** progeny mice that produce only the 2.5 kb PCR product. Explain the absence of this class of progeny.

→ CTFR essential gene (+2)
→ need at least one to survive (+2)

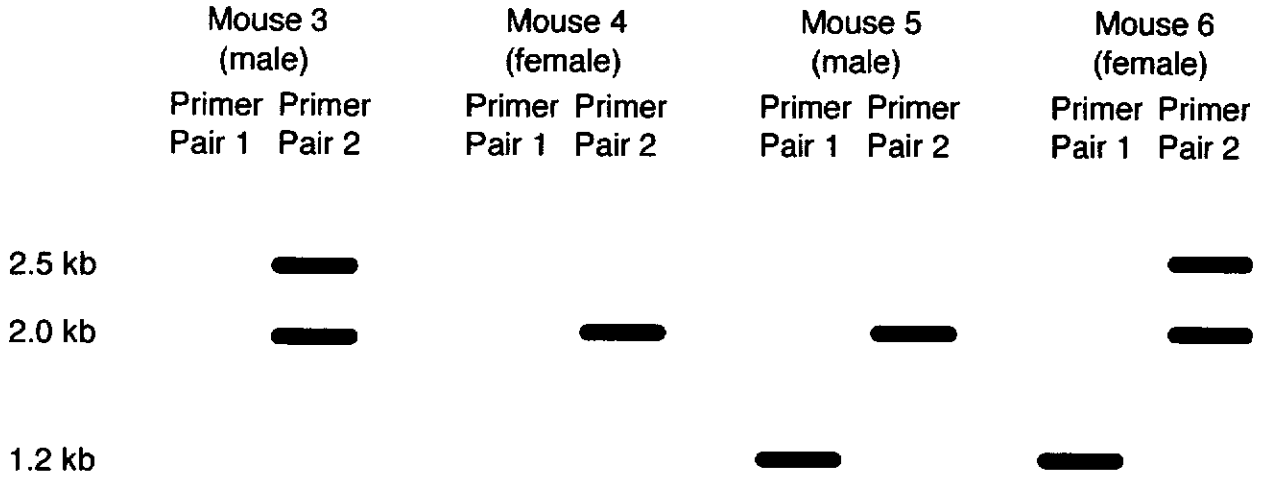
(d 6 pts.) Using the constructs you already have in hand write out the ideal genotype of a mouse that you would like to construct that has the greatest likelihood of exhibiting a mouse version of cystic fibrosis. Be sure to define the geneotypic designations you use, clearly specifing the different versions CFTR (mouse vs. human; wild type vs. ΔF508)

$CFTR_m$ ⇒ normal mouse CFTR
$CFTR_m^-$ ⇒ knockedout CFTR (homologous rec.)
$CFTR_H^{\Delta F508}$ ⇒ human CFTR carrying ΔF508 mut

ideal: $\frac{CFTR_m^- \quad CFTR_H^{\Delta 508}}{CFTR_M^- \quad -}$ (+6)

Name: _____

One of the mice from part **c)** that exhibits both the 2.5 kb and 2.0 kb PCR products is designated Mouse 2. You cross Mouse 1 to Mouse 2 and four progeny from this cross are shown below.



(e 6 pts.) Select two of these mice that could be crossed to produce progeny with the desired genotype from part **d)**. Give the frequency that you would expect mice of the desired genotype among the progeny of the cross.

(+1) Mate 3 and (+1) 6

⇓

50% have $CFTR_{M}^{4FO8}$

25% $CFTR_{M}^{-}/CFTR_{M}^{-}$

⇓

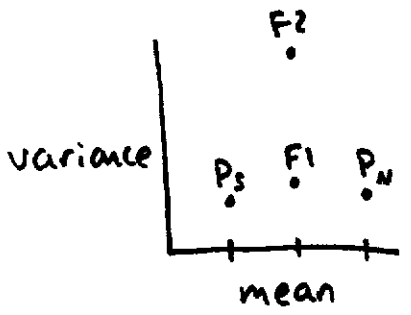
(+4) 1/4 of mice would have ideal genotype

one is lethal

Name: _____

6. You are studying the quantitative trait of birth weight in mice. You start by working with two inbred mice lines, *Skinny* and *Normal*, with mean weights of 10gr and 30gr respectively.

(a 5 points) You cross the mice to generate F1s, and F2s. In the F1 mice, body weight has a mean of 20 and a standard deviation of 2. In the F2s, the mean is 20 and the standard deviation is 3. Are these values roughly consistent with an assumption of additive effects of the genes of the trait (hint: would they fit on a "triangle-like plot")? Why or why not?



These values are consistent with additivity because the means of F1 and F2 are halfway between the means of P_S and P_N, indicating that when all genes are heterozygous, their value in the phenotype is the average of the values of the homozygotes.

(b 5 points) What are the environmental variance (σ_e^2), genetic variance (σ_g^2) and broad-sense heritability (H^2) of birth weight in these mice? Show your calculations. Assume that the trait is affected additively.

$\sigma_P^2 = \sigma_g^2 + \sigma_e^2$ $\quad 9 = \sigma_g^2 + 4$ $\quad H^2 = \frac{\sigma_g^2}{\sigma_P^2} = \frac{5}{9}$

$\sigma_P^2 = \sigma_{F2}^2 = 3^2 = 9$ $\quad \sigma_g^2 = 5$

$\sigma_e^2 = \sigma_{F1}^2 = 2^2 = 4$

(c 5 points) Given those values, what is the estimated minimal number of genes that affect birth body weight in these mice?

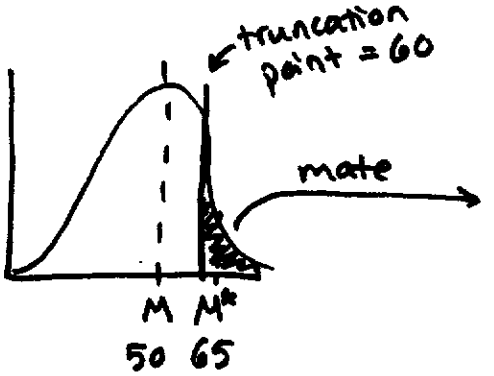
$$n = \frac{D^2}{8\sigma_g^2} = \frac{400}{8 \cdot 5} = 10$$

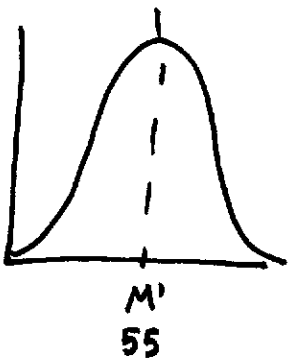
$D = \mu_{P_S} - \mu_{P_N} = -20$

$D^2 = 400$

Name: _____

(d 5 points) You discover an outbred (genetically heterogeneous) population of field mice in Maine, the *HeavyWeights*, with a mean body weight of 50 (substantially higher than that of your previous lines) and a standard deviation of 5. You decide to conduct a breeding experiment. You choose 60 (10 grams above the mean) as the truncation point; the resulting parent population has a mean of 65. The offspring have a mean weight of 55. What is the narrow-sense heritability (h^2) of the trait?





$$h^2 = \frac{M' - M}{M^* - M}$$

$$= \frac{55 - 50}{65 - 50}$$

$$= \frac{5}{15}$$

$$= \frac{1}{3}$$

(e 5 points) Your colleague, a human geneticist, has decided to determine the narrow-sense heritability of birth body weight in humans from a study of parents and daughters. She found that the correlation coefficient between the birth body weight of fathers and their daughters is **1/6**, but that the correlation coefficient between the birth body weight of mothers and daughters is **1/2**. What is the narrow sense heritability based on each of these correlation coefficients? Assuming the weight at birth correlates to adult weight, what could be a reason for the discrepancy between the estimates?

offspring and one parent: $r = \frac{h^2}{2}$: $h^2 = 2r$

fathers and daughters: $h^2 = \frac{1}{3}$

mothers and daughters: $h^2 = 1$

Mother is fetus' environment during gestation.

Name: _____

7. You are interested in two genes that affect the risk for heart disease: Genes *Heart1* and *Heart2*. Your anthropologist colleague in Oxford, UK tells you about a recent finding: a human population on an isolated island in the Pacific ocean, where the Pacific Islander population was invaded by a European population from a shipwreck 5 generations ago. You are excited since you know from previous studies that Pacific Islanders **only** have H1 and H2 risky alleles and Europeans **only** have h1 and h2 protective alleles. Assume the two genes are encoded on the same chromosome, and that the recombination rate between them is r=0.1.

(a 5 points) Explain without any calculation if the two loci were likely in LD or not upon mixture of the two populations.

The loci were in LD because when the European population arrived, there were only two ~~populations~~ haplotypes in the mixed population: H1 H2 and h1 h2. The other combinations are entirely missing.

(b 5 points) Knowing from the Admiralty archives in London that 200 Europeans invaded an island with 1000 Pacific islanders, Calculate D_0 (the degree of LD in the initial mixed population).

haplotype	#
H1 H2	2000
h1 h2	400
total:	2400

$$D = P(H1H2)P(h1h2) - P(H1h2)P(h1H2)$$
$$= \left(\frac{2000}{2400}\right)\left(\frac{400}{2400}\right) = .13886$$

(c 5 points) What is the degree of LD (expressed as D) in the current population? (i.e. D_5)

$$D_n = (1-r)^n D_0$$
$$D_5 = (1-0.1)^5(0.13886)$$
$$= (.9^5)(0.13886)$$
$$= .081995$$

Name: _____

(d 5 points) Since the island population has grown substantially, you decide to examine your assumptions regarding the role of the two genes in heart disease. Fortunately, a big clinical study is being conducted in the island, and careful medical records allow you to assess whether each sampled individual has heart disease or not. You ask your colleague to genotype the Heart1 gene in these subjects. Overall, you obtain information from 100 cases and 200 controls. When you examine their genotypes you come up with the following contingency table:

	Cases (with AMD)	Controls (without AMD)	Totals
Allele H1	**110**	**150**	**260**
Allele h1	**90**	**250**	**340**
Totals	**200**	**400**	**600**

Is allele H1 associated with an increased risk as reflected by an odds ratio? Estimate your confidence in this association (consider P<0.01 as confident). [A table of chi-square values is on the last page of the exam.]

odds ratio (allele H1): $\frac{ad}{bc} = \frac{(110)(250)}{(150)(90)} = 2.04$ ∴ H1 may be associated w/ disease.

χ^2: null hypothesis = allele not associated w/ disease.

$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(b+d)(a+c)} = 16.629 \rightarrow 10^{-4} > p > 10^{-5}$

∴ reject null hypothesis. H1 associated w/ disease.

(e 5 points) As you were writing up your finding, your colleague informed you that the Heart1 genotyping study was in fact done on a genome-wide level, measuring 1 million SNPs (including the two alleles of the Heart1 gene). You are highly enthusiastic and test each of the SNPs for association, hoping to find additional factors. What would be the required significance level in an individual test, to achieve a genome-wide level of p<0.01? Would your result from (d) hold at the genome wide level? Why or why not?

want 0.01 probability for whole family →

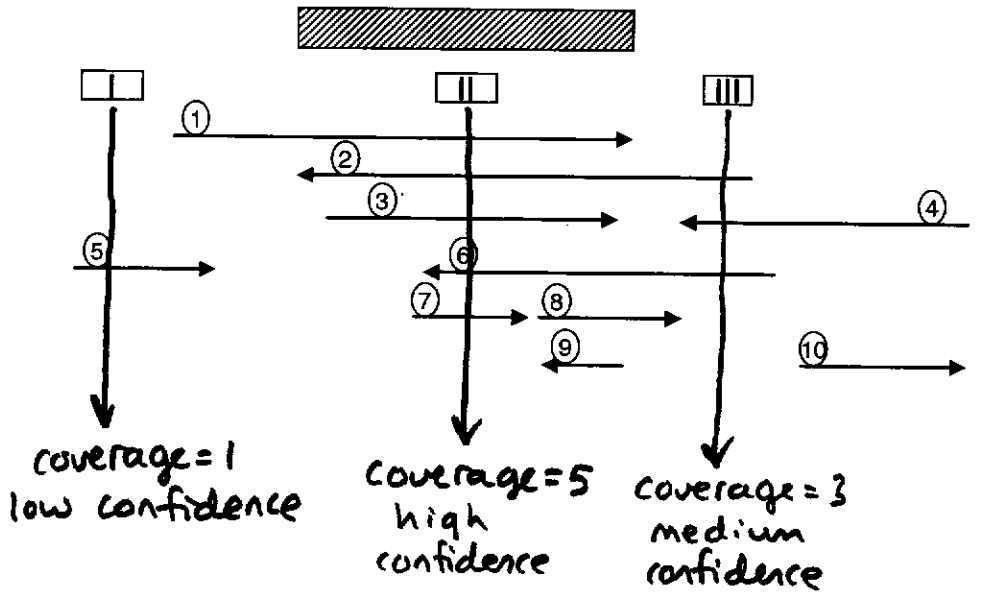
must use $\frac{0.01}{N} = \frac{0.01}{10^6} = 10^{-8}$

∴ above test would not make new threshold and thus result would not hold.

Name: _____

8. You study the filamentous fungus *Fusarium*, a major agricultural pathogen and an emerging human pathogen. In your project, you shotgun sequence both the genome (DNA) sequence and an RNA sample from one strain, known to be highly infectious and pathogenic to tomatoes. For simplicity, we assume the species has only one chromosome.

(a 5 points) You first examine some of your RNA sequencing results. You obtain the following contig. For each of the regions, I, II and III, what is the coverage of the base at the middle of each region? What is your confidence (high, medium, low) in the contig for each of these regions I, II, III?



(b 5 points) By examining your genome sequence, you realized that the sequence covered by the hatched box is a repeat, present at hundreds of thousands of copies in this genome. Does that affect your confidence in the contig above? How and why?

This lowers my confidence.
The sequences to the left and right of the repeat are only held together by the repeat, and thus may not be in the same region at all.

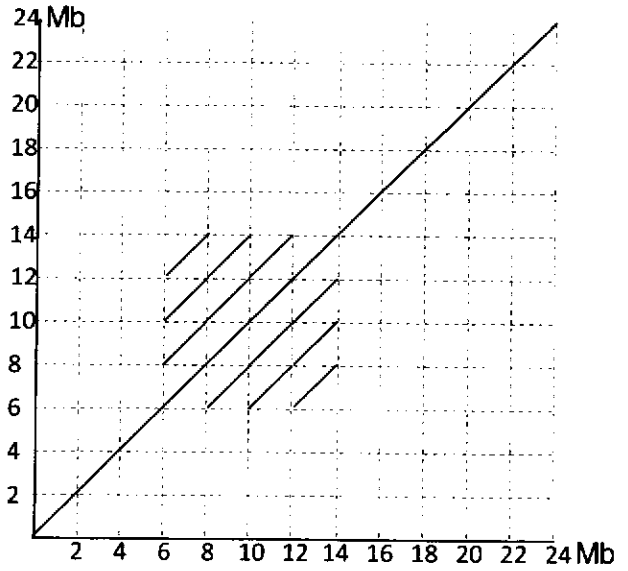
Name: _____

(c 5 points) You now received the paired-end information for your sequencing reads. The paired-end reads are:
Read 5 and 2
Read 1 and 6
Read 3 and 4

How would you revise your confidence in the contig with this additional information?

Increases confidence.
Now know that region to left is in physical proximity to region to right.

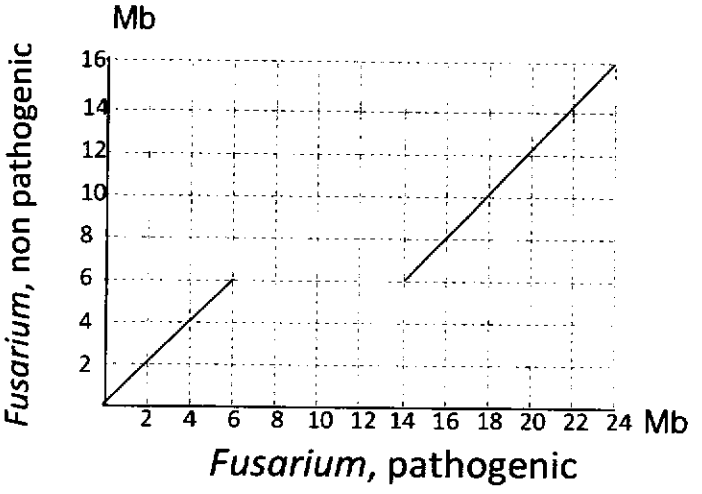
(d 5 points) To examine for the presence of sequences present in multiple copies in your genome you also prepare the following dot plot of the genomic DNA sequence of the chromosome against itself. Given the pattern shown below, what is the size (in Mb) of the repeated unit and how many copies are present?

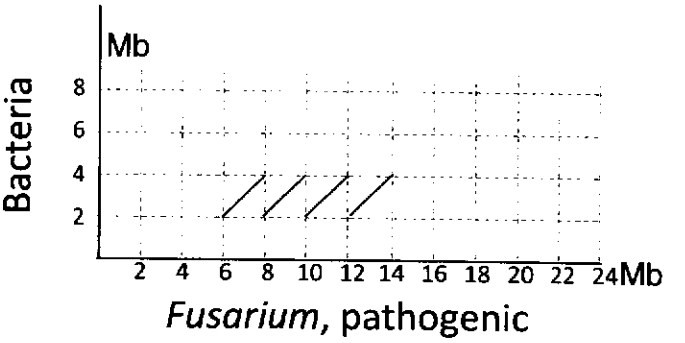


2 Mb repeat
4 repeats.

Name: _____

(e 5 points) You now compare the genome (single chromosome) of your highly pathogenic strain to two other genomic sequences **(1)** The single-chromosome genome of another strain (same species) of *Fusarium*, which is far less pathogenic to tomato. **(2)** The single-chromosome genome of a **bacterium** that infects tomatoes. What can you infer about the role of the event you saw in (d) in pathogenicity? What may be the source of this sequence in the pathogenic strain?





region from 6 to 14 not present in non-pathogenic strain

repeat region from Fusarium found once in bacteria.

Repeat in (d) it probably responsible for the pathogenicity. Pathogenic Fusarium + bacterium both have it, but non-pathogenic Fusarium did not.
This repeat may have come to the pathogenic Fusarium from the pathogenic bacterium.

Appendix: critical values for the chi-square distribution.

P-value	0.1	0.01	10^{-3}	10^{-4}	10^{-5}	10^{-6}
df=1	2.705544	6.634897	10.82757	15.13671	19.51142	23.92813
df=2	4.60517	9.21034	13.81551	18.42068	23.02585	27.63102

Grading section

Question 1 21 points:_____

Question 2 24 points:_____

Question 3 24 points:_____

Question 4 32 points:_____

Question 5 24 points:_____

Question 6 25 points:_____

Question 7 25 points:_____

Question 8 25 points:_____

Total :_____