7.03 Final Exam Review Session - SOLUTIONS

Heritability

Problem 4

You are interested in the heritability of longevity.

1. You start by studying longevity in mice. Since lab strains appear to have little variation in lifespan, you decide to conduct a large scale breeding experiment in a wild population of mice captured in Boston. In your captured population, the mean life span is 2 years, and the standard deviation is 0.5 year. You choose as your truncation point 3 years, and obtain a population with a mean of 4 years. The offspring of the selected animals have a mean life span of 2.6 years. What is the narrow sense heritability of lifespan in your mice?

SOLUTION:

To calculate the narrow sense heritability we use:

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

In our case M=2, M'=2.6, and M*=4

Hence, $h^2=(2.6-2)/(4-2)=0.3$

2. You decide to repeat the selection process, choosing a new truncation of 3.5 years, obtaining a population (from the offspring) with a mean life span of 4.2 years. What is the expected mean life span of their offspring?

SOLUTION:

To calculate the expected life span of the offspring we use the narrow sense measure we calculated in (1)

$M''=M'+h^2(M^{*\prime}-M') = 2.6+0.3*(4.2-2.6) = 3.08$

The mean expected lifespan in the next generation is 3.08 years.

3. You wish to compare your estimates to ones from human studies. You write your colleague, a human geneticist, who sends you data on longevity from a study of identical twins. The correlation coefficient of longevity between identical twins is 0.3. What is the estimated broad-sense heritability of longevity according to this data?

SOLUTION:

The correlation coefficient in a quantitative trait between identical twins is approximately equal to the broad sense heritability. We therefore estimate $H^2=0.3$.

4. While you are preparing your study for publication, another study of longevity in humans is published, this one based on measures from full-siblings. The correlation coefficient reported by that study was 0.1. What is the broad-sense heritability of longevity based on your competitor's study? How can you reconcile your finding in (3) with this new one?

SOLUTION:

The correlation coefficient in a quantitative trait between full siblings is approximately equal to half the broad sense heritability. We therefore estimate H^2=0.2.

One possible source of a higher similarity between identical twins is that their environments are also more similar than those of full siblings. They **(1)** often share embryonic membranes, **(2)** are treated more similarly by other people, **(3)** grew up exactly at the same time, and **(4)** are always the same sex. For example, to control for the factor **(4)** we should compare to measures of heritability between siblings of the same sex. To control for **(3)** we should compare to measures of heritability between fraternal twins.

Problem 2

You study the genetic basis of Crohn's disease. Your colleague at MGH has identified 200 cases (with Crohn's) and 350 controls (without Crohn's). You perform a genome-wide study of 500,000 loci.

1. The best-scoring locus, in chromosome 5, had the following results:

Cases		Controls
11	**55**	**125**
10	**80**	**75**
00	**65**	**150**
Totals	**200**	**350**

Estimate whether the genotype at this locus is significantly associated with Chron's (chi-squared values are below) at a genome-wide significance level of 0.01.

SOLUTION:

Null hypothesis: The locus' genotype is not associated with Crohn's.
Alternative hypothesis: The locus' genotype is associated with Crohn's.

We use the chi-squared test to estimate the deviation of the observed values from the expected ones given no association.

	Case-expected	Controls-expected	(O-E)^2 (cases)	(O-E)^2 (controls)	(O-E)^2/E (cases)	(O-E)^2/E (controls)	Sum((O-E)^2/E)
11	65.45454545	114.5454545	109.30	109.30	1.669823232	0.954184704	21.69256972
10	56.36363636	98.63636364	558.68	558.68	9.91202346	5.664013406	
00	78.18181818	136.8181818	173.76	173.76	2.222515856	1.270009061	

The chi-square value is 21.69. We compare to the critical values with **df=2** and find that this value is above the critical threshold for $p<10^{-4}$ (but below that for 10^{-5}).
However, we need to consider the 500,000 tests we did. To achieve genome-wide significant of 0.01 when performing $5x10^5$ tests, we use the Bonferoni correction and accept only nominal p-values below
$0.01/5x10^5 = 0.2x10^{-7}$

Our test was only significant at a nominal level between 10^{-4} and 10^{-5}. We therefore CANNOT reject the null hypothesis at a genome wide significance level of 0.01.

2. What would have happened if you only worked with alleles, not genotypes? Would allele 1 be significantly associated with Crohn's and if so at which level? Assuming that you did NOT perform multiple tests, what could be the reason to the discrepancy between your result in 1 and 2?

Null hypothesis: Allele 1 is not associated with heart disease.
Alternative hypothesis: Allele 1 is associated with heart disease.
We first derive the appropriate contingency table from the genotype information in (1):

	Cases	Controls
1	**190**	**325**
0	**210**	**375**
Totals	**400**	**700**

We now calculate the OR for allele 1:

OR (Allele 1)= (190*375)/(325*210)= 1.043956044

This suggests that the allele is not associated with risk by this measure. To estimate our confidence we perform a chi-squared test, using our shortcut formula

chi-squared =((190*375-325*210)^2*1100)/(515*585*400*700)= 0.11735837

The chi-square value is 0.11. We compare to the critical values with **df=1** and find that it is below the critical threshold for $p<0.1$. We therefore CANNOT reject the null hypothesis.

If we had conducted only one test each, then the genotype would have been significantly associated, but the allele would not have. This is likely because the heterozygous genotype carries much of the risk, rather than an individual allele.

Name:_____

1. (a 5 pts.**)** You are studying the gene for an enzyme in *E. coli* and you isolate a mutation that has no enzyme activity. When you introduce a gene for an *amber* suppressor tRNA into your mutant strain, enzyme activity is restored. Would you expect activity to be restored if a UGA suppressor tRNA had been introduced instead? Explain. **(**The sequences of the three stop codons are: $^{5'}$UAG$^{3'}$ (amber), $^{5'}$UAA$^{3'}$ (ochre), and $^{5'}$UGA$^{3'}$.)

Since the enzyme activity can be restored with an amber suppressor tRNA, it is likely that the mutation causes one amino acid in the enzyme to change to an amber stop codon. The suppressing tRNA which recognizes the UGA codon will not recognize the UAG amber stop codon. Therefore, you would not expect the UGA suppressing tRNA to restore any activity in the enzyme.

(b 5 pts.**)** Write out the sequence of the anti-codon portion of the amber suppressing tRNA (be sure to label the 5' and 3' ends of the RNA).

5' CUA 3'

(c 8 pts.**)** You sequence the gene for the enzyme in your mutant and by comparing the sequence to wild type you find a single C•G to T•A base-pair change. One strand of a short stretch of sequence containing the mutation is shown below.

```
Wild type:        5′ … G T G T G A T C C A C A T C C … 3′
                                      ↓
Mutant:           5′ … G T G T G A T C T A C A T C C … 3′
```

Is the direction of transcrption left-to-right or right-to-left for this gene segment? Explain your reasoning.

There are 6 possible reading frames associated with this piece of DNA. Transcription must proceed from right-to-left, because the only reading frame in which this mutation causes an amber stop codon (5'UGG3'→5'UAG3') is if the transcription is proceeding from right-to left.

Name: _____

(d 10 pts.**)** Next, you treat your original mutant strain with a mutagen that causes T•A to C•G mutations and isolate revertants that have restored enzyme activity. When you assay enzyme activity you find that the mutants can be grouped into two types. Type 1 has exactly the same enzyme activity as a wild type strain, whereas Type 2 has somewhat less activity than wild type. Write out the DNA sequence of the gene for the enzyme that you would predict for revertants of each type. Write out the sequence of the same DNA strand as shown in the mutant sequence shown below and be sure to label the 5' and 3' ends)

Mutant: 5′ … G T G T G A T C T A C A T C C … 3′

Type 1 Revertant :

5′ … G T G T G A T C C A C A T C C … 3′

Back Mutation: Exactly WT

Type 2 Revertant:

5′ … G T G T G A T C T G C A T C C … 3′

Other A/T base pair in stop codon mutated to turn the stop codon into single amino acid substitution. This substitution has resulted in an enzyme which only has partial activity.

2. (a 8 pts.**)** You have isolated a Tn5 insertion in a wild type *E. coli* strain that you think may be linked to the Lac operon. You grow phage P1 on the Tn5 insertion strain and use the resulting lysate to infect a LacI$^-$ mutant (i.e. a Lac repressor mutant) and select for kanamycin resistance (Kan^r). You examine 12 Kan^r transductants, and find that 9 exhibit normal Lac regulation whereas 3 show constitutive Lac expression. What is the distance between Tn5 and LacI$^-$ expressed as a cotransduction frequency?

9 of the 12 transductants exhibit a wildtype phenotype, and therefore have cotransduced I$^+$ along with the transposon. Therefore, the cotransduction frequency is 9/12 = 75%

Name: _____

(b 6 pts.**)** If the P1 lysate described in part **a)** were used to infect a LacO^c mutant (i.e. an operator constitutive mutant) would you expect the phenotypes and the proportion of different phenotypes among Kan^r transductants to be significantly different from that found for part **a)**? Explain your reasoning.

You would expect both I$^-$ and O^c mutants to have a constitutive phenotype. You would not expect the proportion of phenotypes to be significantly different because the LacI gene is tightly linked to the operator site for the Lac operon. Therefore, they should be contransduced with this Tn5 insertion at similar frequencies.

(c 10 pts.**)** You set up two reciprocal crosses with the Tn5 insertion described in part **a)**. In the first cross you grow P1 phage on a strain with the Tn5 insertion and LacI$^-$ mutation. The resulting lysate is used to infect a LacO^c mutant. Among 100 Kan^r transductants, 5 show normally regulated Lac expression and 95 show constitutive expression. In the second cross you grow P1 phage on a strain with the Tn5 insertion and LacO^c mutation. The resulting lysate is used to infect a LacI$^-$ mutant. For this cross, all 100 Kan^r transductants show constitutive Lac expression. Draw a map showing the relative order of the Tn5, and the LacO^c and LacI$^-$ mutations. Also include any relevant genetic distances that you can from the information in part **a)** and part **c)**. (Warning, points may be deducted for inclusion of incorrect distances)

Tn5 ------ O ----- I Distances: (Tn5, I) = 75% (from part a.)

The second cross does not result in any strains which show normally regulated Lac expression. Therefore, the quadruple crossover class for this cross should result in O$^+$I$^+$. With the transposon on the outside, a quadruple crossover with an order of Tn5----I----O would result in an O^{c}I$^-$ double mutant, whereas the order Tn5-----O-----I results in a quadruple crossover class of O$^+$I$^+$. Since O$^+$I$^+$ is the rare class for this cross, the order is Tn5-----O-----I. You can't determine any distances given this information, since you can't distinguish between the different recombination events which produced constitutive strains, but you can say the cotransduction frequency between Tn5 and I is 75% using the information in part a.

Name:

(d 10 pts.**)** To examine the results from the first cross in part **c)** in more detail, you mate an F' factor carrying the wild type Lac operon (F' Lac^+) into each of the 95 Kan^r transductants that show constitutive Lac expression. Among the 95 constitutive transductants, 25 are still constitutive when they carry F' Lac^+, whereas 70 show normal Lac regulation when they carry F' Lac^+. Based on these results revise your map for part **c)** putting in all of the relevant map distances.

Tn5 ------ O -------- I

Tn5, O = 70% + 5% (WT observed in part c) = 75%
Tn5, I = 70%

Now you have the information to distinguish the different crossovers, and can determine some map distances. Assuming no quadruple crossovers, 70 of the constitutive mutants were recessive, and therefore O^+I^-. 25 of the constitutive mutants were dominant, and therefore O^cI^+. Recall from part d. that 5 strains showed normal regulation, and were therefore O^+I^+. Tn5 and O^+ were cotransduced in the 70 O^+I^- strain and in the 5 WT O^+I^+. Tn5 and I^- were cotransduced in the 70 O^+I^- strains, which results in the above modified cotransduction frequencies.

3. You are studying the regulation of synthesis of the amino acid histidine in a new bacterial species and you find that the last enzyme in the pathway for histidine synthesis (the product of the HisC gene) is synthesized when there is no histidine in the medium, but is not synthesized when histidine is present.

(a 8 pts.**)** You mutagenize the bacteria by generating a collection of random insertions of the transposon Tn5 into the bacterial chromosome. You find a Tn5 insertion, designated $His1^-$, which gives *constitutive* expression of HisC. Classify the $His1^-$ mutation in terms of its likely genetic properties taking into account the type of mutation usually caused by a transposon insertion (explain your reasoning). Propose the type of regulatory function probably encoded by the wild type His1 gene. Finally, diagram a model to explain the effects of histidine and the wild type His1 gene on HisC expression, assuming a linear pathway.

Transposon insertions cause loss of function mutations which completely eliminate the functionality of the gene they insert into. Therefore, you would expect the mutant allele to be recessive. You would expect the mutation to be trans, because an insertion into the HisC gene would cause an uninducible phenotype. In addition, since transposons are so large, an insertion in the cis-regulatory region of HisC would result in a defective promoter (which would be uninducible) if it did occur. The transposon could not disrupt the operator without disrupting the promoter. Therefore the mutation must be trans. His1- is a repressor- mutation: wild-type His1 must negatively regulate HisC.
His --> His1 --| HisC

(b 8 pts.**)** Next, you isolate a second Tn10 insertion mutation, designated His2$^-$, which also shows *constitutive* HisC expression (note that Tn10 confers tetracycline resistance, Tet^r). In a transduction experiment you grow P1 phage on the His2$^-$ Tn10 (Tet^r) strain and use the resulting lysate to infect the His1$^-$ Tn5 (Kan^r) mutant strain, selecting for Tet^r transductants. You find that all of the Tet^r transductants are also Kan^r. What does this result tell you about the relationship of the His1$^-$ and His2$^-$ mutations and what is the significance for understanding the regulatory pathway for HisC?

If cotransduction was observed between the two loci, you would see a strain which was Tet^r but not Kan^r. Since this was not observed, the mutations are unlinked, and are therefore in different genes. Epistatis testing can't be performed using two mutants which have the same phenotype, so since His2- is in a different gene, you cannot determine whether that gene acts upstream or downstream of His1 using the His1- and His2- mutations.

(c 8 pts.**)** Diagram the *two* possible models for linear regulatory pathways for HisC that account for the behavior of the His1 and His2 mutations. Include a role for histidine for each model.

His1 = negative regulator
His2 = negative regulator

His --> His1 --> His2 --| HisC

His --> His2 --> His1 --| HisC

Name: _____

(d 8 pts.**)** You mutagenize the bacteria with a chemical mutagen and find a new mutation that gives *uninducible* HisC expression. Transduction experiments show that this new mutation is closely linked to the Tn5 insertion of His1$^-$. Sequencing of this region of DNA shows that the new mutation is a missense mutation in the same gene that has a Tn5 inserted into it in the His1$^-$ mutation. You therefore call the new mutation His1*. In a transduction experiment you grow P1 phage on the His2$^-$ Tn10 (Tet^r) strain and use the resulting lysate to infect the His1* mutant strain, selecting for Tet^r transductants. You find that all of the Tet^r transductants give constitutive HisC expression. Draw out the model from part (**c**) that is consistent with this new observation? Explain your reasoning.

The double mutant serves as an epistasis test. Constitutive production of HisC in the double mutant tells you that His2 is downstream of His1, so the model with His2 downstream is correct.

His --> His1 --> His2 --| HisC

(e 6 pts.**)** Propose a molecular description of how the His1* mutation might affect the function of the corresponding gene product. Be as specific as possible as to what effect the His1* mutation might have. (For example, a good molecular description of the LacI^s mutation would be: "a mutation in the repressor protein that prevents binding of the inducer lactose" – simply stating "super repressor" would not be adequate.)

The normal function of the His1 gene may be to bind histidine, and then affect the activation of His2. The His1* mutation alters the conformation of His1 such that it always activates His2, whether histidine is present or absent.