



ASSIGNMENT # 7

DEADLINE: RETURN AT THE END OF THE LABORATORY SESSION**OPTION 1:** Deposit a physical copy in the cart (front of the lab)**OPTION 2:** Submit an electronic copy (i.e., PDF file) through the LMS platform (Canvas)

NAME: _____ STUDENT #: _____ DATE: _____

GENE MAPPING BASED ON YOUR COLLECTED DATA

1. Using the class data, construct a linkage map of the yellow (**y**), white (**w**) and miniature (**m**) genes. **Show your work.** You will use the **“Three Point Mapping” virtual tool** available on Canvas or here www.ampossot.com/3point. Follow the instructions in the app and use the table below to summarize the observed data.

Hint: Are you wondering about the classes with the **sepia** phenotype? You have to omit that trait and consider only the **y**, **w** and **m** genes. For example, the number of flies in the class “**sepia**” should be added to the class “**WT**” (because they have the wild-type alleles (**y⁺ w⁺ m⁺**) for these genes). Follow the same rule to the remaining **three** classes showing the sepia trait.

Then, identify parental (**PR**) (larger classes) and double recombinant (**DR**) (smaller classes). The gene whose allele differs between the PR and DR classes is the middle gene.

Phenotype	Observed number	Genotype (F1 female gametes)	*Type of gamete (PR, SR or DR)
Wild type			
Yellow, white, miniature			
Yellow			
Miniature, white			
Yellow, white			
Mini			
White			
Yellow, mini			
Total			

2. Gene order: _____



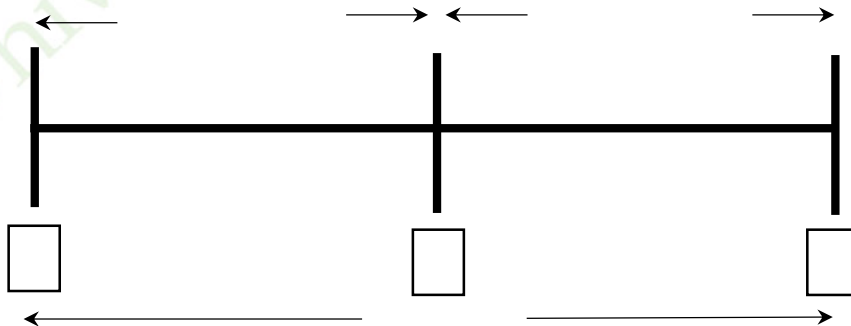
Calculate the genetic distances (**show your work and units**):

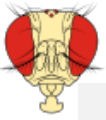
3. $y-w=$

4. $w-m=$

5. $y-m=$

Summary distances:





To calculate Coincidence and Interference, please review the formulas:

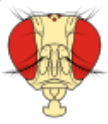
$$C = \frac{\text{Obs. Freq. DR}}{\text{Exp. Freq. DR}}$$

$$I = 1 - C$$

6. C: _____

7. I: _____

8. Please provide a brief explanation of the Coincidence and Interference values that you calculated. Is there any "interference"? What does it mean?

**X² BASED ON YOUR COLLECTED DATA**

9. Set up a X^2 to test the hypothesis that the yellow (**y**) (body color) and miniature (**m**) (wing size) are **recessive mutations** that assort **independently**. The **“Three Point Mapping” virtual tool** available on Canvas (www.ampossot.com/3point) allows you to perform such test after the mapping task.

Hint: Are you wondering about the classes with the **other** phenotypes? You must omit those traits and consider only the **y and m** genes. For example, the number of flies in the class **“sepia”** should be added to the class **“wild”** (because they are **y⁺ m⁺** for these genes). Follow the same rule to the remaining classes.

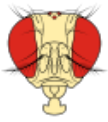
Expected ratios under Ho	Phenotype	Observed number	Expected number
	Wild type		
	Miniature wing		
	Yellow body		
	Yellow body, miniature wing		
	Total		

X^2 Calculated: _____ Degrees of freedom: _____

p=0.05): _____

Reject or fail to reject Ho: _____

10. After the X^2 test, what is your conclusion regarding the mode of inheritance of the **y - m** genes? Please provide a brief explanation.



GENOME WIDE ASSOCIATION STUDY

Conceptual connection: Like three-point mapping, GWAS attempts to identify genomic regions associated with a phenotype. However, GWAS uses allele frequencies in populations rather than recombination frequencies from genetic crosses.

For the following exercise, you will use the data collected from the “GWAS virtual tool” (www.ampossot.com/gwas).

A cohort of human individuals was sequenced using mock next-generation sequencing data. Sequencing reads have already been processed into SNP genotypes. Your task is to identify which genomic region is associated with a hypothetical disease phenotype.

11. In this simulation, you are studying a group of individuals **affected** or **unaffected** by a disease. The goal is to identify genomic regions in which SNP alleles are significantly associated with disease status. If a SNP **increases** disease risk, the risk allele should be more common in affected individuals. Then, the null hypothesis should be that the risk allele frequency is the same in affected and unaffected individuals.

Answer the following questions in the app and record your answers

Step 1: Study Design Prediction

If a SNP is associated with increased disease risk, what pattern should you expect?

Select one answer

What is the null hypothesis for this SNP association analysis?

Select one answer

Data Summary

Individuals	Affected	Unaffected	Chromosomes	SNPs

12. Calculate risk allele frequencies for one candidate SNP.

SNP ID	Chromosome	Affected risk alleles	Affected total alleles	Unaffected risk alleles	Unaffected total alleles

Calculate the risk allele frequency in **affected** individuals (affected risk alleles / affected total alleles). Show your work



Calculate the risk allele frequency in **unaffected** individuals (unaffected risk alleles / unaffected total alleles). Show your work

13. The app will calculate the Odds ratio (OR). This is a statistical measure to estimate the strength of the association between a SNP and the disease. The larger the OR, the stronger the association between the SNP and the disease.

OR = 1 → no association
OR > 1 → allele enriched in affected individuals
OR < 1 → allele may be protective

Answer the following questions in the app and record your answers

What does an odds ratio greater than 1 suggest?

Select one answer

Fill the table below:

Group	Observed <u>risk</u> alleles	Observed non-risk alleles	Total alleles
Affected			
Unaffected			
Total			

Under the null hypothesis, risk alleles should be distributed randomly between affected and unaffected individuals. The **expected** values represent what we would observe if there were no association between the SNP and disease.

Calculate the **Expected** Affected risk alleles. Show your work

(Affected Total alleles x Total Risk alleles) / grand total alleles =



Calculate the **Expected Unaffected risk alleles**. Show your work

(Unaffected Total alleles x Total Risk alleles) / grand total alleles =

14. Interpret association significance. The app calculated a X^2 analysis to test the hypothesis that the risk allele frequency is the **same** in affected and unaffected individuals.

X^2 value: _____ p-value: _____

Answer the following question in the app:

Does this SNP appear associated with disease?

Select one answer

15. Run the genome-wide association scan. The app will run the same X^2 test for the remainder SNPs and provide a list of those with the highest disease association (i.e., lower p-values).

Finally, the Manhattan plot and associated table will indicate the SNPs and location (chromosomes) of those SNPs with the highest disease association

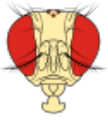
Which chromosome has the strongest association? _____

Which SNP has the strongest association? _____

Does this SNP necessarily cause the disease? Explain your answer.

***IMPORTANT NOTE*:**

Remember to return your assignment at the END of the lab session or to submit an electronic copy (i.e., scanned PDF file) through CANVAS



GENOME-WIDE ASSOCIATION STUDY (GWAS) WORKFLOW

