**NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

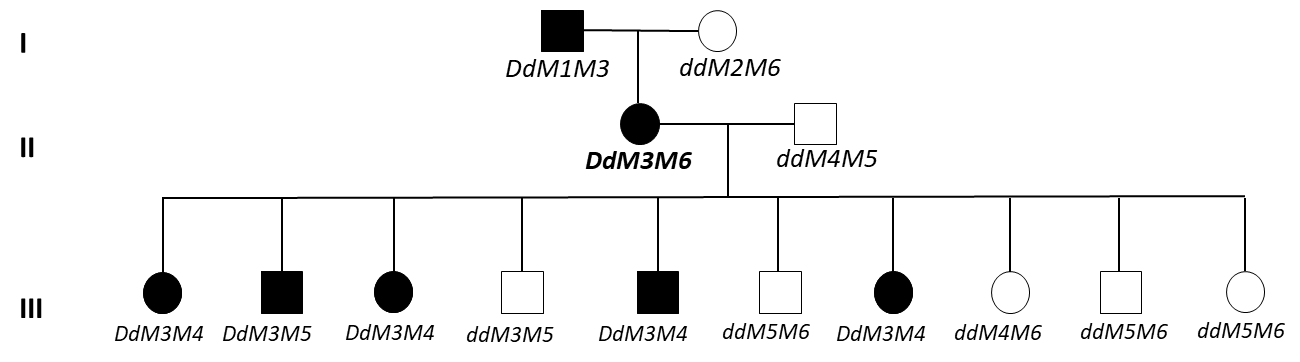
**NAME:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Student #\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Problem Set # 3**

This is the third of six problem sets that will count towards your final grade. The problem set is due **at the start of lecture on Friday, Oct. 11.** You may work in groups of up to three people (and are encouraged to do so). Please hand in one assignment per group with up to three names listed. Late assignments will be penalized 20% per day or part thereof. **Staple multiple pages together** –no paper clips or folded corners as sheets inevitably get lost. **Show your work.** Incorrect answers with correct work will receive part marks; correct answers with no work might not receive full marks.

1. The pedigree below shows the incidence of rare, autosomal dominant disorder called Ehlers-Danlos disease. The pedigree covers three generations of a particular family and also shows individual genotypes at a potential marker locus (M).

1. Indicate the phase of all gen II and III individuals.



1. Which, if any, of the gen III individuals are recombinants?
2. Calculate the LOD score as a test of physical linkage between the marker (M) and the disease locus.
3. What do you conclude about linkage between D and M?

2. DNA profiles can be used to identify unknown victims from a natural disaster. A DNA sample from a toothbrush of a potential victim had the genotype A1A4B2B3C4C5D1D1. Remains from an unidentified victim were found to have the same genotype. Given the allele frequencies in the relevant population in the table below:

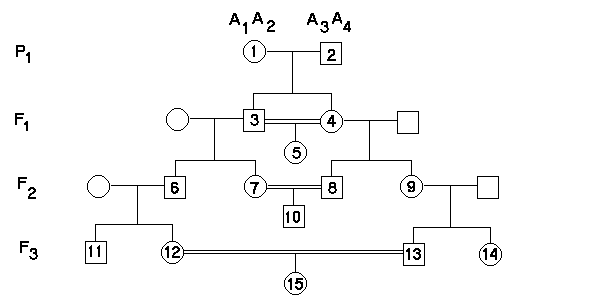
a) What is the probability that these two samples matches by chance?

b) List two assumptions you made in answering this question.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Locus | | | |
| allele | A | B | C | D |
| 1 | 0.2 | 0.1 | 0.15 | 0.7 |
| 2 | 0.1 | 0.6 | 0.1 | 0.3 |
| 3 | 0.05 | 0.1 | 0.1 |  |
| 4 | 0.25 | 0.05 | 0.3 |  |
| 5 | 0.40 | 0.15 | 0.35 |  |

3. Researchers are investigating a newly discovered condition that greatly increases the likelihood of lung damage associated with the use of e-cigarettes. They found that 2% of their study population has the condition. The researchers believe they have found a useful marker locus for assessing the presence of the condition in individuals with unknown family history. They have found that, of those individuals with the condition, 99% have the marker allele M1. Of those without the condition, 99.8% have an alternative allele M2 at the marker locus. What is the probability that an individual has the condition if they have marker allele M1 and what do you conclude about the utility of the marker locus for indirect diagnosis?

4) Consider the pedigree below. Double lines indicate matings between related individuals:



1. Calculate the inbreeding coefficient of individual 10 assuming none of their common ancestors are themselves inbred.
2. Calculate the inbreeding coefficient of individual 15 assuming *f*individual1 = 0.05.

5. In a population genetic study of a species of California wild oat, the following genotype frequencies were found for a trait (hairiness of the leaf sheath) which is controlled by a single locus:

HH 0.571 Hh 0.071 hh 0.358

1. What are the expected frequencies of these genotypes under Hardy-Weinberg?
2. Calculate F.
3. Briefly explain what the value in b) implies regarding heterozygosity and suggest what might be going on in this population to explain this result.

6. Consider a scenario in which allele A2 is recessive and it decreases fitness (i.e. it is deleterious). What is the ratio of the frequency of the deleterious recessive homozygote in an inbred versus a non-inbred (i.e. random mating) population if the frequency of the deleterious allele is *q* = 10-4 and *f* = 0.01?

7. This question will not be marked but I strongly encourage to you to do it nevertheless. It uses a simulation of genetic drift that may help your conceptual understanding of the process.

Download and install Populus (<http://cbs.umn.edu/populus/download-populus>). See here if you have problems installing or running it (<https://cbs.umn.edu/populus/contact>). Press the **Model** button and select **Mendelian Genetics,** then **Genetic Drift**. Select the **Monte Carlo** tab. This is a simulation method used to implement a Fisher-Wright model of drift. (You can read the model background in the Populus help document if you’re interested.) The initial dialog has values for run time (i.e. # of generations), population size, number of loci, a switch to set initial frequencies collectively or individually for each locus, and a switch to permit selfing or not.

a) Run simulations using the parameter value sets given in the table below and tabulate the results in the spaces provided. Set generation time to 600 (it auto-adjusts this if all alleles fix earlier). Set allele frequencies at the 6 loci collectively (i.e. the value in the initial frequency box will apply to all 6 loci). Do not allow selfing. Press the green double-arrow **View** button to see the results graph. **For a given set of parameters (i.e. one row below), run the simulation at least three times** to get an idea of how much each run can differ (the model can be rerun by clicking the blue arrow **Iterate** button in the results window). You can zoom by right-clicking on the plot; to zoom out choose ‘Options -> reset graph’. Enter the value from each of the three replicate runs in each cell, separated by commas as shown in the first example row below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Pop. size (*N*) | Initial freq. of *A* (i.e. *pA*) | # loci fixed for *A* | # loci fixed for *a* | # loci segregating | Time (gen) of 1st fixation of *A*a | Time (gen) of 1st loss of *A*b |
| 10 | 0.2 | 1,1,0 | 5,5,6 | 0,0,0 | 20,43,NA | 1,7,13 |
| 10 | 0.2 |  |  |  |  |  |
| 10 | 0.5 |  |  |  |  |  |
| 10 | 0.8 |  |  |  |  |  |
| 100 | 0.2 |  |  |  |  |  |
| 100 | 0.5 |  |  |  |  |  |
| 100 | 0.8 |  |  |  |  |  |
| 500 | 0.2 |  |  |  |  |  |
| 500 | 0.5 |  |  |  |  |  |
| 500 | 0.8 |  |  |  |  |  |

aThe time (in generations) at which the 1st fixation event of *A* occurs, if one occurs. If not, enter ‘NA’. bThe time (in gen.) for the first loss of *A*, if applicable. If not, enter ‘NA’.

b) What does each line on the graphs represent?

c) What is the relationship between population size and amount of time to fixation or loss based on your data?

d) How does initial allele frequency affect time to fixation or loss?

e) Why do the results of various runs for the SAME set of conditions differ?