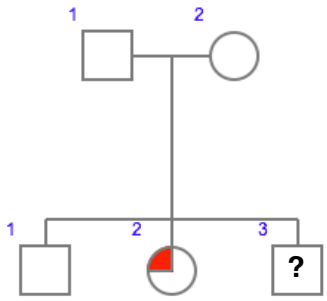
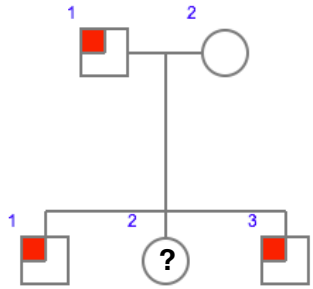
Bio393: Genetic Analysis **Name:**_____
Problem Set #4
Due on Monday, June 1, 2 PM

Question 1:
Each of the families below exhibits a different, extremely rare genetic disorder. Individuals expressing the trait (the disorder) are indicated by symbols with red sections. Assume that no new mutations have arisen in any of the individuals shown. Consider the following possible modes of inheritance: (i) X-linked recessive with complete penetrance, (ii) autosomal recessive with complete penetrance, (iii) autosomal recessive with 70% penetrance, (iv) autosomal dominant with complete penetrance, (v) autosomal dominant with 70% penetrance. For each pedigree state which, if any, of these five modes of inheritance are not possible. For the modes of inheritance that are possible, calculate the probability that the individual indicated by a "?" is affected.

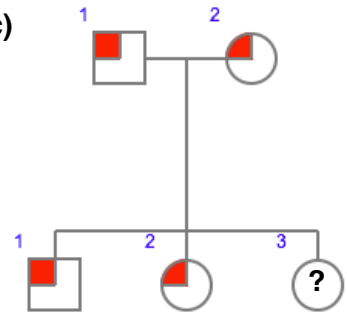
(a)



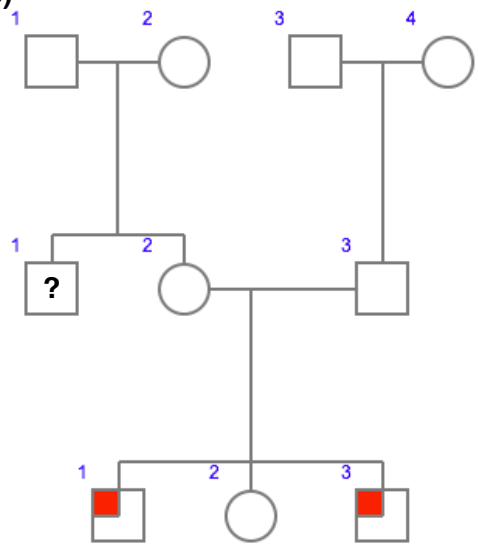
(b)



(c)



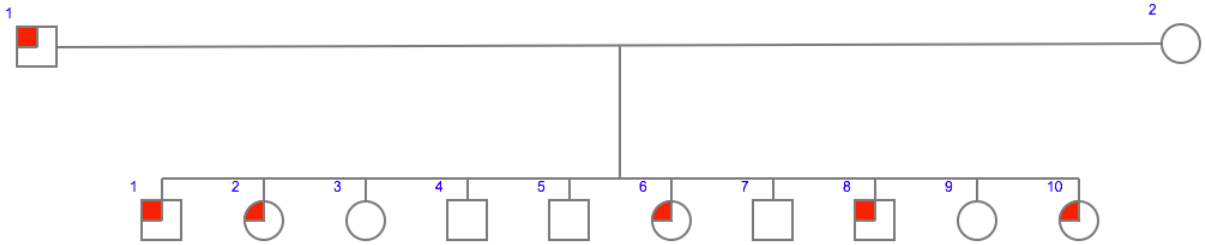
(c)

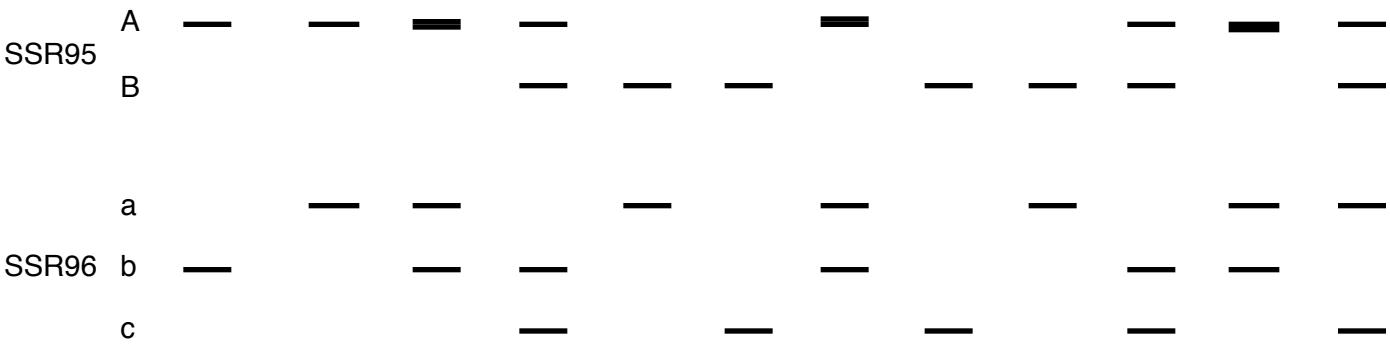


Question 2:
You set out to genetically map color blindness with respect to short-sequence repeat (SSR) markers. Color blindness shows X-linked recessive inheritance and therefore is usually found in males. However, the mutant allele frequency is sufficiently high that colorblind females do occur.

Alleles: + (normal) cb (associated with color blindness)

Here is a family in which some individuals are affected:





(a) Diagram the two possible phase relationships between the SSR95 and SSR96 alleles in the mother.

(b) Calculate the LOD score for linkage at $\theta = 0.1$ between SSR95 and SSR96 in this family.

(c) Identify a value of θ at which this family will yield a higher LOD score for linkage between SSR95 and SSR96. Calculate the LOD score for linkage between SSR95 and SSR96 at that new θ value.

(d) Diagram the two possible phase relationships between the SSR95 and color blindness alleles in the mother.

(e) Calculate a LOD score for linkage at $\theta = 0.1$ between SSR95 and color blindness in this family.

Question 3:
You are studying a rare recessive disease that you have mapped approximately by linkage to simple sequence repeat (SSR) markers. In an effort to localize the disease locus more precisely, you decide to look for linkage disequilibrium (LD) with respect to two dimorphic DNA-based markers (designated A and B) known to be in the vicinity of the disease gene. You first examine a relatively isolated Scandinavian population in which the frequencies of alleles A1 and A2 are 0.9 and 0.1 respectively, and the frequencies of B1 and B2 are both 0.5. By examining the DNA from individuals in the population who have the disease it is possible to determine the frequency of each haplotype, as shown in the table below.

Haplotype	Number of individuals with the disease
A1 B1	10
A1 B2	90
A2 B1	1
A2 B2	10

(a) (i) What can you say about possible linkage disequilibrium between each of the markers and the disease causing allele in this population? (ii) Assume the disease causing allele arose after both of the markers (A and B) were present in the population. Which of the two DNA-based markers is likely to be closer to the disease locus? (iii) Assuming that the disease allele arose only once in this population, what can you say about the haplotype context in which the original disease mutation arose?

(b) Next you examine the genotypes of individuals with the same disease in a large African population. In this population the frequencies of alleles A1 and A2 are both 0.5, and the frequencies of B1 and B2 are also both 0.5. The frequencies of the each haplotype for individuals with the disease in the African population are shown in the table below.

Haplotype	Number of individuals with the disease
A1 B1	26
A1 B2	24
A2 B1	28
A2 B2	22

Give two different explanations for why the linkage disequilibrium results differ between the African and Scandinavian populations.

Question 4:
You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. You have 1000 cases, (480 of which are AA, 400 are AG, and 120 are GG at rs4514) and 1000 controls, (360 of which are AA, 440 are AG, and 200 are GG at rs4514).

(a) Using a chi-squared test, what is the p-value of the association of these alleles with the disease.

(b) Given that you did 500,000 tests, what is your (Bonferroni) corrected threshold for p-value significance (initial $\alpha=0.05$)? Does the rs4514 variant pass "genome-wide significance" for association with Type 2 Diabetes?

(c) What is the odds ratio of this variant in a risk for Type 2 Diabetes?