

Problem Set 1
BENG 183
Fall 2011

Due Tuesday October 4th 2011, 11:59 PM

Deliver in class Tuesday or submit by e-mail to qlma@ucsd.edu.

- 1.) Genetics and probability. Consider two markers 2 cM apart.
 - a.) What is the probability that there will be recombination between them in one generation?
 - b.) What is the probability that there will NOT be recombination between them in one generation?
 - c.) What is the probability there will NOT be recombination after 10, 20, 30, 40, and 50 generations?
 - d.) What is the probability of exactly three recombination events between the markers in the same generation?
- 2.) Accurate calculation of large genetic map differences using mapping functions. In computing the genetic distance between two markers A and B, we discussed in class how the recombination fraction (RF) is used for this purpose. For genetic distances greater than a few cM, however, the RF begins to underestimate the true genetic distance.
 - a.) Explain why this underestimation occurs.

For longer genetic distances, mapping functions allow geneticists to use the observed recombination fraction (RF) for two markers A and B to solve for the average or expected number of crossing over events m between these markers. The average number of crossover events is, in essence, a “corrected” genetic mapping unit. Such a mapping function is based on the Poisson distribution:

$$f(k; m) = (m^k * e^{-m}) / k!$$

where k is the actual number of crossing over events observed in a single trial.

- b.) What is the probability that, in any given meiosis, there will be no crossover events between A and B? Express your answer in terms of m .
- c.) What is the probability of one or more crossover events? Graph this probability as a function of m .

One or more crossover events exactly corresponds to a recombination fraction (RF) of 50% (To see why visit:

<http://www.ncbi.nlm.nih.gov/books/NBK21819/figure/A1116/?report=objectonly>). Therefore, the function you have just graphed is a mapping function to convert from RF to m .

d.) Suppose that we have an RF for two markers of 27.5%. What is the average number of crossovers between the two markers this represents?

3.) You are studying soybeans, in an ongoing effort to make a new kind of tofu that tastes like delicious chocolate. Part of your work involves mapping loci by crossing different strains. Your first cross involves loci S, for size (ss plants are stunted), and A, for food allergen (aa plants are hypoallergenic). You cross a heterozygous plant with genotype AS/as, and a homozygous plant of genotype as/as.

Phenotype	Count	Expected (Mendelian)
A/S	467	?
A/s	32	?
a/S	38	?
a/s	468	?

a.) What numbers would correctly fill the right-hand column in the above table?

b.) Calculate the recombination distance between these two loci.

A second cross considers an additional locus: T, for tasty (tt plants taste bad). One parent plant has genotype AST/ast, the other is ast/ast. (HINT: This is a three point cross).

Phenotype	Count	Expected (Mendelian)
A/S/T	465	?
A/S/t	1	?
A/s/T	19	?
a/S/T	15	?
A/s/t	13	?
a/S/t	20	?
a/s/T	1	?
a/s/t	466	?

c.) What numbers would correctly fill the right-hand column?

d.) What is the order of the three loci along the chromosome?

e.) What are the distances between loci?

f.) Why is the computed value for distance between A and S different in this case? (Assume

it's not because of random variation between experiments)

4.) In class we learned that denatured single-stranded DNA reassociates through defined kinetics, governed by the differential equation:

$$\frac{dC}{dt} = -k \cdot C^2$$

where C is the amount of single-stranded DNA remaining at time t.

a.) Assuming standard metrics, what are the units for k?

b.) Integrate the above differential equation to solve for C / C_0 , the percentage of single stranded DNA remaining, starting from an initial ssDNA concentration of $C = C_0$. For full credit, show each step of your work.

c.) Given $k = 0.5$, plot the famous "COT curve". This curve is drawn on an xy axis, with $y = C / C_0$ and $x = C_0 * t$.

d.) $t_{1/2}$ is called the "reassociation half time", defined as the time at which exactly half of the DNA has reassociated to become double stranded. Solve for $t_{1/2}$ in terms of k. Show each step of your work. Label the reassociation half time on your COT curve from part c.

5.) Chinese Hamster Ovary (CHO) cells are an important cell line for medical research and therapeutic production. Its genome is large and includes many repeats.

Sequencing these repeats is difficult and unrewarding, so researchers are filtering out repetitive sequences through High-Cot filtering.

Assume you have a heated mixture of repetitive DNA, X, and non-repetitive DNA, Y. The initial concentrations X_0 and Y_0 are 8 uM and 3 uM, respectively. The re-association rate constants k_x and k_y are 1000 unit and 50 unit, respectively.

a.) What fraction of the initial mixture is non-repetitive DNA?

b.) What will, the concentration of repetitive DNA, be after three hours?

c.) What will, the concentration of non-repetitive DNA, be after three hours?

d.) After three hours, double-stranded DNA is removed and the remaining single stranded DNA is used for cloning. What fraction of this single-stranded DNA is non-repetitive?

e.) You decide that you want to recover 70% of the non-repetitive single stranded DNA for cloning. How long should you let your mixture of DNA anneal?

6.) Splicing is the process by which pre-mRNA is modified to remove certain stretches of non-

coding sequences called introns; the stretches that remain include protein-coding sequences and are called exons. Sometimes pre-mRNA messages may be spliced in several different ways, allowing a single gene to encode multiple proteins. This process is called alternative splicing.

- a.) A gene contains 85 exons. How many different proteins can this system produce?
- b.) A gene contains 85 exons of which exactly 24 are to be retained in the final mRNA. How many different proteins can this system produce?
- c.) For a gene with 2 exons, what chemical species can be used as evidence that splicing has occurred (other than the final transcript)?

Because there are four nucleotides in DNA, adenine (A), guanine (G), cytosine (C) and thymine (T), there are 64 possible triplets encoding 20 amino acids, and three translation termination (nonsense) codons. Because of this degeneracy, all but two amino acids are encoded by more than one triplet. Different organisms often show particular preferences for one of the several codons that encode the same amino acid.

- d.) *T. maritima* is an organism that thrives at 90 degrees C (very hot). How might it choose to code for the peptide “MAGTIDE” if denaturation of the genome is a major concern for survival? Give the DNA sequence and directionality.

7.) In order to better understand genome-wide association study publications such as the 23andMe research article in PLoS Genetics (Eriksson et al. 2010), some genetic terms must be reviewed.

- a.) What is the difference between a mutation and a polymorphism?
- b.) What are the three types of single nucleotide polymorphisms (SNPs)? How can intergenic SNPs influence cellular function?
- c.) How can haplotypes affect determining proper causality of SNPs?

8.) P. 31 of the book describes a concept known as Linkage Disequilibrium (LD). Consider two loci with allele frequencies:

$$p(A) = 0.8$$

$$p(a) = 0.2$$

$$p(B) = 0.7$$

$$p(b) = 0.3$$

Calculate the expected frequencies of the genotypes AB, Ab, aB, and ab assuming no LD.