Problem 1: Sanger sequencing problems galore!

When you do Sanger sequencing, the end product is a gel like the ones shown below (Figure 1 & 2):

FIGURE 1

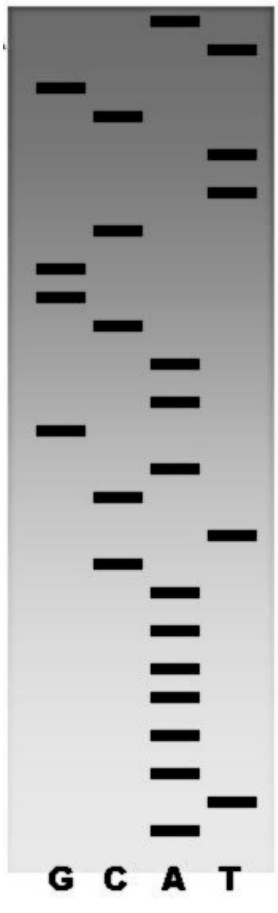


FIGURE 2

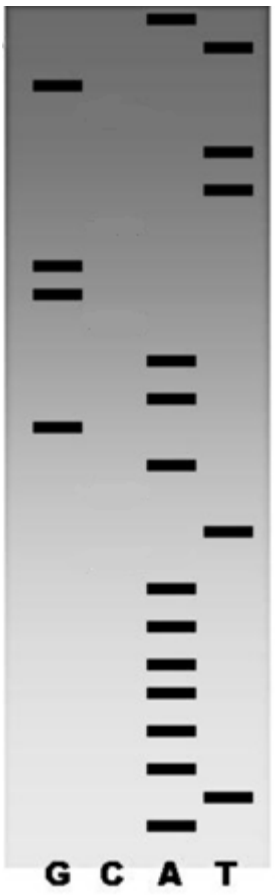
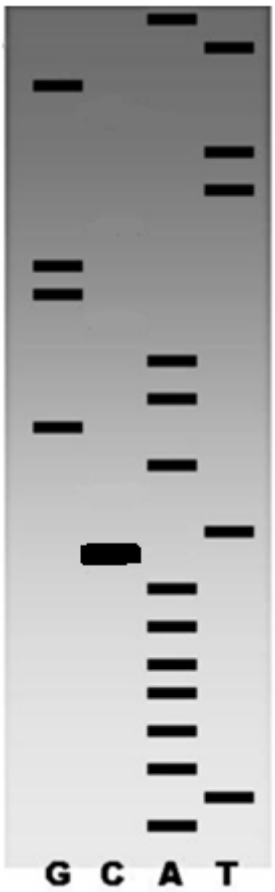


FIGURE 3



a) What are the first three bases of the template strand on the 5' end (Figure 1)? How do you know? *Note, the lane labels correspond to the radiolabel added (G = ddGs added).

b) Your advisor tells you to confirm the exact same sequencing experiment with another run, and you get the picture above (Figure 2). Provide one possible explanation for this anomaly.

c) Your advisor tells you to get it right this time. You get the picture above (Figure 3). Provide one possible explanation for this anomaly. Hint: Perhaps this is an overreaction to an earlier mistake.

d) If the same experiment were run in a thicker matrix for the same amount of time, what changes would you expect to see?

e) Derive a general formula for the probability of seeing a fragment of length N where the fraction of ddNTPs in the mix (versus all NTPs) is P? What kind of distribution is this?

Problem 2: Pyrosequencing

One difficult problem in sequencing is handling heterozygous loci. Typically, PCR amplification of an exon from a heterozygous individual will produce both alleles in roughly equal amounts. You use pyrosequencing on a short piece of DNA that has a known SNP and get the following result:

Base added	A	C	G	T	A	C	G	T	A	C	G
Peak Intensity	0	2	3	1	.5	0	.5	1	0	3	0

a) What are the 2 alleles?
b) For a DNA fragment of length N, what is the smallest peak intensity possible? What is the largest possible?
c) What would a SNP look like in a four-dye chromatogram?

Problem 3: Restriction enzymes

The discovery of restriction enzymes greatly advanced our ability to clone DNA fragment inserts into plasmids for amplification. Let's have some practice to familiarize ourselves with what we expect to see from experiments that utilize restriction digestion.

a) You use Sma I, Pst I, and Xho I to digest a circular plasmid completely. The restriction map of the plasmid is shown in Figure 4. Please draw the band patterns you anticipate to see on a slab gel for the following experiments: (1) Sma I only, (2) Pst I only, and (3) Sma I and Xho I.

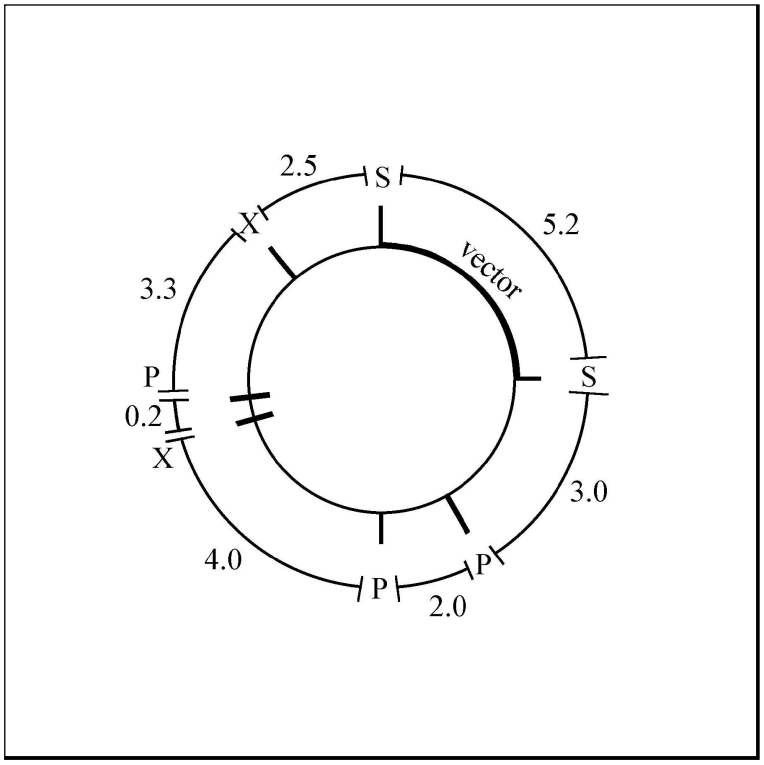


Figure 4. The restriction map of an artificial plasmid.

b) A preparation of linear target DNA to be sequenced is digested to completion with the following combination of enzymes. Gel electrophoresis reveals the size of the fragments after digestion. Using the information given below, construct a restriction map of this DNA.

Enzyme used	**Fragment sizes (kb)***
Bgl II	5 and 10
Hga I	5 and 10
Sma I	2 and 13
Bgl II and *Hga I*	5
Bgl II and *Sma I*	2,5 and 8
Hga I and *Sma I*	2, 3 and 10

*kb is kilobase pairs

c) You mix the above plasmid and linear target DNA in the same tube. You then perform restriction digestion with Sma I followed by sequence ligation. What size(s) of plasmids result?

d) After transfection of a plasmid into E. coli followed by streaking to form single colonies on the proper selective media, how long would you have to culture the cells in order to obtain about 10^9 copies of the plasmid in a single colony?

Problem 4: Solexa and Illumina Sequencing

a) If you have a “finished” human genome sequence where all of the genome sequence is at exactly Q40 accuracy, how many errors occur in your whole genome sequence? How about if it was completed to exactly Q30 accuracy?

In class, we covered how Solexa / Illumina sequencing works in basic mode to yield short reads of 35-70 bp.

b) Through your own reasoning or web research, what is the fundamental reason why Solexa reads are short?

c) Many people use a Solexa machine in "paired end" mode to capture short reads from both ends of a longer cloned sequence. Using your own reasoning, draw a diagram showing how this works along with accompanying text to explain the procedure. As long as you explain yourself and your approach makes sense, you will not be penalized for not guessing the "True" method used by Illumina.