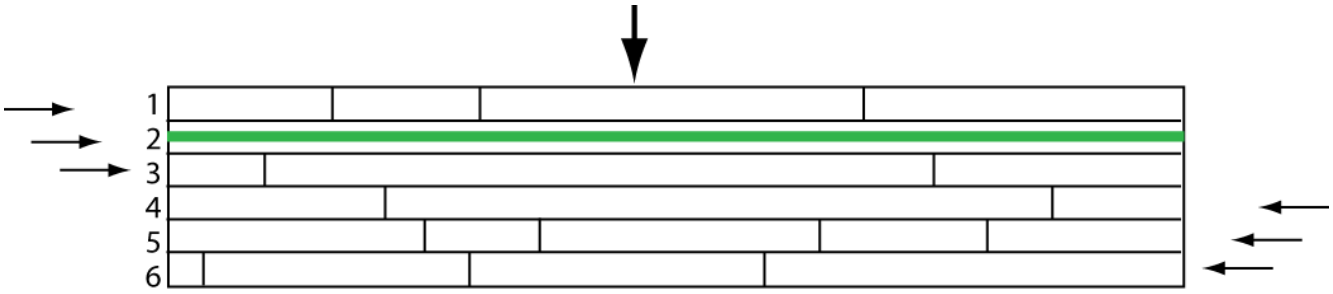
7.03 Problem Set 3

Due before 3 PM on Wednesday, March 16

Hand in answers in recitation section or in the box outside of 68-120

1. The diagram below shows a 300 base pair (100 codon) segment in the middle of a bacterial gene displayed in a six-frame translation program similar to StarORF. Rows 1 - 3 represent the three reading frames read from left to right and rows 4 - 6 represent the three reading frames read from right to left. The arrows show the relationships among the three reading frames in each direction. Vertical marks in each row represent the positions of stop codons in that reading frame. Row 2 has been highlighted in green because it is the open reading frame for the gene.



Reproduce the diagram for this section of the gene showing the positions of all of the stop codons that you know must exist as they would appear in all six frames after each of the following mutations occurred.

a) An amber mutation in the gene at the position of the vertical arrow.

b) An insertion of a single base pair at the position of the vertical arrow.

c) An insertion of two base pairs at the position of the vertical arrow.

d) An insertion of three base pairs at the position of the vertical arrow.

e) Deletion of a single base pair at the position of the vertical arrow.

f) Deletion of two base pairs at the position of the vertical arrow.

2. The base pairing rules constrain the composition of DNA such that the frequency of G = C and the frequency of A = T. However the frequency of G•C base pairs is not constrained to be equal to the frequency A•T base pairs. Although most bacteria have do have a balanced DNA composition in which each of the bases is present at about the same frequency, bacteria that live in environments where they are exposed to heat or UV rays usually have less than 50% A•T in their genomic DNA. This is because thymine (T) residues are particularly susceptible to damage by heat and UV and are selected against under these conditions. Some bacteria which have evolved to live at high temperatures can has an A•T content of only 40%.

a) Calculate the probability of a stop codon in a random DNA sequence as a function of A•T content. A good way to start this problem is to consider that in random DNA sequence (which has 50% A•T

content) the probability of each of the three stop codons is 1/64, and then to develop an expression for the probability of each of the three stop codons as a function of A•T content.

b) Use the value calculated in part **a)** to calculate the probability of an ORF of 200 codons or more in random sequence as a function of A•T content.

c) A consider typical bacterium with a genome of 5 Mbp which encodes a total of 4,000 actual genes (of length greater than 200 codons). Using the probability calculated in part b), calculate the number of fortuitous ORFs of more than 200 codons (which would be difficult to distinguish from real genes) that would arise in a 5 Mbp genome of as a function of A•T content. Next, calculate the proportion of ORFs greater than 200 that are false positives i.e. not real genes that you would expect to find expressed as a function of A•T content. Finally, given the choice would you rather search for real genes in an organism of relatively high A•T content or low A•T content? Explain.

3. The codon for tryptophan is 5'UGG3'. Write out the double stranded sequence of the DNA that codes for the anti-codon loop of the tRNA for tryptophan. Indicate which of the two strands is used as the template for the transcription of the tRNA. Consider an amber suppressor allele of the tRNA for tryptophan. Write out the double stranded sequence of the DNA that codes for the anti-codon loop of the <u>suppressor</u> allele of the tRNA gene for tryptophan.

4. a) You have isolated a new *E. coli* mutant called sm-1 that makes small colonies. You isolate a Tn5 insertion, which carries resistance to kanamycin (Kan^r), that you think may be linked to the sm-1 locus. To test for linkage you grow P1 phage on the Tn5 insertion strain and use the resulting lysate to infect an sm-1 mutant strain. Among 50 Kan^r transductants isolated, 40 are small and 10 are normal. Express the distance between the Tn5 insertion and the sm-1 mutation as a cotransduction frequency using the Poisson distribution to provide error estimates for your measurement.

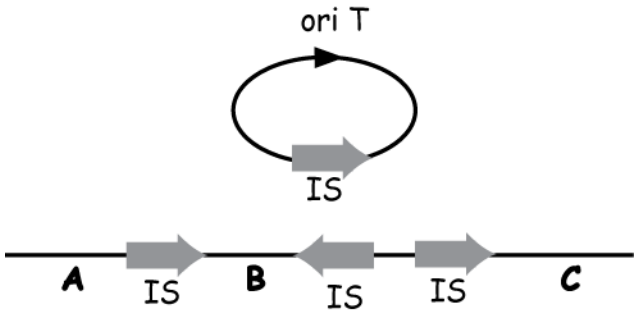
b) Next you isolate a second rough mutant called sm-2. To test for linkage to the same Tn5 insertion as in part **a)**, you grow P1 phage on the Tn5 insertion strain and use the resulting lysate to infect an sm-2 mutant strain. Among 50 Kan^r transductants isolated, 41 are small and 9 are normal. Express the distance between the Tn5 insertion and the sm-1 mutation as a cotransduction frequency with error estimates for your measurement. From these data can you tell the relative order of sm-1, sm-2 and the Tn5 insertion? Explain.

c) To map these mutations relative to one another you grow P1 on a strain with the sm-2 mutation and the Tn5 mutation and then use this lysate to infect an sm-1 mutant selecting for Kan^r. From 100 Kan^r transductants examined, 94 have small colonies and 6 have normal colonies. To do the reciprocal cross you grow P1 on a strain with the sm-1 mutation and the Tn5 mutation and then use this lysate to infect an sm-2 mutant selecting for Kan^r. From 100 Kan^r transductants examined, all have small colonies. Use this information to order the sm-1, sm-2 mutations relative to the Tn5 insertion. Think carefully about how to express the co-transduction distances that you can include in your map.

d) You isolate a Tn10 insertion, which carries resistance to tetracycline (Tet^r), which you think may be near the sm-1 and sm-2 mutations. You grow P1 phage on the Tn10 insertion strain and use the resulting lysate to infect an sm-1 mutant. Among 50 Tet^r transductants isolated, all are small. You also use the same lysate to infect a strain carrying the Tn5 insertion. In this case, from 50 Tet^r transductants isolated, 45 are Kan^r and 5 are kanamycin sensitive. Combine all of this information to

make a map showing the order of the sm-1 and sm-2 mutations as well as the Tn5 and Tn10 insertions. Also give any map distances as co-transduction frequencies.

5. The diagram below shows the F factor and a portion of the *E. coli* chromosome that has three different insertion sequences (IS) of the same type as is carried on F.



Describe how you would isolate an F' factor that carries the B gene. You may assume that each of the markers A^+, B^+ and C^+ can be selected for and that you have mutant strains carrying A^-, B^- and C^- mutations. [Hint: To make the desired F' will require two steps. First you will need to isolate an appropriate Hfr strain. Unfortunately this will require a laborious screen testing one Hfr at a time for the desired transfer properties. If you set up the initial Hfr correctly, you should be able to isolate the desired F' by a relatively easy selection.]