

Name: \_\_\_\_\_

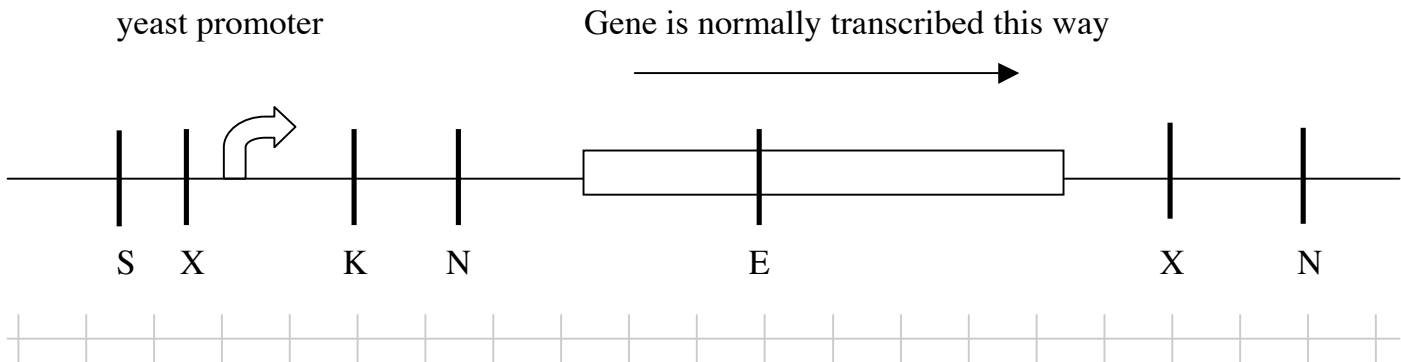
## 2006 7.012 Problem Set 4

Due before 5 PM on FRIDAY, October 27, 2006.

Turn answers in to the box outside of 68-120.

PLEASE WRITE YOUR ANSWERS ON THIS PRINTOUT.

1. You are studying a specific gene in yeast, and you want to express that yeast gene in *E. coli*. Your task is to design a strategy to insert the yeast gene into the bacterial plasmid. Below is a map of the area of the yeast genome surrounding the gene in which you are interested.

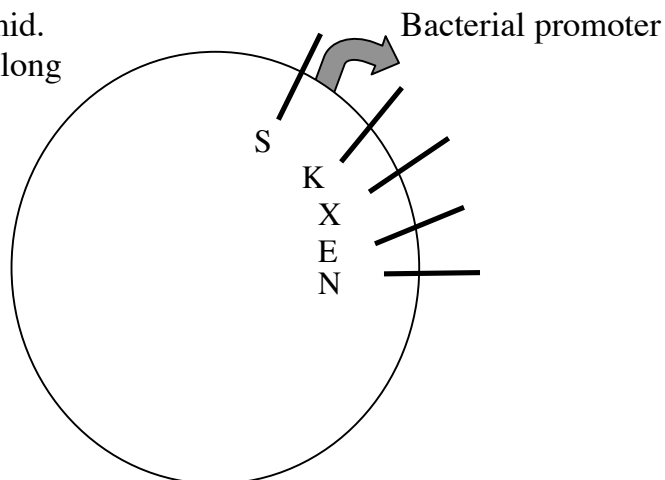


The distance between each tick mark placed on the line above is 100 bases in length.

Below are the enzymes you can use, with their specific cut sites shown  $5' - \text{XXXXXXX} - 3'$   
 $3' - \text{XXXXXXX} - 5'$

<i>Xba</i> I:	<i>Nde</i> I:	<i>Sal</i> I:	<i>EcoR</i> I:	<i>Kpn</i> I:
$\downarrow$ TCTAGA AGATCT $\uparrow$	$\downarrow$ CATATG GTATAC $\uparrow$	$\downarrow$ GTCGAC CAGCTG $\uparrow$	$\downarrow$ GAATTC CTTAAG $\uparrow$	$\downarrow$ GGTACC CCATGG $\uparrow$

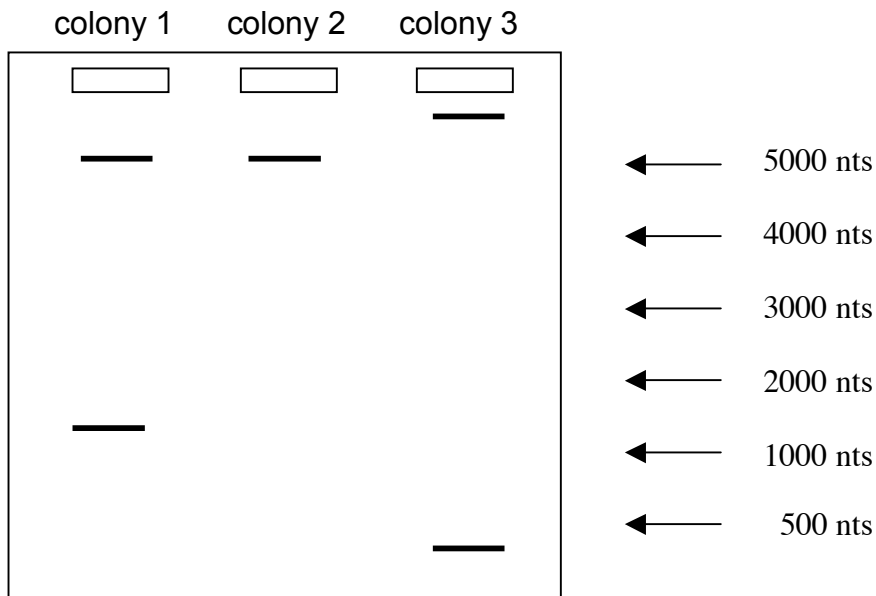
This is the map of the plasmid. The plasmid is 5,000 bases long and the two farthest restriction enzyme sites are 200 bases apart. The plasmid has an ampicillin resistance gene somewhere on the plasmid distal from the restriction cut sites.



Name: \_\_\_\_\_

(a) Which single restriction enzyme is the best choice for you use to design a way to get the insert into the vector if you can only use one single enzyme? Keep in mind that your goal is to have the yeast gene be expressed in the bacterial cells.

(b) You follow the cloning strategy you chose in part (a). You do the digestion of the insert and the vector and then ligate the two digestions together. You then transform the ligation into bacteria and select for ampicillin resistance. You get three colonies on your transformation plate. You isolate plasmid from each one and cut each plasmid with the enzyme XbaI. You then run your three digestions on an agarose gel and see the following patterns of bands. Describe what each plasmid actually was that was contained in each of the three colonies.



Colony 1's plasmid =

Colony 2's plasmid =

Colony 3's plasmid =

Name: \_\_\_\_\_

(c) Which colony's plasmid do you actually want to use for your studies?

(d) Which two restriction enzymes would you use to design a way to get the insert into the vector if you had to use two enzymes simultaneously?

(e) You transform your ligation planned in part (d) into bacteria and plate the bacteria on Petri plates containing ampicillin. (You actually transform six different ligation mixtures, which are described below, into six different populations of cells, and plate each transformation onto a different plate, because you want to do all of the correct controls.) The next day you come in to lab to look at how many colonies of bacteria are on each plate. You are really excited, because the number of colonies you see on each plate tells you that the entire procedure worked! Which of the three following patterns of number of colonies did you see in order to conclude that you had a successful transformation? Circle the correct pattern.

In this table, DV = digested vector. DYG = digested yeast genome.

<u>Ligation Used</u>	<u>Pattern 1</u>	<u>Pattern 2</u>	<u>Pattern 3</u>
DV only + Ligase	200	0	0
DYG only + Ligase	0	200	0
Water + Ligase	0	0	0
DV + DYG + Ligase	200	200	200
DV + DYG (NO ligase)	0	0	0
Undigested vector + Ligase	0	0	200

Name: \_\_\_\_\_

**2.** You are practicing designing primers that you can use in PCR reactions. You want your primers to allow you to amplify the sequence found below.

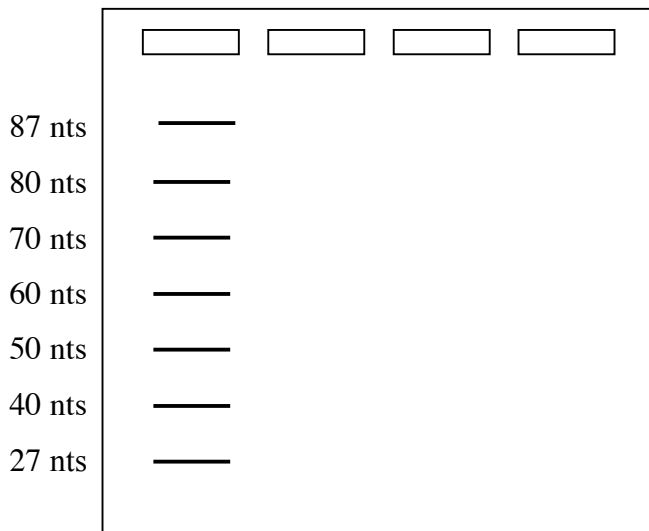
5' - ACTTCGATATGTCTAAAATACGATCGATCTGTGGGGCCTAGCTAGCTAACCAGAGACGCTACCG - 3'  
3' - TGAAGCTATACAGATTTTATGCTAGCTAGACACCCCGGATCGATCGATTGGTCTCTGCGATGGC - 5'

**Left primer should  
anneal to this region**

**Right primer should  
anneal to this region**

Draw into the following gel lanes what size(s) of PCR products you would get if you used the following primers stated in parts (a), (b), and (c) to do a PCR reaction on the template DNA shown above.

Markers    part (a)    part (b)    part (c)



(a) 5'-ACTTCGATATGTCTAAAATAC-3' and 5'-CGGTAGCGTCTCTGGTTAGCT-3'

(b) 5'-TGAAGCTATACAGATTTTATG-3' and 5'-GCCATCGCAGAGACCAATCGA-3'

(c) 5'-GTATTTTAGACATATCGAAGT-3' and 5'-AGCTAACCAGAGACGCTACCG-3'

(d) You are asked to design a 15-nucleotide-long primer that could potentially hybridize to a portion of a specific mRNA that encodes the protein sequence N-Met-Ala-Tyr-Trp-Pro-C. How many different primers would you have to design in order to ensure that one of them will in fact hybridize along its full length to the mRNA?

Name: \_\_\_\_\_

**3.** You are a scientist who is using genomics to currently study a new bacterial species that no one has ever studied before. The following sequence is a piece of DNA within the coding region of a gene that you have recently sequenced.

5' -CCCGTACGTTTACGCCGTATATATCGTCGTAATCCTACGTAGCTCTACGAACA-3'  
3' -GGGCATGCAAATGCGGCATATATAGCAGCATTAGGATGCATCGAGATGCTTGT-5'

**(a)** If you take any bacterial gene sequence, before you begin doing any sequence analysis on it, there are six potential open reading frames. Why are there six?

**(b)** How many of the 6 potential open reading frames are actually open in this sequence shown above?

**(c)** You are using shotgun sequencing to determine the DNA sequence of the genome of this new bacterial species. For one strand of a 30-nucleotide long stretch of DNA, you get the following sequences out of your shotgun sequencing reaction. Assemble the entire 30-nt-long DNA sequence you are trying to sequence, and write the full sequence of the DNA. You only need to write the one strand that is shown; please make sure to label the 5' and 3' ends of that one strand.

5' -GGAGTTCCTC-3'  
5' -CGCGTTGTCACTGAC-3'  
5' -TGGGAGT-3'  
5' -TCCTCAAACGCGTTGT-3'

You put the DNA sequence that you have assembled in part **(c)** into a computer program that tells you that the following piece of DNA, which comes from another bacterium, is a close match to the sequence you have sequenced from your bacterium:

5' -...TGGGCATTTCTCAAGCGGGTTGTAATGGAT...-3'

This 30-nt-long sequence fragment lies in the center of a gene, and that portion of the sequence encodes for this 10-amino acid-long part of a protein:

N-...Trp-Ala-Phe-Leu-Lys-Arg-Val-Val-Met-Asp...-C

Name: \_\_\_\_\_

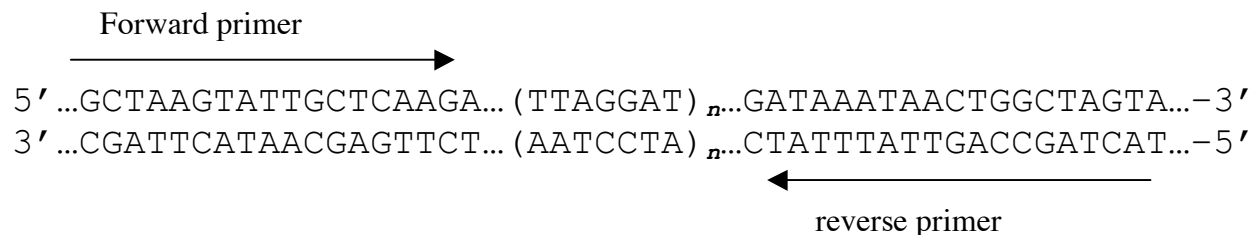
You hypothesize that the sequence you have discovered is another bacterial species' version of the same gene as this previously known gene. To measure how identical the two genes are at the DNA level and/or the two proteins are at the amino acid level, you can calculate a percentage of "identity" for each. This is the percent of nucleotides (for the gene) or the percent of amino acids (for the protein) that are identical between the two sequences.

(d) What is the % identity between the two DNA sequences?

(e) What is the % identity between the two protein sequences?

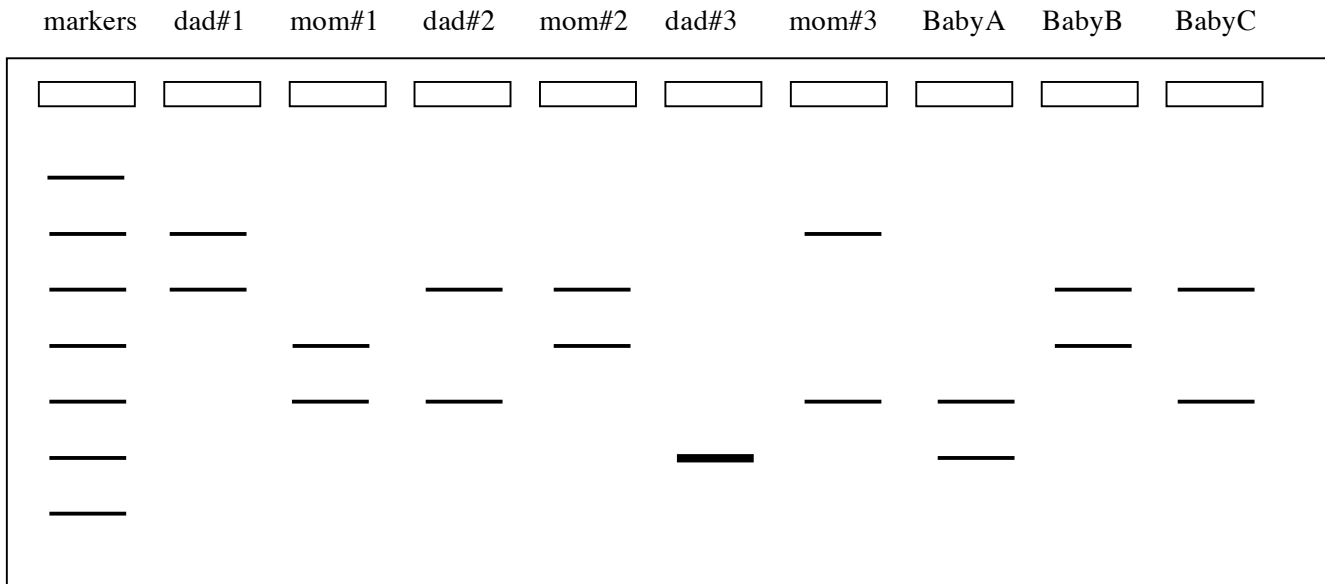
**4.** One of the most common applications of using recombinant DNA techniques like PCR, restriction enzyme digests, and agarose gels is to test samples of human cells such as blood cells to identify people for forensic analysis or paternity testing. This problem is designed to show you how this type of analysis, called DNA fingerprinting, can be used to determine paternity. There are three babies (Baby A, Baby B and Baby C) in a maternity ward, and three sets of confused and worried parents. (Father and Mother #1 are a couple, as are Father and Mother #2, and Father and Mother #3.) This problem will show you how to figure out which baby goes with each set of parents.

As we have spoken about in class, most of the human genome (>95%) is not genes. Most of the DNA sequence differences between humans are found in these non-coding regions. Some of these non-coding regions are just series of DNA sequence repeats found over and over again. Different humans can have different numbers of repeats at these regions (i.e. "n" shown below can vary). The way you assay how many repeats someone has is by doing a PCR reaction using his/her DNA as a template. The primers are designed to the sequences flanking the repeated region. For instance, take the DNA sequence below. Say it is found somewhere on chromosome #15. Different humans differ by the value of "n" (the number of TTAGGAT repeats). You do each PCR reaction and load each one into a separate well of an agarose gel, and then run the gel.



Name: \_\_\_\_\_

You obtain the following results.



(a) Why is it that having more repeats leads to a band that is higher in the gel?

(b) Why is it that some people show only one band?

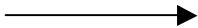
(c) Why is it that some people show two bands?

(d) Given the data so far, for which of the three babies (only Baby A, only Baby B, or only Baby C, all of them, or none of them) can you already conclusively tell who the parents are?

Name: \_\_\_\_\_

Other parts of the non-coding regions in our genome are not genes but they are also not regions of repeats. Humans can vary by DNA sequence at these sites, instead of varying by number of repeats in a row. For instance, take the DNA sequence below. Say it is found somewhere on chromosome #7. Different humans differ by which basepair is found at the position marked in bold below; some people have a T-A basepair, whereas others have an A-T basepair at this bolded position. It just so happens that one version of this site can be cleaved by a restriction enzyme that recognizes the sequence 5'-TTGCAA-3' and cuts between the two Ts. The way you assay which sequence someone has at this region is by doing a PCR reaction using his/her DNA as a template. The primers are designed to the sequences flanking the site of variable sequence; each primer is about 20 nucleotides long. You then treat the PCR product with the restriction enzyme and run the products of the digestion reaction on an agarose gel.

Forward primer

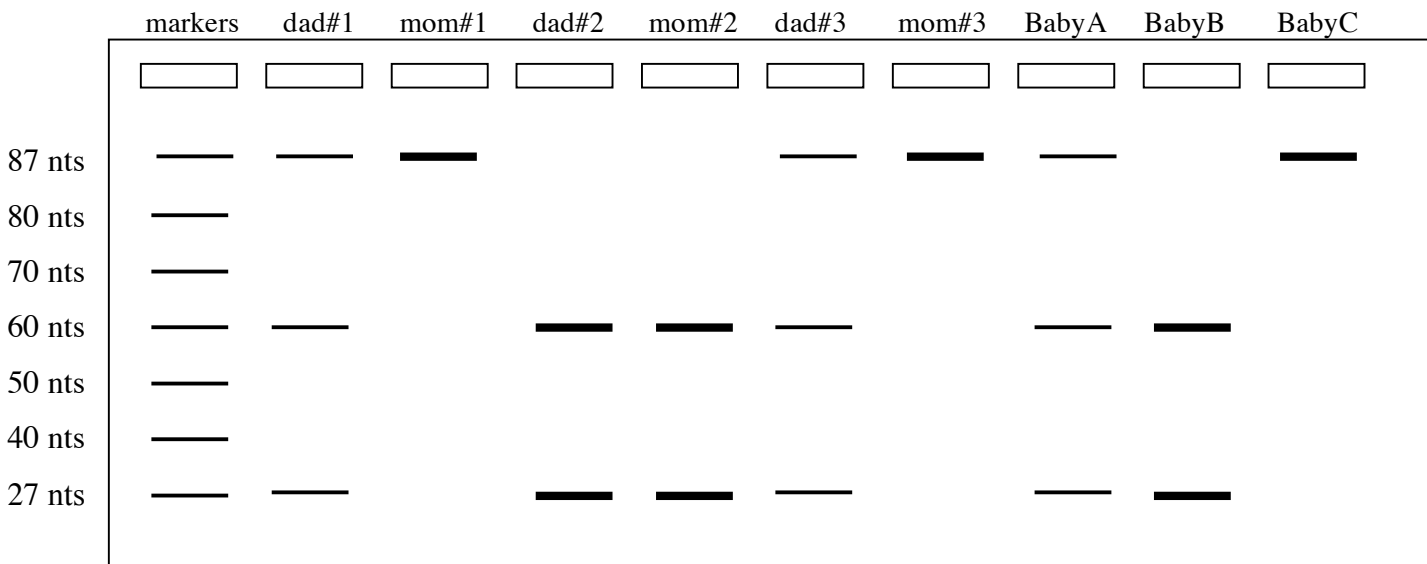


5' -...GATATC**TT**GCAAGTCCATCCTGCATGCACATGCTGATACGCGCAACGGT...-3'

3' -...CTATAG**A**ACGTTTCAGGTAGGACGTACGTGTACGACTATGCGCGTTGCCA...-5'

Reverse primer

You obtain the following results.



(e) Why is it that some people only have one band?

(f) Why is it that some people show two bands?

**Name:** \_\_\_\_\_

(g) Why is it that some people show three bands?

(h) Match the babies up with their actual parents.

**BABY**

**PARENTS**

Baby A

Father and Mother #1

Baby B

Father and Mother #2

Baby C

Father and Mother #3

(i) Now go back and look at the first gel drawn in the problem, and answer the question: which number of repeats did Baby C inherit from his mother?

**5.** Where in a eukaryotic cell do you think you would find each of the following proteins residing when it is actively performing its function? Be as specific as you can in terms of subcellular location.

(a) DNA polymerase

(b) RNA polymerase

(c) the ribosomal proteins

(d) DNA ligase

**Name:** \_\_\_\_\_

**(e)** helicase

**(f)** an activator protein

**(g)** a repressor protein

**(h)** an enzyme in the glycolysis pathway

**(i)** an enzyme in the Krebs/TCA cycle

**(j)** a protein that allows ions to pass in and out of the cell

**(k)** enzymes that splice mRNAs

**(l)** a protein that forms a channel through which mRNAs can be exported into the cytoplasm