Assignment 1

Part I: Dilutions and Concentrations **(3 points/ question for a total of 60 points)**

Your answers must be submitted to two significant figures after the decimal. For example 2.00, 0.020, or 0.0020.

Use the following information to answer questions 1-8

You prepare a solution "X" by adding the following ingredients to 1L of H_2O: 125 mL of solution "A" (1.6 M compound "A"), 50 mL of solution "B" (20 % (m/v) compound "B"), and 25 mL of solution "C" (10g/L of compound "C".) The properties of each ingredient are as follows:

Compound A: MW 200g/mole, density 1.2g/mL, density of 1.6 M soln. 1.05g/mL

Compound B: MW 150g/mole, density 1.3g/mL, density of 20% soln. 1.1g/mL

Compound C: MW 35g/mole, density 1.15g/mL, density of 10g/L soln. 1.03g/mL

Density of solution X: 1.25g/mL

1. By what factor was solution "A" diluted?
2. Given that the volume of solution "B" represented 2 parts, how many parts total are in solution "X"?
3. What is the final molarity of compound "B" in solution X?
4. What is the percentage (m/m) of compound "C" in solution X?
5. What is the percentage (m/v) of compound "A" in solution X?
6. What is the percent of total solutes (m/v) in solution X?
7. What is the parts ratio between solutions "A", "B", "C" and water? (consider your answer for Q. 2)
8. What is the mass ratio between compounds "A", "B", "C" and water? Round off to the nearest whole number.

9. You start with 0.5 L of a KNO_3 stock solution and wish to prepare 10.0 L of 0.1 M KNO_3. What percentage (m/v) would the potassium nitrate stock solution need to be if you were to use it all? (MW of KNO_3: 101g/mole)
10. You have a 2.0 M copper sulfate stock solution and a 5 M NaCl stock solution. You wish to prepare a solution with a final concentration of 0.5 M copper sulfate and 0.25 M NaCl containing 280 mL of water as a solvent. How many milliliters of the copper sulfate stock solution would the solution contain?
11. 40.0 mL of 1.0 M $Fe(NO_3)_3$ is mixed with 20 mL of 5 M $Fe(NO_3)_3$ and 60 mL of water. What is the final molar concentration of NO_3^- ions?
12. You add 2.0 L of a HCl solution of unknown concentration to 2.0 L of 0.5 M HCl and 4.5 L of water. The final concentration of HCl was 1.0 M. What was the unknown concentration of the initial HCl solution?
13. You prepare a 15% NaCl (m/v) solution. (Denisty of NaCl: 2.16 g/mL). How many milliliters of water are in 500 mL of this solution?
14. What is the total molar concentration of ions in a solution prepared by mixing 100.0 mL of 2.0 M KCl with 50.0 mL of 1.50 M $CaCl_2$ and 350 mL of water?

15. The A_{260} nm of a DNA solution is 0.6. How much of this DNA solution and a 6X loading dye should you add to 30.5 µL of water to obtain a sample that contains 75 ng of DNA in 0.5X loading dye? (A_{260} nm of 1.0 = 50 µg/mL DNA)
16. Solution A has 0.20 µg of DNA per µL. Solution B has 0.30 µg of DNA per µL. If you combine 34 µL of solution A with 19 µL of solution B, what would be the absorbance at 260 nm of the final solution?
17. You have a 15% (m/v) solution of glycine (MW: 75g/mole). By what factor must this solution be diluted to obtain a final concentration of 0.05M?
18. A water sample has a phosphate concentration of 6.8 picomolar. Express this as µg/100 L of phosphate. (MW of PO_4^{-3}: 95g/mole)
19. What is the mass per volume (g/L) concentration of total ions in 100 mL of a 0.5M solution of K_2HPO_4? (MW of K_2HPO_4: 174g/mole, K^+: 39g/mole, and PO_4^{-3}: 95g/mole; and H+: 1g/mole)
20. You have three solutions, "A", "B" and "C". Each must be diluted by the following dilution factors: 5X, 15X and 45X respectively. What would be the final volume of a solution prepared by adding each of the ingredients to 100 mL of solvent?

Part II: Lab performance and data analysis **(4 points/question for a total of 40 points)**

Using micropipettes - Part I

1. Generate a standard curve from the data obtained with the volumes ranging from 50-200 µL. (ABS Vs Vol.) Add a trend line of best fit. Determine the R^2 coefficient. (Follow the directives for figures and graphs available on this course's web site)
2. According to your standard curve, what were the average volumes for wells G1-G3 or H1-H3 and wells G4-G6 or H4-H6?

Using micropipettes - Part II

3. What were the accuracy and precision of the measurement obtained using the P-1000, the P-200 and the P-20. Show your calculations.

Preparing solutions

4. Indicate the absorbance readings obtained for each of the following solutions.

    a. A 0.2 mM solution of compound "A".
    b. A 0.72% (m/v) solution of compound "B".
    c. A 5% (v/v) solution of <u>solution</u> I.
    d. A solution containing 0.1 mg of compound "A" and 0.1% (v/v) of compound "B".
    e. A solution with the following ratio: solution I: solution II : water : 2 : 1 : 247

Agarose gel electrophoresis

5. Submit an appropriate figure of your gel electrophoresis of your plasmid isolations and genomic DNA isolation. Include an appropriate legend. (Follow the directives for figures and graphs available on this course's web site)

Spectrophotometric quantification of DNA: Method 1 (UV absorption)

6. Submit a graph representing the A260 readings Vs standard DNA concentrations. Determine the R^2 coefficient. Provide in your legend what was the DNA concentration of the undiluted unknown DNA solution provided?
7. Submit the following information for your yeast genomic DNA isolation:
    - ABS260
    - ABS280
    - ABS260/ABS280 Ratio
    - Concentration in µg/µL of undiluted preparation (show your calculations)
    - Total yield in µg (show your calculations)

Spectrophotometric quantification of DNA: Method 2 (Fluorometry)

8. Submit a graph representing the fluorescence readings Vs standard DNA concentrations. Determine the R^2 coefficient.
9. According to your graph what was the DNA concentration of the undiluted unknown DNA solution provided?
10. Provide the following information for your yeast genomic DNA isolation:
    - Concentration in µg/µL of undiluted preparation (show your calculations)
    - Total yield in µg (show your calculations)

Assignment #2

Part I: Restriction digests and mapping (2 points/ question for a total of 50 points)

1. The nomenclature of restriction enzymes can provide useful information about the source of the enzyme. For example, *Eco*RI indicates that this enzyme was the first enzyme isolated from an *Escherichia coli* strain “R”. From which bacterial genus was *Pst*I isolated from?
2. Define the following terms: Isoschizomer, neoschizomer, and isocaudomer.
3. Amongst the enzymes listed below, which if any, generate compatible ends to each other? (Ex. A and B)

	Enzyme	Recognition Sequence
A	*AciI*	G/CGC
B	*AscI*	GG/CGCGCC
C	EagI	C/GGCCG
D	*HpaII*	C/CGG
E	*MluI*	A/CGCGT
F	*NsiI*	ATGCA/T
G	*PstI*	5CTGC/AG

4. The partial sequences recognized by two restriction enzymes “A” and “B” are indicated below. Complete the sequences such that palindromes are generated for each of the sites. On your palindromes indicate which phosphodiester linkage would have to be cleaved such that 5’ protruding ends of enzyme “A” are compatible with protruding ends of enzyme “B”. **Ex. CT/GCAG**

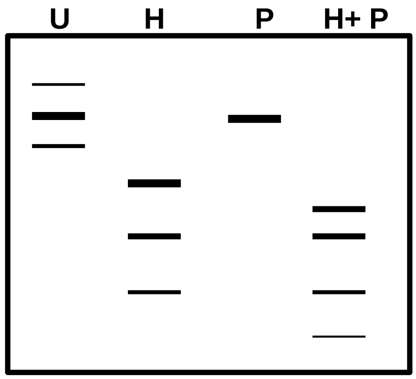
   Enzyme “A”: 5’ATAG— — — — 3’

   Enzyme “B”: 5’GCG — — — 3’

5. Consider your answer to the previous question. A DNA fragment generated with enzyme “A” was ligated to a DNA fragment generated with enzyme “B” as illustrated below. What would be the newly sequence at the junction of the ligated fragments? Which enzymes, A, B, A & B, or neither A or B would cut the ligated DNA?

   ————————A/B————————

6. This picture represents an agarose gel electrophoresis of various restriction digests of a plasmid.



U : Undigested plasmid
H : Plasmid digested with *Hinc*II
P: Plasmid digested with *Pvu*II
H+P: Plasmid digested with *Hinc*II and *Pvu*II

Answer the following questions based on the results obtained:

a. How many times did *PvuII* cut within the plasmid?
b. How many times did *Hinc*II cut within the plasmid?
c. How many times did *Hinc*II cut within the *PvuII* fragment?
d. Which fragment in the *Hinc*II digest was digested by *Pvu*II? (Top, middle or bottom)

7. A linear 12 Kbp DNA **fragment** shown below has cleavage sites for *Bam*HI (B) and *Eco*RI (E). The numbers indicate the position in kilobases. Indicate what fragment sizes would be observed on an agarose gel following digests with *Bam*HI, *Eco*RI, and *Bam*HI + *Eco*RI. Note, if different fragments of the same size are generated, the size should only be indicated once. (For example do not indicate 2Kbp and 2Kbp)

8. What fragment sizes could be generated from a *Bam*HI partial digest? Only indicate the sizes of **intermediate** fragments which would not be obtained following a complete digest. (Fragments which were cut at least once, but which contain one or more *BamHI* site which remain undigested)

9. It was determined that the enzyme *Xho*I cuts at 7.0Kbp on the map shown above. Which *Bam*HI fragment size would be cut by *Xho*I?

10. A complete digest with *Eco*RI + *Bam*HI of 5µg of the above fragment was performed. Indicate the amounts in µg of each fragment size generated.

The table below presents the results of different digests of a plasmid. All sizes are in base pairs.

Digest	Size of Fragments
HindIII	4.00 kb
BamHI	2.35 kb, 1.65 kb
EcoRI	4.00 kb, 2.50 kb,1.50 kb
HindIII + BamHI	1.65 kb, 1.50 kb, 0.85 kb
HindIII + EcoRI	3.50 kb, 3.00 kb, 2.50kb, 1.00 kb, 0.50 kb
BamHI + EcoRI	2.00kb, 1.85kb, 1.65kb, 1.5kb, 0.35kb, 0.50 kb

11. One of the enzymes used only partially digested the DNA. Which enzyme is it?
12. Indicate the size in base pairs of one of the intermediate products, which represents an incompletely digested DNA fragment in the double digests.
13. What is the total plasmid size in base pairs?
14. What are the distances between the *EcoRI* and *HindIII* restriction sites?
15. What are the distances between the *Bam*HI and *HindIII* restriction sites?
16. What are the distances between the *Bam*HI and *EcoRI* restriction sites?

The table below presents the results of different complete digests of a linear DNA fragment. All sizes are in base pairs.

HindIII	5000, 2500
SmaI	5500, 2000
EcoRI	4000, 3500
HindIII + *EcoR*I	3500, 2500, 1500
HindIII + *SmaI*	3000, 2500, 2000
SmaI + *EcoRI*	4000, 2000, 1500

17. What is the distance between the *EcoRI* and *HindIII* restriction sites?
18. What is the distance between the *HindIII and SmaI* restriction sites?
19. Among the fragments generated in the *HindIII and SmaI* digest, which one could be ligated and cloned in a vector digested with *HindIII and SmaI*? Indicate the size of the fragment.
20. The restriction enzyme *Bba*I cleaves the sequence R/AYRTY (R= A or G and Y = C or T). How many different palindromes does *BbaI* recognize?
21. What would be the average fragment size, which would be expected following the digestion of genomic DNA with *Bba*I? Assume an equal distribution of A, G, C, and T.
22. At what frequency would the restriction enzyme *Bba*I be expected to cut in a genome with the following base distribution on one strand: 20% G, 30% C, 20% A, and 30% T?
23. True or false; an overhang generated following the cleavage of a palindrome recognized by *BbaI* would necessarily be compatible with any other palindrome cleaved by *Bba*I?

24. The enzyme *Psp*N4I cleaves the sequence NNGCNN (N= any of the four bases). How many times would you expect *Psp*N4I to digest a 10.752 kb genome?

25. Sequence A, which contains two *Bst*BI sites (TT/CGAA), was digested with *Bst*BI. The resulting fragment was then ligated into the unique *Taq*I restriction site (T/CGA) within vector B.

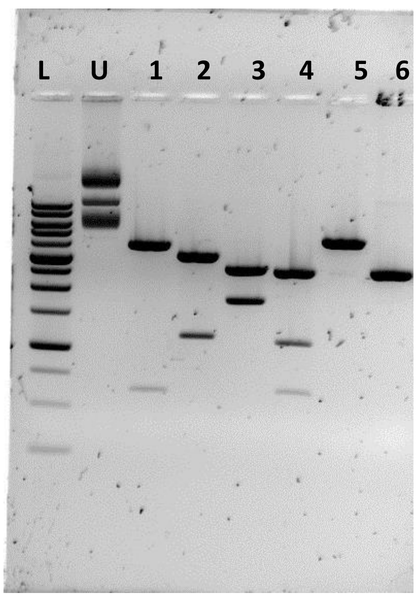
    Sequence A: CAG **TT/CGAA** TTC • • • • • • • • • • • • • • • • GGC **TT/CGAA** AAG

    Vector B: TGG **T/CGA** CAC

    Which enzyme (s) could be used to release the cloned DNA fragment from the recombinant vector B? *Bst*BI, *Taq*I , both, or neither.

Part II: Restriction mapping (4 points/ question for a total of 40 points)

Below is shown an agarose gel electrophoresis of the predigested DNA as you performed in the lab. (This file can be obtained from the course's web site under the heading Data>predigests)

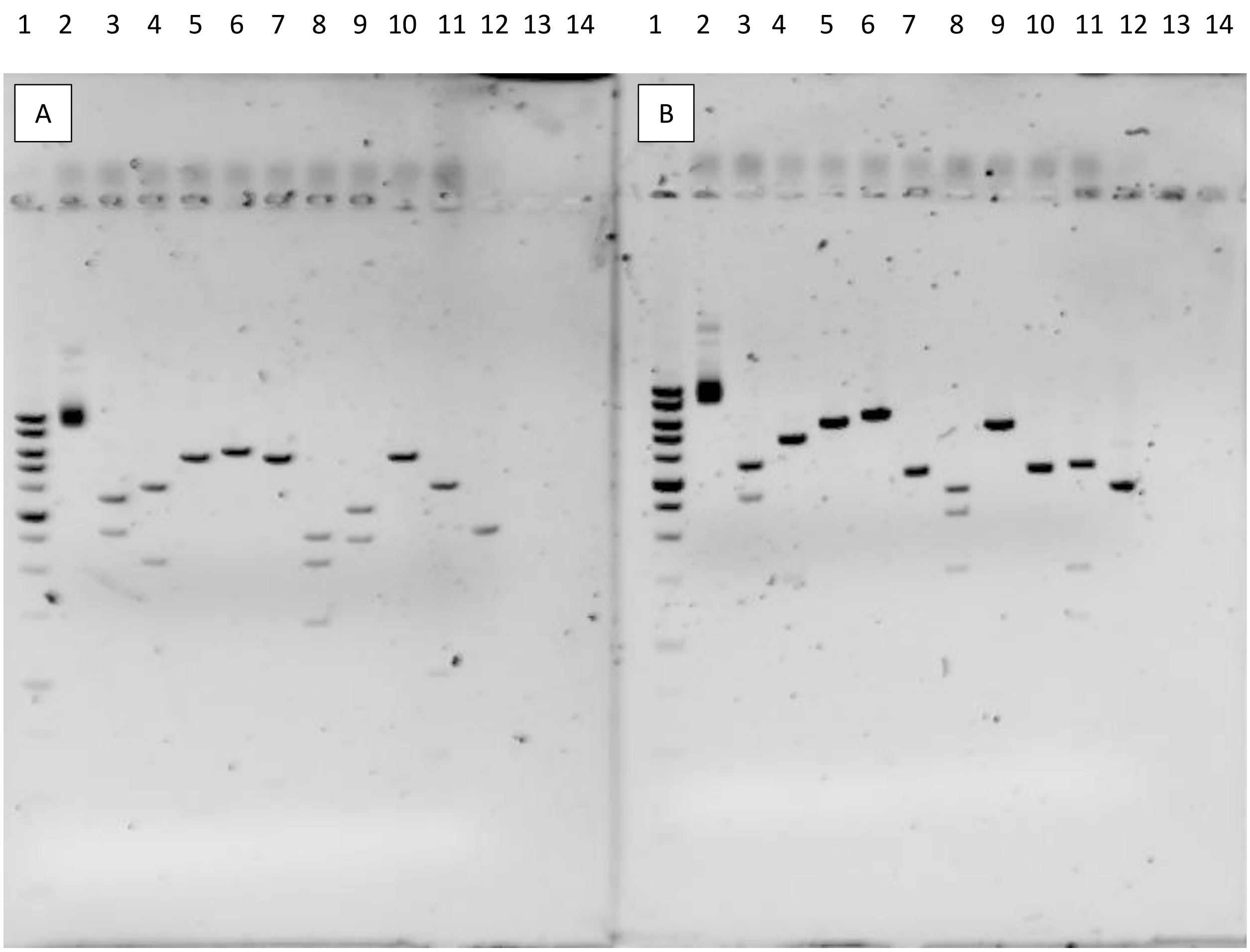


L: 1Kbp molecular weight markers
U: Undigested recombinant **pUC9** plasmid DNA
1: Recombinant **pUC9** plasmid cut with *Bam*HI
2: Recombinant **pUC9** plasmid cut with *Eco*RI
3: Recombinant **pUC9** plasmid cut with *Hind*III
4: Recombinant **pUC9** plasmid cut with *Eco*RI + *Hind*III
5: Recombinant **pUC9** plasmid cut with *Pst*I
6: **pUC9** vector cut with *Bam*HI

1. Submit a standard curve of the molecular weight ladder (Migration distance Vs. Size in bp). Include the R^2 coefficient.
2. Submit a table presenting the results of the restriction digests of the **recombinant** plasmid that includes the following information: Enzyme used, Total number of cuts, Number of cuts in the insert, Number of cuts in the vector, fragment sizes observed. In a caption accompanying the table, indicate the total size of the plasmid, the size of the vector, the size of the insert, and the restriction site (s) in which the insert was introduced in the vector.

3. Submit a figure that represents a possible restriction map of the insert within the multiple cloning site of pUC9. Your map should be linear, to scale and only include the insert within the multiple cloning site. (See directives on this course's web site under the heading graphs and figures)

Below is shown an agarose gel electrophoresis of the digests of the unknown recombinant pUC19 plasmids (A: orientation 1, B: orientation 2) as you performed in the lab. (This file can be obtained from the course's web site under the heading Data>pUC19 recombinant)



1. Mol. wt. marker
2. undigested
3. BamH1
4. EcoR1
5. EcoRV
6. HinDIII
7. PstI
8. HinDIII + EcoR1
9. HinDIII + EcoRV
10. HinDIII + PstI
11. EcoR1 + EcoRV
12. pUC19 Bam
13. empty
14. empty

4. Submit a table presenting the analysis of the restriction digests of the unknown you were provided with. Your table should include: Enzyme (s) used, Total number of cuts, Number of cuts in the vector, Number of cuts in the insert, and Fragments sizes generated.
5. Submit a figure of the restriction map of the insert from the recombinant plasmid you were provided with. Your map must be linear, include the multiple cloning site, the insertion site, the size of the insert, the positions in the multiple cloning site or the insert of all the enzymes tested. Your figure must be to scale. (Make sure that the enzyme indicated as the insertion site is consistent with both orientations) Follow the directives for generating such a figure under the heading Graphs/Figures on this course's web site.
6. Submit a figure of your own agarose gel electrophoresis of the predigested pUC recombinant and the calibration of a restriction enzyme. Make sure to include an appropriate legend. Follow the directives for figures on the Web page of this course and to include all the required information in the legend for the understanding and interpretation of the figure.
7. According to the experiment presented in question 6, what was the most dilute sample which showed a complete digestion (Indicate the dilution)? Based on this information, what was the approximate undiluted enzyme concentration in units/µL? Show how you arrived to this conclusion.
8. Submit a figure of your own agarose gel electrophoresis of the restriction digests of the recombinant pUC plasmid you were provided with. Make sure to include an appropriate legend. Follow the directives for figures on the Web page of this course and to include all the required information in the legend for the understanding and interpretation of the figure.

A Southern hybridization was performed on genomic digests of *Saccharomyces cerevisiae*. Following migration on an agarose gel, the digests were probed with a 3Kbp radioactive probe. The results obtained following the hybridization are presented below.

Enzyme	**Size of hybrids (Kbp)**
BamHI	13
EcoRI	10 + 15
EcoRI + BamHI	6 + 7

9. How many times do the enzymes *Bam*HI and *Eco*RI cut the *S. cerevisiae* genome within the region spanned by the probe?
10. Draw a possible restriction map of this genomic region of *S. cerevisiae*. Indicate the distances between each of the restriction sites.

Bioinformatics 1-2 **(1.25 points/ question for a total of 10 points)**

1. What is the protein accession number and source organism of the first record for the restriction enzyme *BglII* obtained through a general search?
2. Does the record with the accession number M68489 correspond to a nucleotide or a protein sequence?
3. What is the source organism and the name of the gene corresponding to the record M68489?
4. Submit the following information about each of the unknown genes from the first bioinformatics exercise.
   - **Accession number**
   - **Coverage**
   - **Ident.**
   - **E value**
   - **The definition**
   - **The organism from which this sequence was obtained**
   - **The gene name**
   - **The gene's product name**
   - **The protein's accession number**
5. Provide theoretical restriction maps of the three unknown genes available on this course's Web page. Use the enzymes listed on page 35 of the lab manual. Indicate below each map the name of the gene and list the enzymes that do not cut.
6. Compare the theoretical maps generated in the previous question to the experimental map of the unknown insert you analyzed in the lab ("**Restriction mapping of a recombinant plasmid**"). The unknown insert corresponds most closely to which gene?
7. Provide a restriction map showing the positions of the restriction sites *Pst*I, *Sal*I, *Nco*I, and *Xba*I within the region spanned by positions 5683-7676 of the sequence "**Unknown sequence**" available on this course's Web page. Indicate below the map the definition of the gene.
8. According to the map you submitted for question 7, what fragment sizes would be generated following a *SalI-XbaI* double digest of the linear fragment from positions 5683-7676?

Assignment #3

Part I: Cloning and transformations **(3 points/ question for a total of 60 points)**

Sickle-cell anemia is caused by a mutation in the human β-globin gene. The three possible genotypes are homozygous for normal β-globin, heterozygous carrier (having both the normal and sickle - cell alleles), and homozygous for sickle cell anemia. Recombinant DNA technology has been used as the basis for prenatal diagnosis of sickle cell anemia. In a very high percentage of the cases observed, the normal human β-globin gene is carried on a 13000 bp DNA fragment from an *Hpa*I digest, while the sickle-cell gene is carried on a 7600 bp *Hpa*I fragment.

Available to you are:

- A labeled sample of recombinant DNA consisting of a bacterial plasmid vector (4000 bp) carrying the 13000 bp *Hpa*I fragment from the normal human genome.
- A sample of genomic DNA from each member of a couple thought to be carriers of the sickle-cell trait and expecting their first child.
- A sample of the genomic DNA obtained from the fetal cells in the amniotic fluid from the uterus of the pregnant woman.

1. You perform a Southern on *Hpa*I digested DNA using the recombinant plasmid as a probe. What sized band (s) would you expect in each of the following cases?

   - Heterozygous carrier mother
   - Homozygous normal mother
   - Homozygous for sickle cell anemia fetus
   - Combined genomic DNA from a heterozygous carrier mother + homozygous normal father

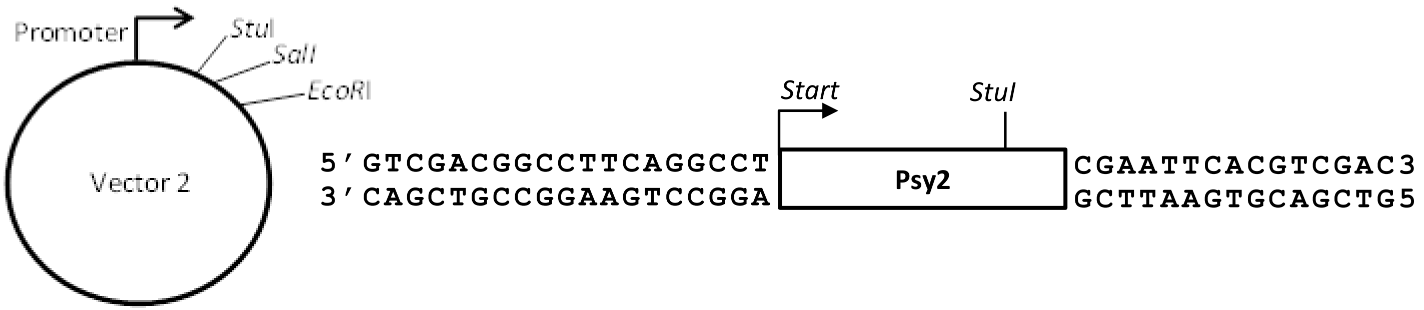
2. You have a partial DNA sequence from the psy2 gene of yeast as shown below. You decide to use PCR to amplify the psy2 sequence based on the flanking sequences shown below.

Indicate the set (s) of primer (s) you could use to amplify the entire psy2 coding sequence.

Set 1:	Set 2:	Set 3:
5’ TCCGGA 3’	5’ AATTCC 3’	5’ GACTCC 3’
5’ GCCGGA 3’	5’ AGTCGA 3’	5’ TTGGAA 3’

```
5’ TCCGGCGGAATTCCAAGGCCT [ psy2 ] CGTCGACTCCGGC 3’
3’ AGGCCGCCTTAAGGTTCCGGA [      ] GCAGCTGAGGCCG 5’
```

Consider the following information to answer questions 3-5:
You successfully amplify the psy2 sequence and wish to clone the PCR fragment into vector 2 in order to express it in yeast. The cloning sites available on this vector are shown below.



*Stu*I: 5'- AGG/CCT-3' *Sal*I: 5'-G/TCGAC-3' *Eco*RI: 5'-G/AATTC-3' *Bam*HI: G/GATCC

There are two different ways to insert the amplified psy2 sequence into vector 2. Give the restriction enzyme(s) that you could use to cut the vector and the psy2 coding sequence for each of these.

3. **Directional cloning:**
   Cut vector and psy2 with:

4. **Non directional cloning:**
   Cut vector and psy2 with:

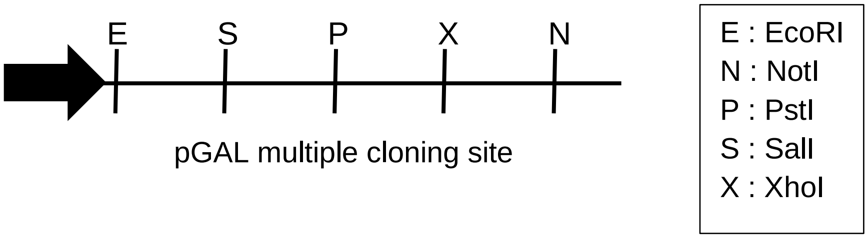
5. Which enzyme (s) could you use to verify the presence and orientation of the insert?

Consider the following information to answer questions 6-9:
Several lab groups independently carried out ligation/transformations similar to those you performed in the lab to introduce GFP into the MCS of digested pUC vector. Both the vector and GFP sequence were digested with the same single enzyme. The results obtained by each of the groups were as follows:

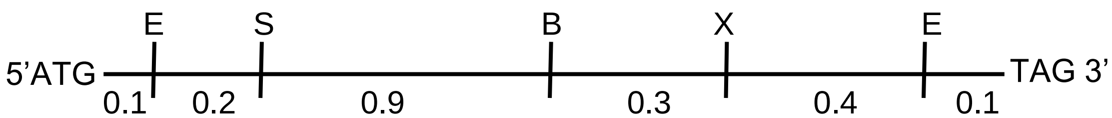
	# of colonies on LB-Amp plates			
Transformation into XL-1	Group 1	Group 2	Group 3	Group 4
a) Phosphatase treated cut pUC + GFP insert	0	2	529	900
b) Phosphatase treated cut pUC	0	2	2	850
c) Non-phosphatase treated cut pUC	0	800	750	890
d) 10μl uncut pUC (0.5 ng/μL)	0	500	500	500
e) No DNA	0	0	0	500

6. Which group obtained the expected/desired pattern of results?
7. Which ligation mixture would be expected to give the greatest percentage of intermolecular ligation events?
8. Comparing ligation reactions “a” and “b”, which group had the highest percentage of intramolecular ligation events?
9. What was the transformation efficiency of the competent XL-1 cells? Indicate the number of transformants per microgram of undigested DNA.

Consider the following information to answer questions 10-14:
You are interested in the directional cloning of the maize heat shock protein gene (2 kb) into the vector pGAL. You need to choose the restriction sites that you will include in your PCR primers and that you will use to cut the vector pGAL (5 kb). The following are maps of the multiple cloning site of pGAL and the DNA sequence that encodes the maize heat shock protein.



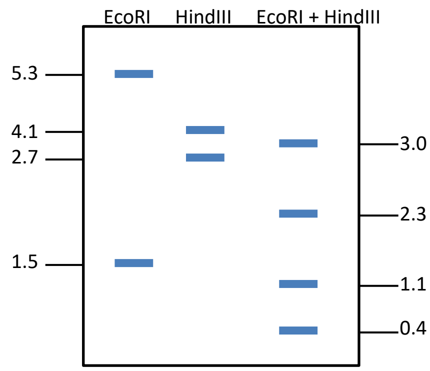
pGAL multiple cloning site



Restriction map of maize heat shock protein gene

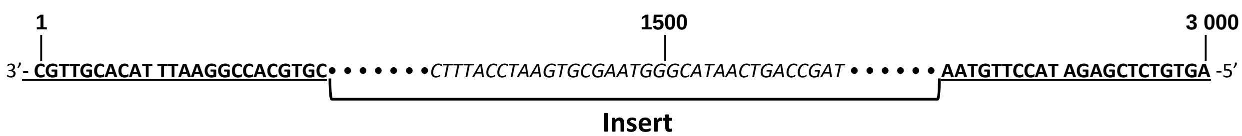
10. Which restriction site would you add to your forward primer?
11. Which restriction site would you add to your reverse primer?
12. You isolate plasmid DNA from a colony that you think has the insert of interest. To verify your assumption, you digest the recombinant plasmid with *Xho*I. What fragment size (s) is (are) expected?
13. To verify the orientation of the insert within the pGAL vector, you perform a *SalI* digest on a positive recombinant. What fragment sizes would you expect?

Consider the following information to answer questions 14-17:
You cloned an *Eco*RI digested DNA fragment into a unique *Eco*RI site of a vector. You identify a recombinant vector that you believe has the DNA of interest. To generate a restriction map of the recombinant plasmid, you take three individual samples of the plasmid and digest them with *Eco*RI, *Hind*III, and with both *Eco*RI and *Hind*III. You then run the digested DNA on an agarose gel to see the fragments. The gel is then subjected to a Southern hybridization using the *Eco*RI insert as a probe. Assume that the insert fragment is smaller than the vector.



14. Which fragment (s) on the gel would hybridize to the probe that represents the inert?
15. Which fragment (s) on the gel would hybridize to a probe representing the 5.3 kb band observed in the *Eco*RI lane?
16. Which fragment (s) in the *Eco*RI lane would hybridize to a probe representing the 2.7 kb band observed in the *Hind*III lane?
17. What new fragment sizes would be observed following a hybridization with the 1.5 kb band of a recombinant with the insert in the opposite orientation?

You cloned an insert of approximately 3 kb into a single restriction site within the multiple cloning site of a vector. The partial flanking sequences of the multiple cloning site (underlined) and of the insert sequence in the desired orientation (in *italics*) are indicated below. You wish to use PCR on various clones to identify those which possess the insert in the correct orientation (as indicated below).



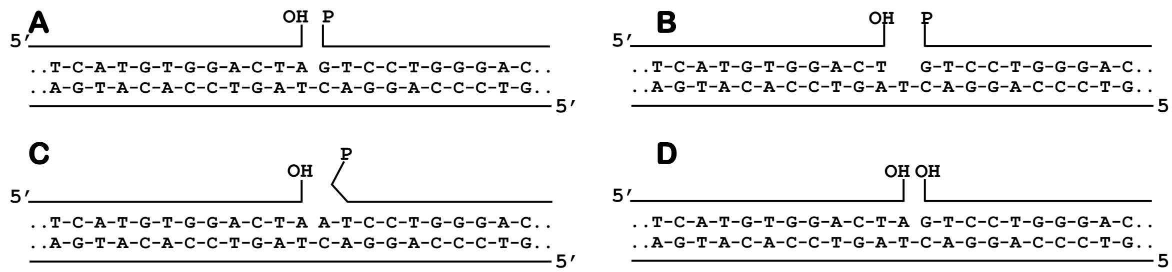
18. Which two primers could you use to identify clones that possess recombinant plasmids with the insert in the correct orientation?

Primers: (All sequences are written 5' to 3')

A. GCAACGTGTA	C. GTCTCGAGAT	E. GCAACGTGTA
B. CCCATTCGCA	D. TTCCATAGAG	F. TAAGCGTGAA

19. You wish to amplify a 1 kb sequence from a linear single stranded 10 kb viral genome. To do so, you use a forward primer that anneals at position 1-25 and a reverse primer that anneals at position 975-1000. How many cycles total would it take to obtain 1µg of amplicon from 2ng of single stranded viral DNA?

20. Below are diagrams of double stranded DNA molecules some of which have mis-matches (“P” represents a 5’ phosphate group) and single strand breaks. Indicate which single strand breaks could efficiently be closed by DNA ligase.



Part II: Cloning (5 points/ question for a total of 25 points)

1. Submit a figure representing your PCR of the GFP gene. Include an appropriate legend, which includes the size of the amplicon observed. (Exercise 4)

2. Submit a table providing the following information for each of the ligation mixtures. Indicate the number of transformants obtained as number of colonies /mL for each ligation reaction as well as for undigested vector.
   - Undigested vector
   - Non-phosphatase treated vector
   - Phosphatase treated vector
   - Phosphatase treated vector + GFP insert

3. What was the transformation efficiency of the competent XL-1 cells provided? Show your calculation. (Note, 20ng of the undigested pUC vector were used to transform XL-1)

4. Submit a figure of your PCR analysis of your plasmid recombinants. Indicate in your figure legend the size of the amplicon observed. Briefly explain if the sizes obtained are those expected and whether the results indicate that the amplicon was cloned in the correct orientation.

5. Submit a figure of your agarose gel electrophoresis of your restriction digests of your plasmid recombinant. Indicate in your figure legend the size of the fragments observed. Briefly explain if the sizes obtained are those expected and whether the results indicate that the amplicon was cloned in the correct orientation.

Bioinformatics 3 **(2 points/ question for a total of 15 points)**

1.
Step 1: Obtain the reverse sequence of the sequence with the accession number **NR_002716**.
Step 2: Obtain the complement sequence of the sequence from in step 1.
Step 3: Obtain the reverse complement sequence of the sequence from step 2.
Step 4: Obtain the reverse sequence of the sequence from step 3.
Step 5: Obtain the complement sequence of the sequence from step 4.
Indicate the first 20 bases of the final sequence obtained. Indicate the 5' and 3' ends.

2. What would be the complement of the reverse sequence of the following sequence? Indicate the 5' and 3' ends.

3'-GATCGGATCCCATCTTATC-5'

3. Map the alignment positions of each of the following primer sequences on the sequence of the pUC19 sequence.

   A. CCTTGAAGATCAGTTGGGTGC
   B. TTTCTTAGACGTCAGGTGGTG
   C. GGGCGCGTTTCGGTGATGACG
   D. AATACGGGATAATACCGCGCC

4. Which of the primers indicated in the previous question, if any, would give an amplification product of at least 200 bp?

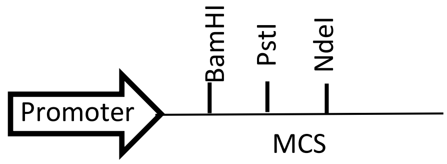
5. **Amplification of GFP:** In exercise 3 of the lab, you performed PCR to amplify the GFP gene using pGFPuv as the template sequence. Use the knowledge gained in bioinformatics to obtain the following information:

   - Alignment positions of the forward and reverse primers.
   - Predicted size of the amplicon.
   - Restriction map for *BamH*I, *EcoR*I, and *Pst*I of the amplicon

6. **Cloning of GFP:** In exercise 5 of the lab, you performed PCR to amplify the GFP gene from potential positive clones. Use the knowledge gained in bioinformatics to obtain the following information:

   - Alignment positions of the forward and reverse primers used for screening the recombinants. Note the reverse primers must be aligned to the GFP amplicon sequence whereas the forward primer must be aligned to the pUC sequence.
     - GFPSc-F1: AGCTCACTCATTAGGCACCCCAGGC
     - GFPSc-R1: GTAGCGACCGGCGCTCAGTTGG
     - GFPSc-R2: CCAACTGAGCGCCGGTCGCTAC
   - Predicted size of the amplicon.
   - Which primer pair will give an amplification product if the GFP insert is in the correct orientation?

7. You wish to clone and express the complete human prolactin coding (mRNA) sequence (from ATG to TAA) into the multiple cloning site of an expression vector: **(3.0 points)**



Using the skills you have acquired in bioinformatics, design a pair of primers that would allow you to amplify the prolactin coding sequence (starting at the ATG and ending at the TAA stop codon). Your primers should also include the sequence of appropriate restriction sites for the directional cloning and expression of the amplicon. Your primers should be 15 bases long including the restriction sites. Provide the following information:

Sequence of forward primer and restriction site added.

Sequence of reverse primer and restriction site added.

Expected size of amplicon.

Assignment #4

Part I (2 points/ question for a total of 30 points)

1. You have been asked to PCR amplify a specific sequence from cDNA that were synthesized from mRNA isolated from brain tissue. After you run your PCR reaction on an agarose gel containing ethidium bromide, you observe no bands when you visualize the gel using ultraviolet light. Why might this be the case? Choose all possible answers.

   A. The gene you are interested in is not expressed in brain tissue.
   B. An oligo dT was used instead of a oligo dA to prime the first strand cDNA synthesis.
   C. A specific primer whose sequence was that of the non-coding strand of the gene was used to prime the first strand cDNA synthesis.
   D. You used reverse transcriptase instead of Taq polymerase to synthesize the first strand of cDNA.

2. You perform a first strand cDNA synthesis from three molecules of an mRNA template (2000 bases) using an oligo-dT primer. A pair of forward and reverse primers (spanning positions 560-900), are then used for the subsequent PCR reaction. How many **PCR** cycles (do not include RT reaction) are required to obtain 48 molecules of products with ends defined by both primers? By this, I mean that you count only double stranded molecules that begin and end at the primer binding sites but lack any other sequence.

3. The DNA sequence of one strand of the extremely tiny gene Lilliputian is shown below, with the location of its start and stop codons underlined. Is this the sense or antisense strand?

   5'CAG GAC ATGC <u>TTA</u> TCG TAC TAT GGG TGC AAT GCC CAT TAA GGG TGC <u>CAT</u> ACC GAT GAT GCC TCA3'

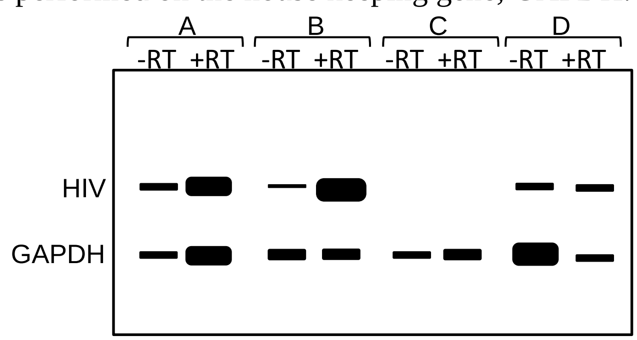
4. I want to use RT-PCR to make a copy of the <u>entire</u> Lilliputian mRNA. Indicate, the sequence of a primer of 6 nucleotides that could be used for the first strand cDNA synthesis.

Consider the following information to answer questions 5-8.

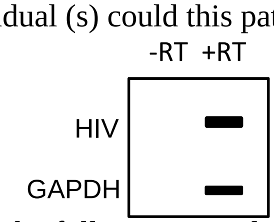
The Lilliputian gene is known to be induced 5 fold when cells are grown in glucose as compared to growth in glycerol. To study the expression of this gene, RT-PCR was performed for a wild type strain as well as various mutants. RT-PCR reactions were done simultaneously for the Lilliputian gene, as well as a house keeping gene; GAPDH. A densitometric analysis was then performed. The densitometric values obtained for a wild type strain of the organism grown in glucose was 500 for the Lilliputian mRNA and 250 for the GAPDH mRNA. Indicate the expected densitometric values for the Lilliputian mRNA in each of the following scenarios:

5. Wild type grown in glycerol (GAPDH value = 100)
6. Mutant strain with a mutation which enhances the promoter's activity 2 fold grown in glycerol (GAPDH value = 500)
7. Mutant strain with a mutation which changes the ATG codon to CTG grown in glycerol (GAPDH value = 250)
8. Mutant strain with a mutation which represses the promoter's activity 5 fold grown in glucose (GAPDH value = 500)

Consider the following information to answer questions 9-13.
RT-PCR is a commonly used diagnostic method for the detection of HIV infections, caused by an RNA virus. Briefly, following infection, the viral RNA genome is converted to a double stranded DNA that integrates into the genome. Following integration, the virus may remain dormant for several years. Following activation, the viral DNA is transcribed into RNA and initiates its replicative cycle; spreading from cell to cell. The following figure illustrates RT-PCR reactions done on the blood from different individuals. Reactions were carried out in the presence and absence of reverse transcriptase in the reaction mixture (+RT and –RT respectively). Simultaneous RT-reactions were performed on the house keeping gene; GAPDH.



9. Which individual (s) have an HIV specific amplification product derived from DNA but not RNA?
10. Which individual (s) have an HIV specific amplification product derived from RNA?
11. Which individual shows the highest expression of HIV RNA?
12. Which individual has more copies of the HIV genome integrated in their genome?
13. The above experiment was repeated as previously, but the samples were initially treated with DNAse before performing the reverse transcriptase reactions. One of the patterns observed is shown below. To which individual (s) could this pattern correspond?



14. Which polymerase could read the following template and synthesize a complementary strand using the primer 5' ACCCTTAATGGG 3'? Choose all that apply.

5'UUCAGGACAUGCUUAUCGUACUAUGGGUGCAAUGCCCAUUAAGGGUGCCA-3'

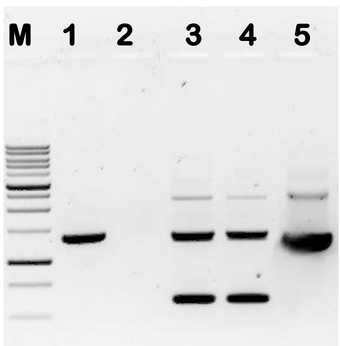
A. A DNA dependent DNA polymerase
B. A DNA dependent RNA polymerase
C. An RNA dependent RNA polymerase
D. An RNA dependent DNA polymerase

15. Indicate the sequence in the 5' to 3' direction of the first 6 DNA bases which would be added to the primer.

Part II: Gene expression - Transcription **(5 points/ question for a total of 40 points)**

1. Indicate the ABS_{260}, ABS_{280}, the ABS_{260}/ABS_{280} ratio and total yield of your RNA preparation.
2. Submit a figure of your RNA gel generated in lab exercise 6. Include an appropriate figure legend.
3. Submit a figure of the RT-PCR gel 1 generated in lab exercise 7. Include an appropriate figure legend, which indicates the sizes of the products observed in each lane.

Below is a gel of RT-PCR reactions similar to the ones you performed in exercise 5. (You may obtain a copy of these results on this course's Web page under the heading data>rtpcr)



Reaction	Template
PCR-1	**R** (Step I)
PCR-2	**DR** (step II)
PCR-3	**RT-R** (step III)
PCR-4	**RT-DR** (step III)
PCR-5	Yeast genomic DNA

4. What is the purpose of the DNAse treatment?
5. What is the purpose of the PCR reaction using RNA <u>NOT</u> treated with either DNAse or RT?
6. What were the sizes of the RT dependent and RT independent PCR products?
7. What does the difference in size between the PCR products tell you about the gene from which this RNA is derived from?
8. Submit a figure of the RT-PCR gel 2 generated in lab exercise 7. Include an appropriate figure legend, which indicates the sizes of the products observed in each lane.

Bioinformatics 4 & 5 **(1.2 point/ question for a total of 30 points)**

1. What is accession number of the first nucleotide record for phospholipase A2 mRNA from *Mus musculus*?
2. What is accession number of the first nucleotide record for phospholipase A2 mRNA from *Alligator mississippiensis*?
3. What is accession number of the first nucleotide record for phospholipase A2 mRNA from *Canis lupus familiaris*?
4. What is the percentage of identity at the nucleotide level between the phospholipase A2 genes from each of the following pairs of organisms:
    A. Phospholipase A2 from *Homo sapiens* to *Mus musculus*
    B. Phospholipase A2 from *Homo sapiens* to *Alligator mississippiensis*
    C. Phospholipase A2 from *Homo sapiens* to *Canis lupus familiaris*
    D. Phospholipase A2 from *Mus musculus* to *Canis lupus familiaris*
5. Which pair (s) of sequences, if any, are analogues?
6. Which pair (s) of sequences, if any, are orthologues?
7. What is the percentage of identity between the phospholipase A2 proteins from each of the following pairs of organisms:
    A. Phospholipase A2 from *Homo sapiens* to *Mus musculus*
    B. Phospholipase A2 from *Homo sapiens* to *Alligator mississippiensis*
    C. Phospholipase A2 from *Homo sapiens* to *Canis lupus familiaris*
    D. Phospholipase A2 from *Mus musculus* to *Canis lupus familiaris*
8. What is the accession number for the serine/threonine-protein kinase protein record of *Bos taurus*?
9. What is the accession number for the serine/threonine-protein kinase of *Mus musculus*?
10. What is the accession number for the serine/threonine-protein kinase of *Monoraphidium neglectum* obtained through the general search?
11. Which pair of amino acid sequences, if any, are analogues?
12. Which pair of amino acid sequences, if any, are orthologues?
13. What are the protein and nucleotide accession numbers for the homologues of NP_002732 from *Danio rerio*?
14. What type of homologues are the nucleotide sequences from *Danio rerio* and NM_002741?
15. What type of homologues are the protein sequences from *Danio rerio* and NP_002732?
16. What protein from *Homo sapiens* shares the highest level of identity with the unknown protein from *Danio rerio*? Indicate the percent identity and coverage.

17. What is the name of the first domain shared between unknown protein from Danio rerio and Homo sapiens?
18. The protein from which of the different organisms (Rattus norvegicus (Norway rat), Arabidopsis thaliana (thale cress), and Escherichia coli) has the highest level of similarity with the protein from *Danio rerio*?
19. From which organism does the viral 1 sequence come from?
20. What is the protein product and function of the longest ORF from the viral 1 sequence?
21. In which reading frame was the longest ORF of the viral 1 sequence found?
22. What is the percentage identity between the translated ORF from viral 1 and the most closely related protein from a different organism?
23. When comparing the longest translated ORF from the viral 1 sequence to the most closely related protein from a different organism, what percentage of the amino acids represent conserved substitutions?
24. What snp has the viral 2 sequence acquired? Indicate the position and base change. Ex. C118 to A.
25. What type of amino acid change, if any, does this snp lead to?

Assignment #5

Part I

Refer to the following description to answer questions 1-6 (2 points/ question for a total of 50 points)

Below are 210 consecutive base pairs of DNA that includes only the beginning of the sequence of gene X. The underlined sequence (from position 20-54) represents the promoter for gene X and the underlined and italicized sequence (from position 71-90) encodes the gene X ribosome binding (RBS) site. Transcription begins at and includes the T/A base pair at position 60 (underlined).

```
   1         10        20        30        40        50        60        70
   I--------I---------I---------I---------I---------I---------I---------I
5′ ATCGGTCTCGGCTACTACATAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTAGGCTACTAC
3′ TAGCCAGAGCCGATGATGTATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGATCCGATGATG
                                                    Promoter
            80        90        100       110       120       130       140
            I---------I---------I---------I---------I---------I---------I
5′ CAGGTATCGGTCTGATCTAGCTAGATGCTCTTCTCTCGAACCCGCGGGGGCTGTACTAGCATGCGTCG
3′ GTCCATAGCCAGACTAGATCGATCTACGAGAAGAGAGCTTGGGCGCCCCCGACATGATAGTACGCAGC
          RBS
           150       160       170       180       190       200       210
   ---------I---------I---------I---------I---------I---------I---------I
5′ TCTCGGCTACTACGTAAACGCGCGCATATATCGATATCTAGCTAGCTATCGGTCTCGGCTACTACGTAAA
3′ AGAGCCGATGATGCATTTGCGCGCGTATATAGCTATAGATCGATCGATAGCCAGAGCCGATGATGCATTT
```

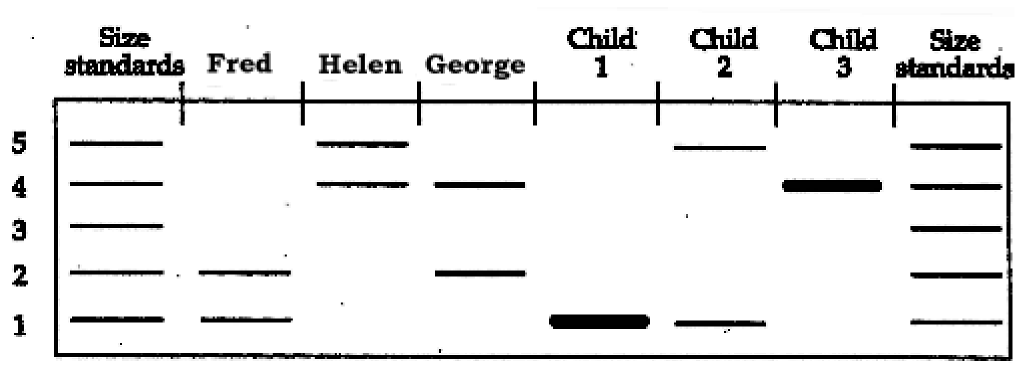
1. What are the first 6 nucleotides of the mRNA from gene X?
2. What are the first 4 amino acids encoded by gene X? Use the one letter amino acid code to indicate your answer. Ex. TRCG
3. You have found a mutant of gene X. The mutation represents an SNP which changes the G/C base pair at position 110 (underlined) to T/A. Would the mRNA expressed from this version of gene X be longer, shorter, or the same as that produced from the normal gene X?
4. If the mutant mRNA can be translated, would you expect the protein to be longer, shorter, or the same as that produced from the normal gene X?
5. You have found another mutant of gene X. The mutation represents an indel which removes the G/C base pair at position 110 (underlined). Do you expect the protein to be longer, shorter, or the same as that produced from the normal gene X?
6. What is the length of the 5’ UTR?

Refer to the following description to answer questions 7-11

The following schematic represents a yeast gene and its various elements:

Indicate the predicted sizes (bases for RNA and amino acids for protein) of each of the following:

7. Pre-mRNA (unprocessed)
8. mRNA
9. Protein
10. mRNA from a gene with an snp at position 1652 creating a stop codon (1652-TAG-11654)
11. Atomic mass (kDa) of the protein from a gene with an snp at position 1652 creating a stop codon (1652-TAG-1654)
12. Fred is married to Helen, who was previously married to George, now deceased. George and Helen conceived one child together and adopted one child. Fred and Helen have also conceived one child. All members of Helen's current family have had DNA fingerprinting done at a single VNTR locus. Unfortunately, the sheet that identified each child has been misplaced.



Identify which fingerprint in each lane (lanes 5, 6, and 7) correspond to each child.

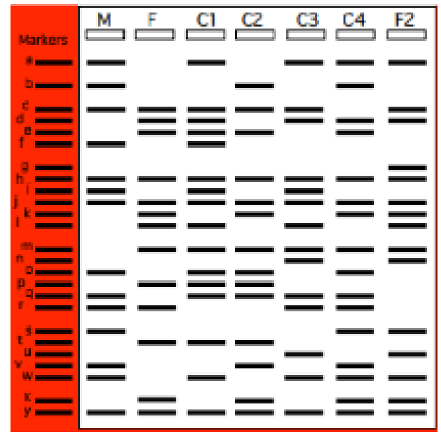
Fred and Helen's conceived child Child ____

George and Helen's conceived child Child ____

George and Helen's adopted child   Child ____

Refer to the following description to answer questions 14-17

The following diagram shows a Southern blot of restriction digested genomic DNA from an elf Mother (M) and elf Father (F) and four potential elf children (C1 to C4) probed with a VNTR DNA sequence. The restriction enzyme used was *Not* I. Another elf (F2) is claiming to be the father of child C4. Assume the mother, M, is the real mother of these four children.



13. Which child is the biological child of the elf Father (F)?
14. Which children if any, could be elf F2's children?
15. According to all the fingerprints generated what are the minimum and maximum number of loci being examined?
16. According to the number of loci indicated in the previous question, the profile of the elf Father (F) must include a minimum of how many homozygous loci?

Refer to the following description to answer questions 17-21

The following schematic shows the region of a gene associated with hairy toes (HTO). The hairy toes phenotype is due to a recessive mutation which creates an RFLP removing a *Hae*III restriction site at position 550 (*Hae*III sites are indicated by arrows). Note that there may be other snps, but these are unrelated to the phenotype of hairy toes. This region of the genome was examined in different individuals by performing a Southern analysis on genomic DNA samples digested with *Hae*III and probed with a DNA fragment spanning the region from 425-650. The results of the hybridization are shown in the table below.

Individual	Sizes of bands observed (bp)
A	450 and 625
B	475, 300 and 150
C	300 and 150
D	475, 450 and 150

17. Which individual (s) have a genotype which would give them the hairy toes phenotype?
18. Which individual (s) have a genotype which would give them the normal phenotype?
19. Which individual (s) is (are) a carrier of the hairy toes genotype, but have a normal phenotype?
20. How many different alleles of the region were detected among the four individuals?

21. Below are the sequences of portions of the yfg1 mRNA and three yfg1 mRNA variants (A, B, and C) from zebrafish embryos.

    5’...GAUGAAAGAUCAGGUCUGAAUGUAU...3’ yfg1 mRNA

    5’...GAUGGUUGAUCAGGUCUGAAUCUAU...3’ Variant A

    5’...GAUGAAAGAUCACAUCUCAAUGUAU...3’ Variant B

    5’...GAUGAAAGAUCAACUCUGAAUGUAU...3’ Variant C

    You’ve created a probe which is 100% complementary to the sequence of the yfg1 mRNA: 5’...TTCAGACCTGATCTTTCATC...3’. You hybridize the probe to total zebra fish RNA at room temperature (20˚C), and then want to adjust the stringency by washes at 48˚C. Which of the four mRNAs will remain hybridized to the probe following the wash?

Indicate how each of the following conditions would affect the Tm of nucleic acid hybrids; increase, decrease or no effect.

22. Increased temperature
23. Decreased formamide concentration
24. Higher A:T content
25. You wish to use a probe of 500 bases which is 45% G/C to perform a hybridization at 40^{o}C in a buffer containing 0.2 M NaCl. Assuming 100% complementarity, your hybridization buffer should contain what maximum percent concentration of formamide? Assume optimal hybridization will occur at 10^{o}C below the Tm.

Part II: (5 points/ question for a total of 20 points)

Amplification of the yeast ADH gene as a function of stringency

1. Submit a figure of the amplification of the ADH gene as a function of varying concentrations of K**+** (Taq buffer) and Mg**++** concentration ($MgCl_2$). Include an appropriate figure legend.

2. Provide an analysis of the results presented in the previous question. Your analysis should include the following:

    - Which conditions theoretically represented the lowest and highest stringency conditions? Explain your reasoning.
    - Compare the stringency conditions achieved with varying Mg++ as compared to K+.
    - Were the results observed consistent with the first two points? Why or why not. Compare predicted to expected results

PCR amplification of *Apo*C2 VNTR

3. Submit a figure of the VNTR profiles of the class. Include an appropriate figure legend, which includes a brief analysis. The analysis should include the sizes of the different alleles observed identified alphanumerically and whether each individual was homozygous or heterozygous.

PCR amplification of *Apo*B RFLP

4. Submit a figure of the RFLP profiles of the class. Include an appropriate figure legend, which includes a brief analysis. The analysis should include the different alleles observed labelled alphanumerically, and whether each individual was homozygous or heterozygous.

Bioinformatics (2 points/ question for a total of 30 points)

You should now be quite familiar with the NCBI site and be able to complete the following exercise with relatively few directives. Consider this a practice run for the bioinformatics section on the final exam.

Use the following sequence to answer questions 1-8

```
CCATGAACCCTCCCTCTACAAAGGTCCCCTGGGCCGCCGTGACTCTGCTGCTGCTCCTCTTGTTGCCGCC
CGCGCTGCTGTCGCCCGGGGCGGCCGCGCAGCCCTTGCCCGACTGCTGCCGCCAGAAGACGTGCTCCTGC
CGCCTCTACGAGCTGCTGCACGGCGCGGGCAACCACGCGGCCGGCATCCTCACGCTGGGCAAGCGGCGGC
CCGGACCCCCGGGCCTCCAGGGCCGGCTGCAGCGCCTCCTGCAGGCCAGCGGCAACCATGCGGCCGGCAT
CCTGACCATGGGCCGCCGCGCAGGCGCAGAGCCAGCGCCGCGCCCCTGTCCCGGGCGCAGATGTCCCGTG
GTGGCCGTCCCCTCTGCAGCGCCTGGAGGGCGGTCGGGAGTCTGAGCTGACGCTCGCGCCGCGTCCTGGC
CCTGCCCTCTCCTCGCTGCCCGCCGGA
```

1. What is the name of the organism and of the **gene** (NOT predicted) from which this sequence was obtained?

2. What is the accession number of the first predicted record of a nucleotide homologue from *Felis catus*?

3. What is the name of the protein product encoded by the *Felis catus* sequence?

4. What is the percentage identity between the protein product from the *Felis catus* and the protein product encoded by the gene found in question 1?

5. What is the description of a conserved domain associated with the *Felis catus* protein?

6. The sequence from the organism in question 1 was cloned into the *BamH*I site of pUC19. What fragment sizes would be expected following a digest with *Pst*I of recombinants containing the insert of interest in either orientation?

7. Which of the following primers could be used to perform a reverse transcriptase reaction on the mRNA sequence from *Felis catus*?

   A. AGACGAGTCTCTGTTAGTGAA
   B. TTCACTAACAGAGACTCGTTG
   C. GGAACTTTGAGGCTGTTGACT
   D. ACTAGGAGAGGGCAGGGTCA

8. Using the initial sequence as a translated query to search a protein database, which organism of a different genus has the most similar protein with a similar function?

9. A Southern hybridization was done on *Hinc*II digested genomic DNA from a new variant of the Hepatitis B virus. The digested DNA was probed with the region corresponding to bases 807 – 1894 of the original Hepatitis B genome (Accession number NC_003977). The Southern hybridization revealed bands of 1465 bases and 483 bases. Which band (s) is (were) unexpected?

10. Consider the information given in the previous question. The mutation (s) most probably occurred in the gene which codes for which **protein** of the Hepatitis B virus?

Refer to the record with the accession number NM_010030 to answer questions 11-15

11. What is the accession number of a *Mus musculus* nucleotide homologue of this sequence which has a different gene name?
12. What type of homologues are the proteins encoded by the original query sequence and that obtained in question 11?
13. What is the percentage amino acid identity between the two proteins discussed in the previous question?
14. What type of amino acid substitution is most abundant between the two homologues (discussed in question 12?
15. The following sequence: GGACCGTCTTCCTCCTGTTTACCGAGACTGTGTCATGG represents the original sequence, the reverse sequence, the complement sequence, or the reverse complement sequence of the sequence with the accession number NM_010030?