BPS3101

Genomics

Midterm 1

Oct 6, 2017 11:30 am - 12:50 pm

NAME: ___________**KEY**_________________

STUDENT ID#: _____________________________________

Read carefully:

Cellular phones, unauthorized electronic devices or course notes (unless an open-book exam) are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: academic fraud allegations will be filed which may result in your obtaining a **0** (zero) for the exam.

By signing below, you acknowledge that you have read and ensured that you are complying with the above statement.

Signature: _________________________________

Multiple choice questions. Note there may be more than 1 correct answer. Select all that are correct!

1. The human genome is __________ Mb in size and has approximately ___________ genes.
    a) 3, 35 000
    b) 30, 22 000
    c) 300, 30 000
    d) 3000, 20 000

2. The most abundant RNA in a cell is:
    a) microRNA
    b) messenger RNA
    c) ribosomal RNA
    d) snRNA
    e) siRNA

3. Isoschizomers:
    a) generate different sticky ends.
    b) are restriction enzymes that only cut one strand of DNA.
    c) recognize the same DNA sequence.
    d) were isolated from the fungi *Schizosaccharomyces pombe.*
    e) generate identical cohesive ends.
    f) are selective for RNA over DNA.

4. A Northern blot can use a __________ probe to detect ____________.
    a) DNA, DNA
    b) RNA, DNA
    c) RNA, peptide nucleic acid
    d) Oligonucleotide, DNA
    e) Oligonucleotide, RNA

5. A Southwestern blot requires:
    a) a primary and secondary antibody.
    b) SDS-PAGE to separate proteins
    c) proteins to be denatured before they bind RNA.
    d) proteins to be renatured before they bind DNA.
    e) was developed in Arizona.

6. The disadvantage of using a poly(T) primer for cDNA synthesis is that it
    a) only works for bacterial transcripts.
    b) will enrich samples with 5' UTR sequences.
    c) may fail to amplify past the 3' UTR.
    d) will not capture the introns

7. A genetic map is based on ______________ and a physical map is based on ______________.
    a) Sequences, banding patterns of chromosomes
    b) Recombination frequency, heterozygous genes
    c) STS, recombination frequency
    d) Recombination frequency, sequence
    e) Sequences, SNPs

8. SNPs are:
    a) Small nuclear proteins.
    b) Single nucleotide polymorphisms.
    c) never missense mutations.
    d) typically in non-coding DNA.

9. The first gene that was sequenced was:
    a) MS2 coat protein
    b) Alanine tRNA
    c) a DNA gene
    d) 504 nt long
    e) sequenced by Fred Sanger

10. Illumina sequencing uses _______________ to generate the input DNA library
    a) A plasmid clone library
    b) emPCR on a glass slide
    c) A λ phage library
    d) RT-PCR on a glass slide
    e) Bridging PCR on a glass slide

Long answer questions

1. Two techniques were discussed in class for quantifying how many copies of a specific DNA sequence was present. Name both techniques (4pts). Describe how one of these techniques works (6pts).

qPCR and ddPCR

qPCR explaination

real-time PCR/qPCR

need gene specific primers polymerase and a dye to detect dsDNA

Thermocycle to amplify DNA and measure product after each cycle

Cycle when product crosses the threshold for detectability is called the Ct are relates to initial [DNA]. Figure can help.

ddPCR explaination

uses emulsion PCR (emPCR) with one template per droplet. Need gene specific primers polymerase, and dye to detected dsDNA . Instrument counts dye positive and negative droplets. Positive droplets over total droplets correlataes to initial [DNA].

2. Suppose that the yeast gene for protein X is located within the sequence shown below.

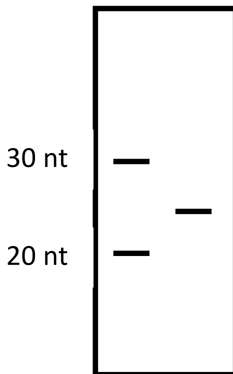
```
5’ …GTATTTAAAGCGCCTTCATTAAAGGCATCGTATTAGTCGCGA… 3’
3’ …CATAAATTTCGCGGAAGTAATTTCCGTAGCATAATCAGCGCT… 5’
```

a) Give the amino acid sequence of protein X and a brief rationale as to why this was protein X. Note there is a codon table at the back of the exam. (3 pts)

MPLMKAL or Met-Pro-Leu-Met-Lys-Ala-Leu

Rationale – only reading frame with with start and stop codons.

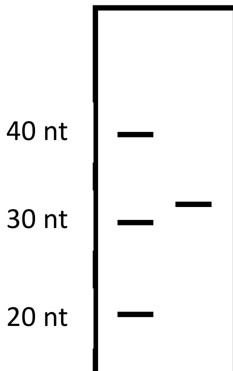
b) Design an oligomer probe (6’mer in length) for a northern hybridization experiment studying gene X expression. Give its sequence with 5’ and 3’ designations and show your expected hybridization results in the box below with a brief explanation. Include size markers. (4pts)



TTAAAGCGCCTTCATTAAAGGCAT any six of these

Probe would detect mRNA – no UTRs info so just used from start to stop.

c) You then use the same probe from b in a Southern hybridization experiment where the DNA has been digested with CfoI (5’ GCGC 3’ recognition sequence). Show the expected result with size markers and give your rationale. (3pts)



Single CfoI site.

Cuts dsDNA into 10 and 32 bp fragments – use +/-2 in length

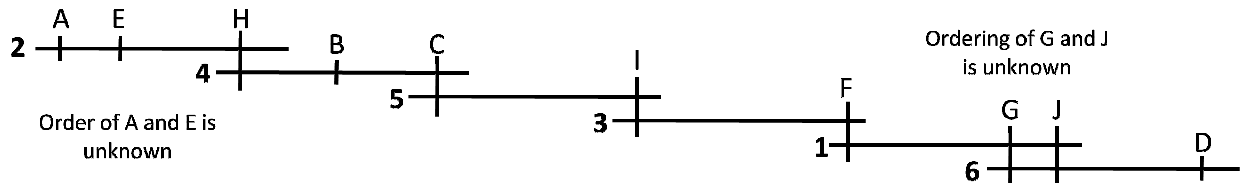
Depending on where their probe is it should either show

a 32 or 10 bp fragment

3. A PhD student is characterizing the genome of a new bacteria, *Pseudomonas sp,* that produces a powerful antifungal agent. She has a clone library of six BAC clones (1-6) and has identified ten STSs (A-J). In the chart below a + symbol indicates that the BAC clone has that STS.

BAC	A	B	C	D	E	F	G	H	I	J
1	-	-	-	-	-	+	+	-	-	+
2	+	-	-	-	+	-	-	+	-	-
3	-	-	-	-	-	+	-	-	+	-
4	-	+	+	-	-	-	-	+	-	-
5	-	-	+	-	-	-	-	-	+	-
6	-	-	-	+	-	-	+	-	-	+

a) Explain how you would use these data in mapping the Pseudomonas sp genome and show your results in a clearly labelled diagram. Does this complete the mapping project? If not, what strategy would you use next (6pts)



No. Orders of A&E and G&J cannot be completely determined. If its a closed circular genome no clone from D to A/E.

Make a second library and rescreen for order of A or E with H but not both. Same for GJ.

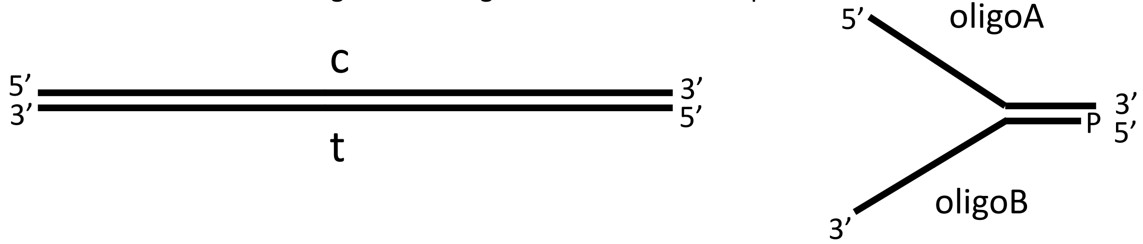
b) Suppose the *Pseudomonas sp* genome is 500 times smaller than the human genome. How many *AvaI* restriction sites would, you expect it to have. The AvaI restriction site is CPyCGPuG (Py = pyrimidine, Pu = purine). Show your calculations and briefly explain what factors might make your calculated value differ from the true value? (4pts)

6 -6.5 Mb genome. Frequency of AvaI site: 1 every 4x2x4x4x2x4 = 1024bp

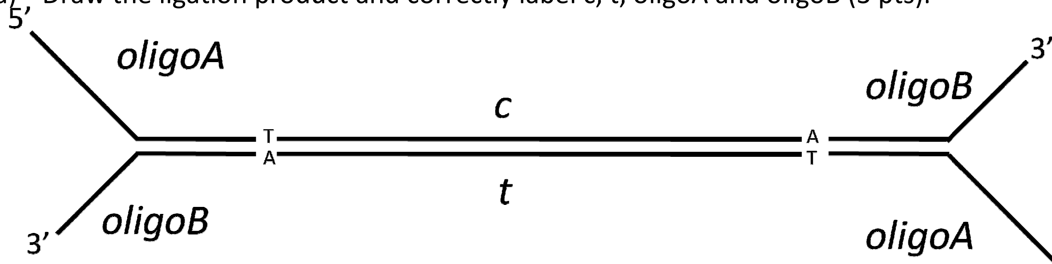
Total number of sites = any number between 5750-6250

If the GC content is > 50% than there will be more AvaI sites than predicted by this approximation.

4. You are constructing a library for illumina sequencing. Your end repaired phosphorylated dsDNA is shown. Top strand is labelled c for coding and the bottom strand t for template. You A-tail it with the Klenow fragment and ligate on the forked adaptor shown.

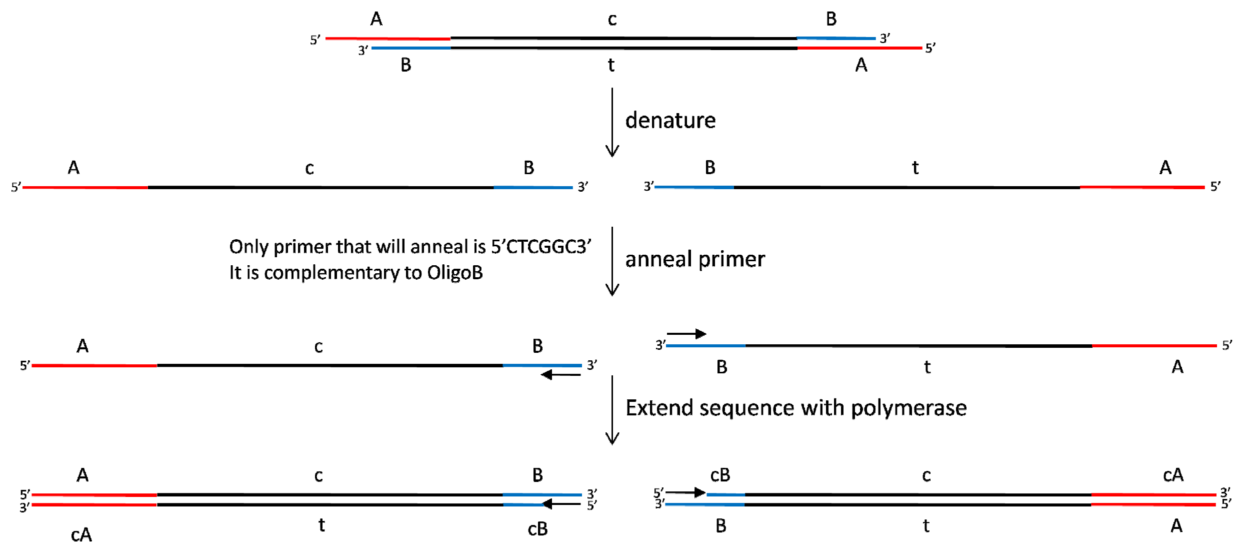


a) Draw the ligation product and correctly label c, t, oligoA and oligoB (3 pts).



b) If this is the sequence of the forked adaptors, draw a cycle of denature, anneal and elongate with this primer pair 5'CTCGGC3' and 5'ACACTC3' on the ligation product from above. Show each step. Label your ssDNA and dsDNA from your drawing with c, t, oligoA, oligoA_comp, oligoB, and oligoB_comp as appropriate. Note: you do not need to write out the sequences for oligo A and B. (7pts)

```
Oligo A   5' ACACTCTTTCCCTACACGACGCTCTTCCGATC*T 3'
                                     ||||||||||||
Oligo B   3' GAGCCGTAAGGACGACTTGGCGAGAAGGCTAG-P 5'
```



Extra page – can be used as scrap paper or if you need more space to answer any question.

Codon table

1st position	2nd position: U	C	A	G	3rd position
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	stop	stop	A
	Leu	Ser	stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Amino Acids