29 pts 1. Genome Evolution and Phylogenetics:

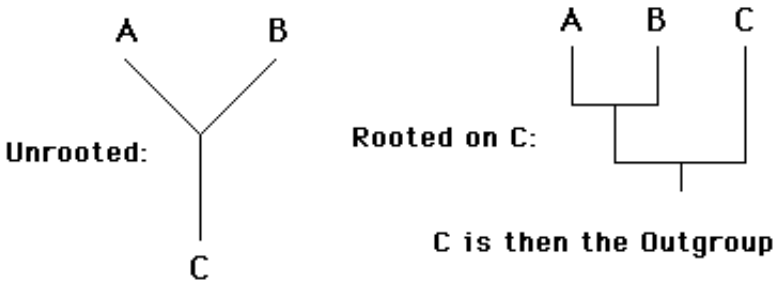
(12) a. What are the three main ways by which organisms duplicate genes? Provide an example of each.

Duplication by
Entire Genome, e.g. Yeast
Single Chromosome, e.g. Human Trisomy 20 (Mongolism)
Individual or Groups of Genes, e.g. Human globin genes

(4) b. Briefly describe how transposons can promote gene duplication.

The main way is that when transposons hop from one regions to another in a given genome, they yield repeated regions. These can then be used for gene duplication via unequal crossing-over or by unequal sister chromatid exchange.

(8) c. Draw an unrooted tree with three external nodes A, B, C, and draw the rooted tree where the root is on the branch leading to node C. Which node is then the outgroup?



(5) d. Briefly explain how the Lucy phylogenetic analysis supported the hypothesis that Homo sapiens evolved from Homo erectus in Africa.

The Lucy phylogenetic analysis purports to show that the origins of Homo sapiens reside once and only once in a black female in Africa.

This indicates that Homo sapiens evolved from Homo erectus in Africa and then migrated out of Africa, rather than evolving from Homo erectus several times in various parts of the world, eg Europe and Asia AFTER Homo Erectus migrated out of Africa.

20 pts 2. Post-Genomic Molecular Biology – Lander's Ten Goals:

(4) a. One goal is systematic identification of common variants in genes involved in human disease, which will lead to an Association Analysis in determination of disease susceptibility. What does this mean?

This means that the assumption that only a few polymorphisms in genes involved in human disease account for a majority of the disease, that these will be determined via this proposed identification of common variants, and that, once determined, these variants can be used upon birth with any human being to determine their life-long susceptibilities to these diseases by determination of all of their human disease gene variations.

(6) b. What two concepts underly development of "High ThroughPut" tools?

1) Minaturization in size, eg capilary gels for sequencing, grids for microarrays

2) minaturization in volumes, ie ml -> nl (nanoliters), with greatly reduced times for reactions

Result: automation of procedures; use of robots for this automation

(4) c. Define Proteomics.

Proteomics: complete determination of protein states throughout life cycle of a given cell: which protein are made at any given differentiation stage; concentrations; all post-traslation modified forms and their concentrations; all interactions of each variation of the protein with any and all other proteins and other molecules

(6) d. What is ELSI and why is it important?

ELSI: Ethical, Legal and Social Issues
These issues are clearly of great importance regarding intrusion on the private lifes of individuals, on their ability to obtain health care, on ownership of genome property, eg human genes, and on the impact of these technologies on human beings and their quality of life, eg GM-foods

24 pts 3. DNA microarrays:

(6) a. Briefly describe how DNA expression arrays differ from genomic oligonucleotide DNA arrays.

DNA expression arrays are 2D-grid arrays of ssDNA obtained from a cDNA library and immobilized via spotting usually on a glass slide or a nylon membrane. The number of spots is typically rather small.

Oligo DNA arrays are typically genomic arrays, and are 2D-grids at high density of 25-mers or so directly synthesized on the grid. This synthesis is done via masking procedures developed by the silicon valley industry

(4) b. Briefly explain how each gene is assayed more than once in an oligonucleotide array.

Oligos are designed via knowledge of the genome sequence to several parts of the exons of a given gene. These oligos are all to the same gene and are constructed side by side on the oligo array

(4) c. What are the "control" oligonucleotides used in an oligonucleotide array?

"Control" oligos used in an oligo array are oligos which are constructed with a one base mismatch near the middle of the 25-mer, thereby greatly reducing the probability that hybridization will occur to the correct sequence. These then measure the "background" hybridization in the experiment.

(4) d. A tree approach is used in a "Clustering" analysis of expression array approach. Briefly explain how the lengths of the tree are determined.

The tree lengths, leading to the branch lengths, are directly proportional to the similarity of the expression array signals shown by two genes. Thus, if these signals are nearly identical, the tree lengths are short; if they are very different, the tree lengths are long. Algorithms have been developed that relate one to the other.

(4) e. What relationship do these "Clusters" have to gene expression changes that are seen in an expression array experiment?

Groups of genes that belong to a given Cluster, ie show a similar expression profile in response to a given change in environment, are often found to share common metabolic function, ie are often found to be in common metabolic pathways, or to share common physiological function, eg all be involved in protein biosynthesis.

(2) f. What is the primary advantage of a cDNA microarray experiment over a Northern experiment?

A cDNA microarray experiment can monitor the response of ALL genes, at the mRNA level, to a given environmental change.

A Northern can only compare the number of mRNA species or conditions that can fit within the number of columns of a given gel, eg about 15. This is often at the level of a single gene.

21 pts 4. Post-Translational Modification:

(6) a. What are chaperones and what is their function?

Chaperones are proteins whose function is to catalyze 3D folding of other proteins by helping these other proteins find their correct 3D structure.

(9) b. What are the three main ways in which proteins are modified after their synthesis?

1. Proteolytic cleavage: 5' end, sometimes 3'end, sometimes internal (polyprotein)
2. Intein splicing
3. Chemical modification, e.g. phosphorylation events

(6) c. Give an example with brief descripion of how each of the following types of protein modification is important in the function of the protein:
1) phosphorylation 2) acetylation

Many possibilities ... two examples:

1) phosphorylation: essential to function of Thr/Ser kinases or Tyr kinases and protein cascades in signal transduction

2) acetylation: acetylation of histones in nucleosomes is important in regulation of what regions of chromosomes are transcribed.

(12) 5. Briefly describe a Two-hybrid System experiment.

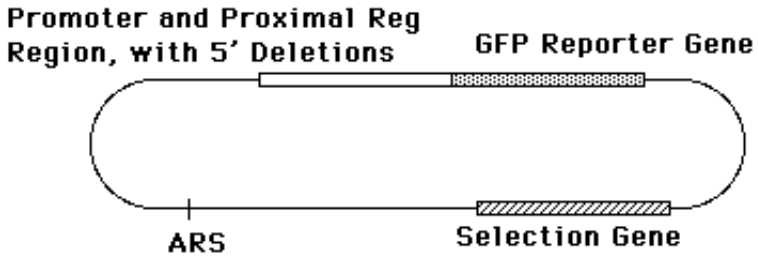
Fusion proteins are expressed from two chimeric plasmids, one containing a fusion of gene encoding "bait" protein, the one for which interacting proteins are sought, with the DNA-binding domain of a transcription factor (TF), and the other containing a fusion of cDNAs encoding "fish" proteins, those that might interact with "bait" protein, with the activation domain of the TF. A yeast host is used that has the receptor element (DNA binding site) for the TF controlling expression of a reporter gene, eg His gene, GFP, lacZ. This host is transformed with the two plasmids, selecting for the two plasmids, and then host cells are screened for expression of the reporter gene, eg for His gene, growth in absence of histidine. Cells which express the reporter gene likely have a cloned cDNA encoding a protein that interacts with the "bait" protein.
See figures at Web site for BIMM100 at:
http://www.biology.ucsd.edu/classes/bimm100.FA00/20.EukGeneReg.html#C4

45 pts 6. You perform a 5'-deletion analysis of the proximal regulator region of a eukaryotic gene.

(2) a. What is the primary reason for doing such an experiment?

To determine the nucleotide positions and sequence of Response Elements or DNA binding sites in the proximal regulator region

(10) b. To assay the effects of your 5'-deletion mutants, you make reporter plasmids, each of which contains one of your deletion mutants. Diagram your reporter plasmid and label ALL important parts of the reporter plasmid, using GFP (gene for Green Fluorescent Protein) as the reporter.



(6) c. What is meant by the phrase "using GFP as the reporter" and what is the assay?

This means that the gene encoding GFP is used as the gene under control of expression by the proximal regulator region in which the 5'-deletion mutations were generated.

The assay is for expression of the GFP gene, and this is done via the Green Fluorescence shown by the GFP protein when irradiated with nearUV or visible light.

You find GFP expression for 5'-deletions up to -200, partial GFP expression for a -195 deletion mutant, and no GFP expression for 5'-deletions -185 or more.

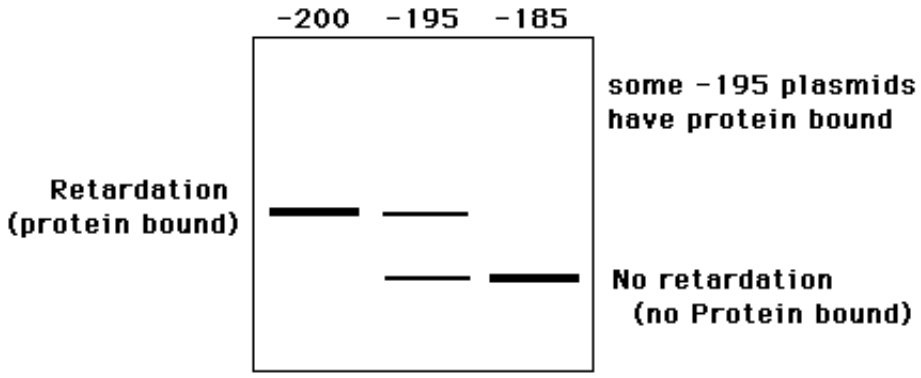
(2) d. What do the negative numbers mean?

The negative numbers mean the nucleotide positions in the proximal regulator region UPSTREAM of the Initiator (Inr) site for transcription, which is given position 1 by definition.

(4) e. What is the correct interpretation of these GFP expression results?

These data indicate that a binding site or Response Element for a transcription factor that acts as an Activator is present in the proximal regulator region between nucleotides -200 and -185 (partial binding of the Activator occurs if nucleotides -195 to -185 are present on the reporter plasmid).

(9) f. Diagram below the agarose gel results of a gel retardation experiment with these 3 plasmids and the regulatory protein involved.



(6) g. The regulatory protein has two primary domains. What are the functions of these domains?

1) Transcription activation
2) DNA binding

(6) h. What results in the Karin et al Metallothionein gene paper led to the conclusion that there were TWO MRE regions?

1) 5'-deletion mutations could come all the way in to less than -50 and the Cd++ response was still present.

2) 3'-deletion mutations could go all the way out to about -150 and the Cd++ response was still present.

Since these regions overlap, the simplest explanation was that TWO MRE sites existed and that BOTH had to be deleted before the Cd++ response was lost.

23 pts 7. Post-Transcriptional Modification:

(8) a. Briefly describe the G capping reaction in eukaryotes.

1) GTP is added in 5'-5' linkage to 5' nucleotide, with loss of PPi from GTP and Pi from the mRNA, yielding 2 phosphates between the G nucleoside and the 1st mRNA nuc.

2) Cap 0: a methyl group is added to the G-7 position

3) Cap 1: the 2'-OH of the ribose of the 1st nuc is methylated about 85% of the time

4) Cap 2: the 2'-OH of the ribose of nuc 2 is methylated about 10-15% of the time.

Some other methylations can also take place.

(6) b. What are the major differences in transcription termination between prokaryotes and eukaryotes?

Prokaryotes: transcription termination occurs at specific sites in either a rho-dependent or a rho-independent manner. Protein rho is used at rho-dependent sites, and a stem-loop + row of U's is used as mRNA structure in rho-indep sites.

Eukaryotes: transcription just terminates at pause sites with no protein factors or mRNA structure at many possible sites that are past (on 3' side) the PolyA cleavage site for the gene. (pre-mRNA is processed via cleavage at the polyA cleavage site, with addition of ~ A residues at 3' end).

(9) c. What are the three major steps in splicing catalyzed by a spliceosome?

1) 1st transesterification reaction: 2'-OH of A internal to intron attacks the 5' end of the intron.

2) 2nd transesterification reaction: 3'-OH of exon 5' to intron attacks the 3' end of the intron, joining the two exons and releasing the intron as a circularized lariat molecule.

3) Released lariat form of intron is now degraded in a series of steps.

20 pts 8. Regulation of prokaryotic gene expression:

(6) a. What are the primary differences in repressor regulation of a catabolic (breakdown) operon and an anabolic (biosynthetic) operon?

Catabolic: regulator is Repressor protein that is <u>inactivated</u> by a small molecule Inducer molecule (usually the sugar involved)

Anabolic: regulator is Repressor proteint that is <u>activated</u> by a small molecule CoRepressor molecule (often the end product made by the proteins encoded by the anabolic operon, eg trp)

(6) b. Briefly explain how Catabolite Activation works. Include why this is also called the "Glucose Effect"

Catabolite Activation: cAMP plus Catablite Activation Protein (CAP) are required in a complex to activate many catabolic operons, eg lac operon. In low glucose, cAMP conc is high, and vice versa; hence, control of lac operon expression via glucose levels, and the term "Glucose Effect". (Thus, when glucose is present, lac operon is turned off, and vice versa).

(4) c. Briefly explain why the lacI^s repressor is dominant to lacI$^+$ and lacI$^-$ repressors.

In a lacI^s/lacI$^+$ merodiploid, the LacI^s repressor will bind to the lacO operators even in the presence of Inducer (it is a SuperRepressor). Thus, the merodiploid cell looks phenotypically like a lacI^s cell (no induction), and lacI^s is dominant to lacI$^+$.

In a lacI^s/lacI$^-$ merodiploid, the LacI^s repressor will bind to the lacO operators even in the presence of Inducer (it is a SuperRepressor). Thus, the merodiploid cell looks phenotypically like a lacI^s cell (no induction), and lacI^s is dominant to lacI$^-$.

(4) d. What would the dominance relations be for the repressors in (c) if lacI encoded an activator protein?

Here, lacI^s would encode a SuperActivator, and the lac operon would be constitutively expressed. A lacI$^-$ cell would encode a nonfunctional Activator, and the lac operon would be noninducible. A lacI$^+$ cell would induce the lac operon via Inducer activation of the LacI$^+$ Activator Protein.

Thus, in a lacI^s/lacI$^+$ merodiploid, the LacI SuperActivator would activate in absence of inducer, yielding constitutive expression. And: lacI^s is dominant to lacI$^+$

And, in a lacI^s/lacI$^-$ merodiploid, the LacI activator would be inducible, with the cell phenotypically looking like a LacI cell. Thus: lacI$^+$ is dominant to lacI$^-$

NOTE: same dominance relations as for case where lacI encodes a Repressor !!

37 pts 9. tRNA:

(12) a. Briefly describe, or draw, the secondary AND tertiary structure of a tRNA molecule. Include names of structural features.

Secondary Structure: see Brown, Fig 10.2

Tertiary Structure: see Brown, Fig 10.3

(9) b. What are 3 major differences between how Group I and Group II aminoacyl tRNA synthetases function with a tRNA molecule.

Group I: 1) binds minor groves of acceptor and anticodon stems; 2) ATP binding domain a nucleotide binding domain at N-term end of enzyme; 3) anticodon arm binding domain at C term end; 4) synthetase attaches aa to the 2'-OH of the ribose of the 3 terminal nucleotide (the A of the -CCA) of the tRNA

Group II: 1) binds major groves of acceptor and anticodon stems; 2) ATP binding domain more C-term; 3) anticodon arm bound in major groove by N-term domain; 4) synthetase attaches aa to the 3'-OH of the ribose of the 3 terminal nucleotide (the A of the -CCA) of the tRNA

(4) c. What is the Wobble Hypothesis?

Hypothesis that base pairing of 3rd posn bases of the codon and the 1st posn of the anticodon can be "nonstandard" due to "wobble" of the 1st posn nucleotide of the anticodon. Such wobble is due to the curvature of the anticodon in the loop of the anticodon arm of the tRNA molecule.

tRNA mutations can sometimes suppress nonsense mutations.

(6) d. Define suppression and give an example of intragenic suppression using frameshift mutations.

Suppression mutation is a Second Mutation that reverses partially or completely the effects of a First Mutation.

Example of Intragenic Suppression using Frameshift mutations:
Intragenic suppression: both mutations are in the same gene.
Frameshift mutations: insertion or deletion mutations of (usually) one nucleotide
Suppression: insertion suppresses (reverses effects of) deletion.
Thus, those Crick +- or -+ double mutants that showed a wildtype phenotype.

(6) e. What is a nonsense mutation and how can they be suppressed by such tRNA mutations?

Nonsense mutation: mutation of a sense codon (encodes an amino acid) to a nonsense or stop codon.

tRNA mutants can sometimes suppress nonsense mutations via an anticodon mutation in the tRNA gene such that the mutated tRNA is loaded with an amino acid at acceptor arm, but anticodon reconizes the stop codon, at least part of the time.
Result: the amino acid is now put into the protein at this codon, rather than have the remature termination of synthesis of the protein.

22 pts 10. Translation:

(12) a. What are the primary differences in translation initiation between prokaryotes and eukaryotes? Do NOT list initiation factors involved.

Prokaryotes: Small Ribosome binds mRNA at Shine-Delgarno sequence in mRNA, THEN fMet-tRNAf is bound in P site of small ribosome. Met attached to tRNAf is formylated.

Eukaryotes: Met-tRNAi is bound to P site of small ribosome, THEN mRNA is bound, at 5'-end of mRNA. Met-tRNAi-Ribo complex moves 5'->3' down mRNA until start codon is found (usually the first AUG). Met attached to tRNAi is NOT formylated

(6) b. During translation elongation, what two groups of the complete ribosome complex move relative to each other?

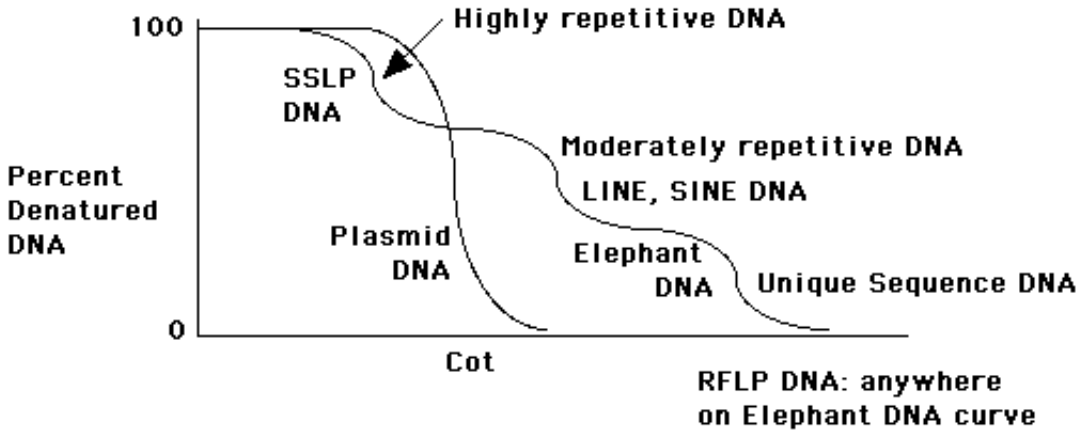
the peptide-tRNA-mRNA complex moves relative to the Small-Large Ribosome complex.

(4) c. What is the molecule called Peptidyl Transferase and what is its function?

Peptidyl Transferase is the enzyme that catalyses peptide bond formation during protein biosynthesis. The molecule that does this is now thought to be the large rRNA species in the Large Ribosome (23S in prokaryotes, 45S in eukaryotes), which then is a Ribozyme.

24 pts 11. Cot curves:

(9) a. Draw a Cot curve for elephant DNA, assuming a DNA complexity similar to human DNA, and indicate what each part of the curve means in terms of complexity of the DNA. Label your axes.



(3) b. Now superimpose on your drawing a Cot curve for a typical plasmid cloning vehicle. Label your curves.

(12) c. State, or show on your Cot curve drawing, where you would find SINE, LINE, RFLP, and SSLP DNA species, and briefly explain why.

SINE: Short INterspersed Element, eg Alu fragments ... part of Moderately Repetitive DNA, due to many copies present in genome

LINE: Long INterspersed Element, eg retroposons ... also part of Moderately Repetitive DNA, due to transposition and many copies present in genome

RFLP: Restriction Fragment Length Polymorphism ... usually point mutation in a restriction site ... can occur in ANY class of DNA, hence part of Highly Repetitive, Moderately Repetitive, and Unique Sequence DNA classes.

SSLP: Simple Sequence Length Polymorphisms ... polymorphisms that involve variable number of tandem repeats, where the repeats are short (2-7 bp) ... hence part of Highly Repetitive DNA.

23 pts 12. DNA metabolism:

(9) a. What are the three major parts of a eukaryotic DNA replication origin and what are their functions?

See the DePamphilis J. Article 3 for the course

Three major parts:

1) A + B1 region: bind ORC proteins to yield initiation complex

2) B2 region or DUE region: DNA Unwinding Element - A,T-rich region that unwinds to provide site (OBR - Origin of Bidirectional Replication) for initiation of daughter strand synthesis.

3) Sites, eg yeast B3, on one side or other or both sides of OBR for binding of Transcription Factors, that either activate or inhibit initiation from this rep origin

(4) b. Briefly describe transcription-coupled repair.

See the Hanawalt J. Article 4 on Transcription-Coupled Repair.

Transcription coupled repair is the selective repair of DNA damage found:
1) with genes and exons encoding proteins, and
2) found on the DNA strand used as template for mRNA synthesis

This is done by a coupling of nucleotide excision repair (Exinuclease; UvrABC) with transcription. RNA pol pauses at damage in DNA, backs up, repair occurs. In prokaryotes, RNA pol falls off and begins transcription again. In eukaryotes, RNA pol II resumes original transcription. TFIIH, as part of both RNA pol II and Exinuclease, is probably important in this coupling process.

(4) c. Telomerase is essentially absent in mature adult human cells. How is this detremental to humans?

The fact that Telomerase is essentially absent in mature adult human cells means that the ends of the DNA for each chromosome get progressively shorter with each round of DNA replication and cell division. Eventually, the telomeres will be lost, and genes will start to disappear, genes that are needed by humans. Such loss may contribute to the aging process

(6) d. Define a ribozyme and briefly explain why Telomerase is NOT a ribozyme.

A ribozyme is an RNA molecule that has catalytic activity, ie an RNA molecule which by itself is an Enzyme.

Telomerase is NOT a ribozyme because, although Telomerase is a Ribonucleoprotein (part protein and part RNA), the catalytic activity resides solely with the Protein component of Telomerase (the RNA serves a template role only for the synthesis of the telomere DNA T,G-rich repeats).

HAPPY HOLIDAYS !!!