34 pts 1. Cloning Vehicles:

(8) a. Define a Cloning Vehicle. Include required features of a Cloning Vehicle in your answer.

Cloning Vehicle: a DNA molecule capable of replication / self-duplication (a replicon) used for Cloning purposes.

Required features: 1) origin of replication
2) selection marker, eg antibiotic resistance
3) at least one cloning site

(6) b. What features does a YAC cloning vehicle require that a simple Plasmid vector does not?

Additional YAC required features: 1) CEN marker - centromere, for meiotic stability
2) SEPARATE selection markers for EACH of the two YAC arms
3) TEL (telomere) regions at non-cloning ends of each YAC arm, for linear chromosome stability

(12) c. Give four major reasons why Phage Lambda is a good cloning vehicle for genomic DNA libraries.

Any four of the following:
1) clone large inserts
2) in vitro packaging yields intact phage
3) in vitro packaging selects for insert
4) in vitro packaging selects for insert size
5) store nearly indefinitely (many years) in refrigerator
6) Store phage in liquid at high titer
7) Infection highly efficient
8) Screen plaques at high density
9) Easy to isolate phage, purify DNA

(8) d. Briefly describe the experiment in the Burke-Olson YAC paper in which it was shown that YAC HY1 contained a single SmaI insert. Draw the results of the Southern gel, indicating assay used in the gel.

DNA was gently isolated from yeast cells containing YAC HY1 and either cut with SmaI or not. Cut and uncut DNA were run out on a Southern gel, and probed either with pBR322 DNA or with total human DNA. The human DNA probe showed a single band in either the cut or uncut DNA, indicating that only one SmaI human DNA fragment had been cloned.

See the Soft Reserves key for figure of the Southern gel

20 pts 2. Finding genes in Genomes:

(6) a. Define an ORF. Why are ORFs useful in finding genes in Genomes?

ORF - Open Reading Frame: one of the six reading frames in a DNA molecule which contains NO Stop Codons but ends with a Stop Codon.

A Gene, with or without Exon structure, must be an ORF. Finding ORFs then delineates genome regions that contain the Genes. This eliminates a lot of DNA regions and a lot of Reading Frames in all DNA regions in the search for genes.

(4) b. Briefly describe how Homologies are useful in finding and identifying genes.

Given an ORF, if one now finds a Homologue from another organism (a gene or cognate protein that shows a convincingly high enough level of sequence similarity to the ORF), then this gives strong evidence that the ORF is actually a gene, one that descended from an ancestral gene common to the homologous gene found in the other organism.

Further, if one knows the function of the gene in the other organism, it is highly likely that the function of the ORF will be similar. Thus, one gets predictive information regarding function and identification of the ORF as well.

(6) c. What is the difference between Homology and Similarity?

For Homology, two genes are Homologues if they descended from a common ancestor. This is a qualitative characteristic: they either have a common ancestor, or they don't. Further, there is no methodology to determine homology directly; rather, it is deduced from other properties such as degree of similarity of two sequences.

Similarity is a measure of degree of likeness between two gene sequences, using specific criteria for comparison of the sequences. This is a quantitative measure, provided by the criteria, permitting one to determine a percent similarity. One often uses such level of similarity to make deductions concerning homology.

(4) d. Briefly describe the BEST way of finding genes in Genomes experimentally.

Genes are expressed via RNA into protein or structural RNA. Hence, the best way of identifying genes in genomes is to compare the sequence of the expressed regions, via cDNA sequences, with the genomic sequences. cDNA sequences give exon sequences, which by comparison with genomic sequences provides complete experimental information on positions of Exons and Introns.

18 pts 3. Genome Sequencing:

(6) a. Briefly describe the dideoxy sequencing procedure. Why is it limited to sequence of about 500 bp?

In didexoy sequencing, 4 DNA polymerase reactions are carried, each with DNA primer and template for synthesis of the DNA to be sequenced, and each with a different ddNTP. Incorporation of the ddNTP causes chain termination, resulting in a "nested set" of ssDNA fragments after denaturation, each terminating in one base in each of the 4 reactions. One of the 4 dNTPs is radioactively labeled, yielding labeled synthesized DNA fragments. These are size fractionacted on a polyacrylamide sequencing gel, and the sequence is read from the gel.

This is limited to about 500 bp, because the DNA fragments get increasingly closer together as they get larger. 500 bp is the limiting length: band width is close to distance between bands.

(6) b. In Shotgun Sequencing, one desires about a 10-fold Redundancy in sequencing each nucleotide. What is meant by Shotgun Sequencing and why is this Redundancy desirable?

Shotgun sequencing: break a DNA molecule into many fragments, sequence each fragment, construct the sequence of the original DNA molecule by finding "overlaps" in sequence between sequences of each fragment; joining such overlaps yields larger "contigs" (Brown, Fig 2.1).

Redundancy is a measure of how many times each nucleotide in the original DNA molecule is present in cloned pieces, each of which is sequenced. A high redundancy is desirable because shotgun sequencing is a random process and the higher the redundancy the greater the probability that each nucleotide will be present on a sequenced fragment (no gaps; one final contig).

(6) c. What is the difference between Directed Shotgun Assembly and Clone Contig Assembly? What is the major similarity between the two Assembly procedures?

Directed Shotgun Assembly: shotgun the entire genome. Join contigs using Anchored Markers such as STSs.

Clone Contig Assembly: break genome down into mapped smaller segments, e.g. mapped cosmids or BACs. Shotgun sequence these smaller segments.

Major similarity: each uses Shotgun Sequencing and each uses Anchor Markers in the mapping parts of the Genome Assembly.

20 pts 4. Draw a C:G base pair as found in DNA. Label all parts of the base pair, and label 5' and 3' ends.

See Soft Reserves key or posted key for figure.

Also see Brown, Fig 1.1 and Fig 7.3

18 pts 5. DNA Microarrays and their uses:

(6) a. Briefly describe a cDNA Microarray and how it is used.

A cDNA Microarray is a 2D solid surface on which has been spotted and immobilized ssDNA from separate clones from a cDNA libary.

This is used by hybridizing to this 2D "grid" of spots of single-stranded cDNA cDNA made from mRNA isolated from cells treated in some way. This cDNA is fluorescently labeled. The extent to which hybridization occurs under different treatments provides information on genes expressed under these different treatments.

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

A cDNA microarray experiment detects changes in thousands of genes simultaneously in one experiment due to a given treatment, whereas in a Northern one determines changes typically in only one gene, that to which your radioactive probe is directed, due to the given treatment.

(6) c. Define a SNP, and briefly explain how SNPs are related to genetic disease.

A SNP is a Single Nucleotide Polymorphism, a base pair that is different among different individuals in a population.

Genetic disease is often caused by single base mutations within a gene, leading to loss of function, or overexpression, of a given protein. Such mutations are SNPs. In such case, individuals with the disease would have one allele of the SNP, unaffected individuals would have another allele.

(4) d. Briefly describe how Oligonucleotide Microarrays can be used to assay SNPs.

The Oligo Array would have oligonucleotides complementary to both alleles of a SNP, usually one under the other. Stringent hybridization, requiring a perfect sequence match for hybridization, is then used to determine which alleles a given individual has in their DNA.

18 pts 6. R.enzymes, DNA libraries, and Reverse Genetics:

(4) a. Briefly describe what is meant by a Reverse Genetics experiment.

Reverse Genetics proceeds from a protein to the gene encoding the protein

The main steps in such an experiment:
1) Determine a partial amino acid sequence of the protein
2) Use this to generate oligonucleotide probes
(a degenerate set, to account for degeneracy of the genetic code)
3) Use these probes to select a genomic clone from a genomic library.
4) Sequence the DNA from the selected clone and find the coding sequences.

(4) b. How would a Lambda Genomic Library be used in such an experiment to isolate a gene?

If the Genomic Library is a Lambda library:
1) the lambda library is plated out at high density on a bacterial lawn on plates.
2) The resulting "lawn of plaques" is copied to a membrane filter.
3) The phage on the membrane are disrupted and the DNA denatured by alkali.
4) The oligo probes are used to identify position of the desired Lambda clone.
5) This clone is isolated using this approach.
6) The phage containing the cloned DNA are grown up, the DNA isolated, and the DNA sequence determined.

Sau3A (R.site: 'GATC) genomic fragments are cloned into the BglII (R.site: A'GATCT) site of a Lambda Cloning Vehicle used for 20 kb inserts to yield a Lambda Genomic Library.

(2) c. Why would this yield an approximately "random" distribution of genomic fragments in the library?

Sau3A sites are on average separated by 256 bp since Sau3A is a 4-cutter. 256 bp is much less than 20 kb; hence the Sau3A end points of the 20 kb fragments are distributed approximately "at random" on the genome.

(2) d. Why can one use these two enzymes in this library construction?

Both Sau3A and BglII yield single-strand 5'-overhangs which are GATC overhangs. Hence, Sau3A fragments will hybridize to BglII fragments, permitting the chimeric DNA construction.

(6) e. In vitro packaging is used in the library construction. What selection procedures are used for the resulting chimeric DNA molecules, and what size inserts does one obtain?

In vitro lambda packaging selects for:
1) lambda DNA that has a cos site
2) "headfuls" of DNA; hence, insert DNA, since the lambda cloning vehicle is not big enough.

Size of cloned inserts:
1) Total chimeric DNA must be 50 kb plus/minus 10%, and 10% of 50 kb is 5 kb.
2) The Lambda Cloning Vehicle is designed to clone on average 20 kb inserts.
3) Hence, the insert sizes will be 20 kb plus / minus 5 kb

12 pts 7. Physical vs Genetic Maps:

(5) a. How can a Genomic Clone Library with Microsatellite markers be used to generate a GENETIC map?

GENETIC maps are determined via Recombination Frequencies, or probabilities that two "markers" are close together. If two Microsatellite markers are close together, they will have a relatively high probability of appearing on the same clone in a genomic clone library. If they are far apart, this probability will be much lower. Hence, the relative probabilities of two markers appearing on the same Clone in a Genomic Clone Library can be used to generate a Genetic Map.

(5) b. How can a Genomic Clone Library with R.site markers be used to generate a PHYSICAL map?

One identifies the R.sites on different clones in the Clone Library, and then uses the positions of these R.sites to identify which clones overlap which other clones. When this is done for the entire genome, one has mapped the clones to the genome, and one has a Physical Map of the genome consisting of a Restriction Map. For example, see Brown, Fig 4.16A.

(2) c. What is the highest resolution genome physical map?

The DNA sequence of the genome

10 pts 8. Microsatellites:

(2) a. Write down an example of a Microsatellite of repeat length 2.

CACACACACACACACACACA

(4) b. What is a polymorphism, and what feature of Microsatellites gives rise to polymorphisms?

A Polymorphism is a marker on the genomic DNA which is present in a naturally-occurring population in two or more forms or Alleles.

The feature of Microsatellites that gives rise to polymorphisms is the number of repeats in the Microsatellite.

(4) c. Microsatellites are better than Minisatellites for mapping purposes. Why?

Compared with Minisatellites, Microsatellites:
1) are found distributed more randomly throughout higher eukaryotic genomes.
2) are often found with several different number of repeats; hence, they are found in a naturally-occurring population in several different allelic forms.
3) have a higher Heterozygosity, i.e. different alleles will be found on each of the two chromosomes when the microsatellite is found on an autosome. This higher heterozygosity is due to the presence of several different allelic forms.