10 Genomics, Proteomics, and Genetic Engineering

Key Concepts

- In recombinant DNA (gene cloning), DNA fragments are isolated, inserted into suitable vector molecules, and introduced into host cells (usually bacteria or yeast), where they are replicated.
- Large-scale automated DNA sequencing has resulted in the complete sequence of the genomes of many species of bacteria, archaeons, and eukaryotes including the human genome.
- Functional genomics using DNA microarrays enables the level of gene expression of all genes in the genome to be assayed simultaneously, which allows global patterns and coordinated regulation of gene expression to be investigated.
- Two-hybrid analysis of proteins allows protein-protein interaction networks to be identified.
- Recombinant DNA is widely used in research, medical diagnostics, and the manufacture of drugs and other commercial products.
- Transgenic organisms carry DNA sequences that have been introduced by germ-line transformation or other methods.

Key Terms

1. sticky ends
2. vector
3. library
4. reverse transcriptase
5. cDNA
6. bioinformatics
7. DNA microarray
8. two-hybrid system
9. reporter gene
10. gene targeting
11. transformation rescue
12. transgenic organism

Concepts in Action

10.1. A processed eukaryotic mRNA generally includes a poly-A tail at the 3′ end. Poly(dT) will anneal to this tail and prime DNA synthesis. The method does not work with a prokaryotic mRNA because there is no poly-A tail.

10.2. It occurs 23 times there also, because the sequence is a "palindrome."

10.3. **(a)** The *tet-r* gene is not cleaved with *Bgl* II, so the addition of tetracycline to the medium requires that the colonies be tetracycline-resistant (Tet-r) and hence con-

tain the plasmid. **(b)** Cells with the phenotypes Tet-r Kan-r or Tet-r Kan-s will form colonies. **(c)** Colonies with the phenotype Tet-r Kan-s contain inserts within the cleaved *kan-r* gene.

10.4. One must use cDNA whose transcription is driven by a suitable promoter that functions in a prokaryotic system. Eukaryotic genomic DNA includes regulatory elements that will not work in bacteria and introns that cannot be spliced out of RNA in prokaryotic cells.

10.5. Assuming a random distribution of bases and an equal content of each of the four nucleotide bases, a specific DNA sequence of n nucleotides is expected to occur, on average, every 4^n nucleotides. Hence, a recognition site for an eight-base cutter, such as *Sse*I, would be expected every $4^8 = 65{,}536$ bases, yielding fragments approximately 65.5 kb in length. To separate such long fragments efficiently, you would need to employ either conventional gel electrophoresis using an extremely low agarose concentration (≥ 0.3%, which is very inconvenient to work with) or to separate the fragments with a special electrophoresis apparatus that separates large DNA fragments.

10.6. The primer sequences are included in the PCR product at both ends of the gene fragment. The total length of the amplified piece of DNA is therefore 19 + 303 + 19 = 341 base pairs.

10.7. If the cloning site is inside an antibiotic-resistance gene, then a transformed cell that contains a vector with a DNA fragment inserted at the site will be sensitive to the antibiotic. The second antibiotic-resistance marker is necessary to provide a basis for selection of transformed colonies.

10.8. The 3.1 kb fragment has inverted repeats at the ends; it is probably a DNA transposable element.

10.9. The cDNA reveals the coding sequence in its open reading frame; the genomic sequence includes the introns in the gene as well as upstream and downstream regions that are not included in the primary transcript.

10.10. The E–B fragment contains a repressor-binding site and the H–S fragment contains a repressor-binding site; these act cooperatively to produce full repression.

10.11. *Not*I or any other eight-base cutter would be expected to make a cut on the average every 66 kb ($4^8 = 65{,}536$). Therefore, with the *Arabidopsis* haploid genome at about 100 Mb, one would expect, on average, about 1500 fragments (1 Mb/66 kb = 1515). This calculation assumes equal base ratios and a completely random sequence.

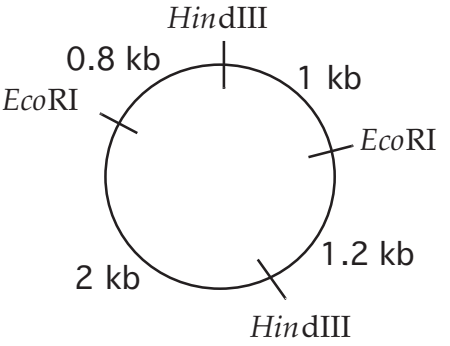
10.12. Six-base cutters, such as *Bam*HI, find their recognition sequences every 4096 (4^6) bases, on the average. In a genome of 4,600,000 bases, about 1123 fragments are expected, assuming that A, G, C, and T are in equal proportions in the genome and are randomly distributed.

10.13. **(a)** The probability of a *Taq*I site is $1/6 \times 1/3 \times 1/3 \times 1/6 = 1/324$, so *Taq*I cleavage is expected every 324 base pairs; the probability of a *Mae*III site is $1/3 \times 1/6 \times 1 \times 1/6 \times 1/3 = 1/324$, so *Mae*III cleavage is expected every 324 base pairs. **(b)** Answers are the same as in part (a), because for both *Taq*I and *Mae*III sites, the number of A + T is equal to the number of G + C.

10.14. The hint says that if the genome were represented x times in the library, the probability that a particular sequence would be missing is e^{-x}, which we want to equal 0.01. Hence, $x = 4.6$. Because one haploid representation of the genome equals $(6 \times 10^9)/2 = 3 \times 10^9$ base pairs, and the average insert size is 2×10^4 base pairs, one requires $([3 \times 10^9]/[2 \times 10^4]) \times 4.6 = 6.9 \times 10^5$ clones.

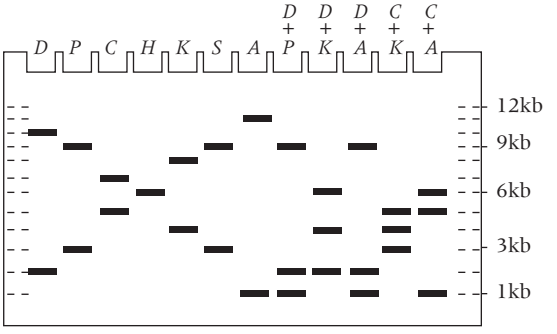
10.15. Although the sticky ends of *Sau*3A and *Bam*HI are compatible, the *Bam*HI restriction site is reconstituted only if the genomic fragment has C next to the *Sau*3A site. The probability of a C at this position is 0.25, assuming equal amounts of four bases and their random distribution in the genome. The probability of recreating *Bam*HI sites at both ends of the cloned fragment (and, therefore, of being able to isolate the cloned fragment from the vector with *Bam* HI) is $0.25^2 = 0.0625$. Hence, only 6.25% of the cloned fragments would be flanked by *Bam*HI sites.

10.16. The only restriction map consistent with the data is a circular one:



10.17. Interpretation in terms of level of expression in the experimental sample relative to the control sample: **(a)** pronounced overexpression; **(b)** pronounced underexpression; **(c)** approximately equal expression; **(d)** moderate overexpression; **(e)** moderate underexpression.

10.18. The resulting electrophoresis gel would have bands in the positions indicated in the accompanying diagram.



10.19. Genes encoding proteins that are used in the synthesis of amino acids, nucleosides and nucleotides, vitamins, and other small molecules that are not present in minimal medium but are present in complete medium.

10.20. The tryptophan operon is expressed equally under all growth conditions. The *lacI* and *crp* genes are expressed constitutively. Although the cyclic AMP receptor protein co-regulates the *lac* operon, it does so through variation in the levels of cAMP under various growth conditions rather than through variation in the expression of *crp*. Transcription of *lacZ* and *lacY* into a polycistronic mRNA is induced by growth in lactose medium.

Experimental minimal medium	Control minimal medium	Transcript				
		trpE	*lacI*	*lacZ*	*lacY*	*crp*
Glucose	Glucose	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ
Glucose	Glycerol	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ
Glycerol	Glucose	Ⓨ	Ⓨ	Ⓨ	Ⓨ	Ⓨ
Lactose	Glucose	Ⓨ	Ⓨ	Ⓡ	Ⓡ	Ⓨ
Glucose	Lactose	Ⓨ	Ⓨ	Ⓖ	Ⓖ	Ⓨ
Lactose	Glycerol	Ⓨ	Ⓨ	Ⓡ	Ⓡ	Ⓨ
Glycerol	Lactose	Ⓨ	Ⓨ	Ⓖ	Ⓖ	Ⓨ

Y = yellow
R = red
G = green

Study Questions

10.S1. Restriction enzyme *Mlu*I recognizes and cleaves specific six-base pairs sequence. Assuming equal and random frequencies of each of the four nucleotides in *E. coli* chromosome, what is the average distance between two *Mlu*I restriction sites? *E. coli* DNA molecule contains 4.6×10^6 base pairs

10.S2. Restriction enzyme *Bal*I cleaves both DNA strands at the center of symmetry of the recognition site 5′-TGGCCA-3′. What kind of ends does it produce?

10.S3. The ability of an introduced DNA fragment to correct a genetic defect in a mutant organism is called a(n) _____.

10.S4. In gene-targeting experiments, the replacement of the wildtype gene in the genome with the completely _____ gene results in a knockout mutation.

10.S5. 5′-TTT___-3′ is a half of a palindromic restriction site for *Dra*I. What is the complete sequence?

10.S6. A _____ is a DNA sequence, present once per haploid genome, that can be amplified with a suitable pair of oligonucleotide primers by means of the polymerase chain reaction.

10.S7. The restriction site of *Nde*I is 5′-CATATG-3′. Is 3′-CATATG-5′ also a *Nde*I restriction site? Why or why not?

10.S8. The genomic DNA of an organism has a base composition of 30 percent A–T base pairs and 70 percent G–C base pairs. Assuming a random sequence of bases, what is the expected frequency of the following restriction sites:

a. *Aat*II 5′-GACGTC-3′

b. *Sfo*I 5′-GGCGCC-3′

c. *Eco*NI 5′-CCTNNNNNAGG-3′

10.S9. Restriction enzyme *ApaL*I cleaves site 5′-G ↓ TGCAC-3′ and *Sfc*I cleaves 5′-C ↓ TGCAG-3′. Is it possible to ligate two DNA sequences, one digested with *Apa*LI and another with *Sfc*I?

10.S10. Restriction enzymes *Age*I (5′-A ↓ CCGGT-3′) and *Ava*I (5′-C ↓ CCGGG-3′) produce compatible "sticky" ends. Will these enzymes be able to recleave a DNA fragment that resulted from ligation of two sequences, one digested with *Age*I and another with *Ava*I?