


GENOMES 5

T. A. Brown

 **CRC Press**
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **Informa** business

 **Garland Science**
Taylor & Francis Group

CONTENTS

PREFACE

ACKNOWLEDGEMENTS

PART 1 HOW GENOMES ARE STUDIED

CHAPTER 1 GENOMES, TRANSCRIPTOMES, AND PROTEOMES

1.1 DNA

Genes are made of DNA

DNA is a polymer of nucleotides

The discovery of the double helix

The double helix is stabilized by base-pairing and base-stacking

The double helix has structural flexibility

1.2 RNA AND THE TRANSCRIPTOME

RNA is a second type of polynucleotide

The RNA content of the cell

Many RNAs are synthesized as precursor molecules

There are different definitions of the transcriptome

1.3 PROTEINS AND THE PROTEOME

There are four hierarchical levels of protein structure

Amino acid diversity underlies protein diversity

The link between the transcriptome and the proteome

The genetic code is not universal

The link between the proteome and the biochemistry of the cell

SUMMARY

SHORT ANSWER QUESTIONS

IN-DEPTH PROBLEMS

FURTHER READING

CHAPTER 2 STUDYING DNA

2.1 ENZYMES FOR DNA MANIPULATION

XIII
XVII

The mode of action of a template-dependent DNA polymerase 26

The types of DNA polymerase used in research 28

Restriction endonucleases enable DNA molecules to be cut at defined positions 29

Gel electrophoresis is used to examine the results of a restriction digest 32

Interesting DNA fragments can be identified by Southern hybridization 33

Ligases join DNA fragments together 34

End-modification enzymes 35

1 **2.2 THE POLYMERASE CHAIN REACTION** 35
Carrying out a PCR 36

2 The rate of product formation can be followed during a PCR 37

3 PCR has many and diverse applications 38

5 **2.3 DNA CLONING** 38
Why is gene cloning important? 39

6 The simplest cloning vectors are based on *E. coli* plasmids 39

8 Bacteriophages can also be used as cloning vectors 41

9 Vectors for longer pieces of DNA 44

11 DNA can be cloned in organisms other than *E. coli* 45

11 **SUMMARY** 47

16 **SHORT ANSWER QUESTIONS** 48

16 **IN-DEPTH PROBLEMS** 48

16 **FURTHER READING** 49

18
19 **CHAPTER 3**
20 **MAPPING GENOMES** 51

22 **3.1 WHY A GENOME MAP IS IMPORTANT** 51

Genome maps are needed in order to sequence the more complex genomes 51

Genome maps are not just sequencing aids 52

23 **3.2 MARKERS FOR GENETIC MAPPING** 53

Genes were the first markers to be used 54

RFLPs and SSLPs are examples of DNA markers 55

Single-nucleotide polymorphisms are the most useful type of DNA marker 57

25 **3.3 THE BASIS TO GENETIC MAPPING** 59

26 The principles of inheritance and the discovery of linkage 59

Partial linkage is explained by the behavior of chromosomes during meiosis	60	Many prokaryotic genomes have been sequenced by the shotgun method	95
From partial linkage to genetic mapping	63	Shotgun sequencing of eukaryotic genomes requires sophisticated assembly programs	95
3.4 LINKAGE ANALYSIS WITH DIFFERENT TYPES OF ORGANISM		From contigs to scaffolds	97
Linkage analysis when planned breeding experiments are possible	64	64 What is a 'genome sequence' and do we always need one?	99
Gene mapping by human pedigree analysis	66	4.3 SEQUENCING THE HUMAN GENOME	101
Genetic mapping in bacteria	67	The Human Genome Project – genome sequencing in the heroic age	102
The limitations of linkage analysis	69	The human genome – genome sequencing in the modern age	104
3.5 PHYSICAL MAPPING BY DIRECT EXAMINATION OF DNA MOLECULES		70 The Neanderthal genome – assembly of an extinct genome using the human sequence as a reference	106
Conventional restriction mapping is only applicable to small DNA molecules	71	The human genome – new challenges	107
Optical mapping can locate restriction sites in longer DNA molecules	71	SUMMARY	108
Optical mapping with fluorescent probes	74	SHORT ANSWER QUESTIONS	109
Further innovations extend the scope of optical mapping	75	IN-DEPTH PROBLEMS	110
3.6 PHYSICAL MAPPING BY ASSIGNING MARKERS TO DNA FRAGMENTS		FURTHER READING	110
Any unique sequence can be used as an STS	77	CHAPTER 5	
DNA fragments for STS mapping can be obtained as radiation hybrids	77	GENOME ANNOTATION	113
A clone library can be used as the mapping reagent	78	5.1 GENOME ANNOTATION BY COMPUTER ANALYSIS OF THE DNA SEQUENCE	113
SUMMARY	80	The coding regions of genes are open reading frames	113
SHORT ANSWER QUESTIONS	80	Simple ORF scans are less effective with genomes of higher eukaryotes	114
IN-DEPTH PROBLEMS	81	Locating genes for noncoding RNA	116
FURTHER READING	81	Homology searches and comparative genomics give an extra dimension to gene prediction	117
CHAPTER 4		5.2 GENOME ANNOTATION BY ANALYSIS OF GENE TRANSCRIPTS	119
SEQUENCING GENOMES	83	Hybridization tests can determine if a fragment contains one or more genes	119
4.1 METHODOLOGY FOR DNA SEQUENCING	83	Methods are available for precise mapping of the ends of transcripts	120
Chain-termination sequencing of PCR products	83	Exon–intron boundaries can also be located with precision	121
Illumina sequencing is the most popular short-read method	86	5.3 ANNOTATION BY GENOME-WIDE RNA MAPPING	121
A variety of other short-read sequencing methods have been devised	88	Tiling arrays enable transcripts to be mapped on to chromosomes or entire genomes	122
Single-molecule real-time sequencing provides reads up to 200 kb in length	90	Transcript sequences can be directly mapped onto a genome	123
Nanopore sequencing is currently the longest long-read method	92	Obtaining transcript sequences by SAGE and CAGE	125
4.2 HOW TO SEQUENCE A GENOME	93		
The potential of the shotgun method was proven by the <i>Haemophilus influenzae</i> sequence	93		

5.4 GENOME BROWSERS	126	IN-DEPTH PROBLEMS	151
SUMMARY	128	FURTHER READING	152
SHORT ANSWER QUESTIONS	128		
IN-DEPTH PROBLEMS	129	PART 2 GENOME ANATOMIES	
FURTHER READING	129		
CHAPTER 6		CHAPTER 7	
IDENTIFYING GENE FUNCTIONS	131	EUKARYOTIC NUCLEAR GENOMES	153
6.1 COMPUTER ANALYSIS OF GENE FUNCTION	131	7.1 NUCLEAR GENOMES ARE CONTAINED IN CHROMOSOMES	153
Homology reflects evolutionary relationships	131	Chromosomes are made of DNA and protein	153
Homology analysis can provide information on the function of a gene	132	The special features of metaphase chromosomes	155
Identification of protein domains can help to assign function to an unknown gene	133	Centromeres and telomeres have distinctive DNA sequences	157
Annotation of gene function requires a common terminology	134		
6.2 ASSIGNING FUNCTION BY GENE INACTIVATION AND OVEREXPRESSION	135	7.2 THE GENETIC FEATURES OF NUCLEAR GENOMES	158
Functional analysis by gene inactivation	136	Gene numbers can be misleading	158
Gene inactivation by genome editing	136	Genes are not evenly distributed within a genome	160
Gene inactivation by homologous recombination	137	A segment of the human genome	161
Gene inactivation by transposon tagging and RNA interference	138	The yeast genome is very compact	163
Gene overexpression can also be used to assess function	139	Gene organization in other eukaryotes	165
The phenotypic effect of gene inactivation or overexpression may be difficult to discern	140	Families of genes	166
		Pseudogenes and other evolutionary relics	167
6.3 UNDERSTANDING GENE FUNCTION BY STUDIES OF ITS EXPRESSION PATTERN AND PROTEIN PRODUCT	142	7.3 THE REPETITIVE DNA CONTENT OF EUKARYOTIC NUCLEAR GENOMES	169
Reporter genes and immunocytochemistry can be used to locate where and when genes are expressed	142	Tandemly repeated DNA is found at centromeres and elsewhere in eukaryotic chromosomes	169
CRISPR can be used to make specific changes in a gene and the protein it encodes	143	Minisatellites and microsatellites	170
Other methods for site-directed mutagenesis	145	Interspersed repeats	171
		SUMMARY	171
6.4 USING CONVENTIONAL GENETIC ANALYSIS TO IDENTIFY GENE FUNCTION	147	SHORT ANSWER QUESTIONS	172
Identification of human genes responsible for inherited diseases	147	IN-DEPTH PROBLEMS	173
Genome-wide association studies can also identify genes for diseases and other traits	149	FURTHER READING	173
SUMMARY	150	CHAPTER 8	
SHORT ANSWER QUESTIONS	151	GENOMES OF PROKARYOTES AND EUKARYOTIC ORGANELLES	175
		8.1 THE PHYSICAL FEATURES OF PROKARYOTIC GENOMES	175
		The traditional view of the prokaryotic chromosome	175
		Some bacteria have linear or multipartite genomes	177
		8.2 THE GENETIC FEATURES OF PROKARYOTIC GENOMES	180

Gene organization in the <i>E. coli</i> K12 genome	180	PART 3 HOW GENOMES ARE EXPRESSED	
Operons are characteristic features of prokaryotic genomes	182		
Prokaryotic genome sizes and gene numbers vary according to biological complexity	184		
Genome sizes and gene numbers vary within individual species	185		
Distinctions between prokaryotic species are further blurred by horizontal gene transfer	186		
Metagenomes describe the members of a community	188	CHAPTER 10	
8.3 EUKARYOTIC ORGANELLE GENOMES	189	ACCESSING THE GENOME	215
The endosymbiont theory explains the origin of organelle genomes	190	10.1 INSIDE THE NUCLEUS	215
The physical and genetic features of organelle genomes	191	The nucleus has an ordered internal structure	216
SUMMARY	195	Chromosomal DNA displays different degrees of packaging	217
SHORT ANSWER QUESTIONS	195	The nuclear matrix is a dynamic structure	218
IN-DEPTH PROBLEMS	196	Each chromosome has its own territory within the nucleus	220
FURTHER READING	196	Chromosomal DNA is organized into topologically associating domains	221
CHAPTER 9		Insulators prevent crosstalk between segments of chromosomal DNA	223
VIRUS GENOMES AND MOBILE GENETIC ELEMENTS	199	10.2 NUCLEOSOME MODIFICATIONS AND GENOME EXPRESSION	224
9.1 THE GENOMES OF BACTERIOPHAGES AND EUKARYOTIC VIRUSES	199	Acetylation of histones influences many nuclear activities, including genome expression	225
Bacteriophage genomes have diverse structures and organizations	199	Histone deacetylation represses active regions of the genome	226
Replication strategies for bacteriophage genomes	201	Acetylation is not the only type of histone modification	227
Structures and replication strategies for eukaryotic viral genomes	202	Nucleosome repositioning also influences gene expression	230
Some retroviruses cause cancer	204	10.3 DNA MODIFICATION AND GENOME EXPRESSION	231
Genomes at the edge of life	205	Genome silencing by DNA methylation	231
9.2 MOBILE GENETIC ELEMENTS	206	Methylation is involved in genomic imprinting and X inactivation	232
RNA transposons with long terminal repeats are related to viral retroelements	206	SUMMARY	234
Some RNA transposons lack LTRs	208	SHORT ANSWER QUESTIONS	235
DNA transposons are common in prokaryotic genomes	209	IN-DEPTH PROBLEMS	235
DNA transposons are less common in eukaryotic genomes	211	FURTHER READING	236
SUMMARY	212	CHAPTER 11	
SHORT ANSWER QUESTIONS	213	THE ROLE OF DNA-BINDING PROTEINS IN GENOME EXPRESSION	239
IN-DEPTH PROBLEMS	213	11.1 METHODS FOR STUDYING DNA-BINDING PROTEINS AND THEIR ATTACHMENT SITES	239
FURTHER READING	214	X-ray crystallography provides structural data for any protein that can be crystallized	239

NMR spectroscopy is used to study the structures of small proteins	240	Synthesis of bacterial RNA is regulated by repressor and activator proteins	273
Gel retardation identifies DNA fragments that bind to proteins	241	Synthesis of bacterial RNA is also regulated by control over transcription termination	276
Protection assays pinpoint binding sites with greater accuracy	242	Synthesis of eukaryotic RNA is regulated primarily by activator proteins	277
Modification interference identifies nucleotides central to protein binding	244		
Genome-wide scans for protein attachment sites	245	12.4 THE INFLUENCE OF RNA SPLICING ON THE COMPOSITION OF A TRANSCRIPTOME	280
11.2 THE SPECIAL FEATURES OF DNA-BINDING PROTEINS	245	The splicing pathway for eukaryotic pre-mRNA introns	281
The helix–turn–helix motif is present in prokaryotic and eukaryotic proteins	246	The splicing process must have a high degree of precision	282
Zinc fingers are common in eukaryotic proteins	248	Enhancer and silencer elements specify alternative splicing pathways	284
Other nucleic acid-binding motifs	248	Backsplicing gives rise to circular RNAs	286
11.3 THE INTERACTION BETWEEN DNA AND ITS BINDING PROTEINS	249	12.5 THE INFLUENCE OF CHEMICAL MODIFICATION ON THE COMPOSITION OF A TRANSCRIPTOME	287
Contacts between DNA and proteins	250	RNA editing alters the coding properties of some transcripts	287
Direct readout of the nucleotide sequence	250	Chemical modifications that do not affect the sequence of an mRNA	289
The conformation of the helix also influences protein binding	251	12.6 DEGRADATION OF THE COMPONENTS OF THE TRANSCRIPTOME	290
SUMMARY	252	Several processes are known for nonspecific RNA turnover	291
SHORT ANSWER QUESTIONS	253	RNA silencing was first identified as a means of destroying invading viral RNA	292
IN-DEPTH PROBLEMS	253	MicroRNAs regulate genome expression by causing specific target mRNAs to be degraded	293
FURTHER READING	254	SUMMARY	294
CHAPTER 12	257	SHORT ANSWER QUESTIONS	295
TRANSCRIPTOMES		IN-DEPTH PROBLEMS	295
12.1 THE COMPONENTS OF THE TRANSCRIPTOME	257	FURTHER READING	296
The mRNA fraction of a transcriptome is small but complex	257		
Short noncoding RNAs have diverse functions	258	CHAPTER 13	299
Long noncoding RNAs are enigmatic transcripts	260	PROTEOMES	
12.2 TRANSCRIPTOMICS: CATALOGING THE TRANSCRIPTOMES OF CELLS AND TISSUES	262	13.1 STUDYING THE COMPOSITION OF A PROTEOME	299
Microarray analysis and RNA sequencing are used to study the contents of transcriptomes	262	The separation stage of a protein profiling project	300
Single-cell studies add greater precision to transcriptomics	264	The identification stage of a protein profiling project	303
Spatial transcriptomics enables transcripts to be mapped directly in tissues and cells	266	Comparing the compositions of two proteomes	305
12.3 SYNTHESIS OF THE COMPONENTS OF THE TRANSCRIPTOME	268	Analytical protein arrays offer an alternative approach to protein profiling	306
RNA polymerases are molecular machines for making RNA	268	13.2 IDENTIFYING PROTEINS THAT INTERACT WITH ONE ANOTHER	307
Transcription start-points are indicated by promoter sequences	270	Identifying pairs of interacting proteins	307
		Identifying the components of multiprotein complexes	309

Identifying proteins with functional interactions	311		
Protein interaction maps display the interactions within a proteome	311		
13.3 SYNTHESIS AND DEGRADATION OF THE COMPONENTS OF THE PROTEOME	313		
Ribosomes are molecular machines for making proteins	313		
During stress, bacteria inactivate their ribosomes in order to downsize the proteome	316		
Initiation factors mediate large-scale remodeling of eukaryotic proteomes	317		
The translation of individual mRNAs can also be regulated	318		
Degradation of the components of the proteome	320		
13.4 THE INFLUENCE OF PROTEIN PROCESSING ON THE COMPOSITION OF THE PROTEOME	320		
The amino acid sequence contains instructions for protein folding	321		
Some proteins undergo proteolytic cleavage	324		
Important changes in protein activity can be brought about by chemical modification	325		
13.5 BEYOND THE PROTEOME	326		
The metabolome is the complete set of metabolites present in a cell	327		
Systems biology provides an integrated description of cellular activity	327		
SUMMARY	330		
SHORT ANSWER QUESTIONS	331		
IN-DEPTH PROBLEMS	332		
FURTHER READING	332		
CHAPTER 14	335		
GENOME EXPRESSION IN THE CONTEXT OF CELL AND ORGANISM			
14.1 THE RESPONSE OF THE GENOME TO EXTERNAL SIGNALS	335		
Signal transmission by import of the extracellular signaling compound	336		
Receptor proteins transmit signals across cell membranes	337		
Some signal transduction pathways have few steps between receptor and genome	339		
Some signal transduction pathways have many steps between receptor and genome	340		
Some signal transduction pathways operate via second messengers	341		
14.2 CHANGES IN GENOME ACTIVITY RESULTING IN CELLULAR DIFFERENTIATION	341		
Some differentiation processes involve changes to chromatin structure	341		
Yeast mating types are determined by gene conversion events	343		
Genome rearrangements are responsible for immunoglobulin and T-cell receptor diversities	344		
14.3 CHANGES IN GENOME ACTIVITY UNDERLYING DEVELOPMENT	346		
Bacteriophage λ : a genetic switch enables a choice to be made between alternative developmental pathways	347		
<i>Bacillus</i> sporulation: coordination of activities in two distinct cell types	348		
<i>Caenorhabditis elegans</i> : the genetic basis to positional information and the determination of cell fate	351		
Fruit flies: conversion of positional information into a segmented body plan	353		
Homeotic selector genes are universal features of higher eukaryotic development	354		
Homeotic genes also underlie plant development	356		
SUMMARY	357		
SHORT ANSWER QUESTIONS	358		
IN-DEPTH PROBLEMS	358		
FURTHER READING	359		
PART 4 HOW GENOMES REPLICATE AND EVOLVE			
CHAPTER 15	361		
GENOME REPLICATION			
15.1 THE TOPOLOGY OF GENOME REPLICATION	361		
The double-helix structure complicates the replication process	362		
The Meselson–Stahl experiment proved that replication is semiconservative	363		
DNA topoisomerases provide a solution to the topological problem	365		
Variations on the semiconservative theme	367		
15.2 THE INITIATION PHASE OF GENOME REPLICATION	368		
Initiation at the <i>E. coli</i> origin of replication	368		

Origins of replication have been clearly defined in yeast	369	Homologous recombination pathways in eukaryotes	399
Origins in higher eukaryotes have been less easy to identify	370	16.2 SITE-SPECIFIC RECOMBINATION	400
15.3 EVENTS AT THE REPLICATION FORK	371	Bacteriophage λ uses site-specific recombination during the lysogenic infection cycle	400
DNA polymerases are molecular machines for making (and degrading) DNA	371	Site-specific recombination is an aid in construction of genetically modified plants	401
DNA polymerases have limitations that complicate genome replication	373	16.3 TRANSPOSITION	402
Okazaki fragments must be joined together to complete lagging-strand replication	374	Replicative and conservative transposition of DNA transposons	402
15.4 TERMINATION OF GENOME REPLICATION	376	Retroelements transpose replicatively via an RNA intermediate	403
Replication of the <i>E. coli</i> genome terminates within a defined region	376	SUMMARY	405
Completion of genome replication	378	SHORT ANSWER QUESTIONS	406
Telomerase completes replication of chromosomal DNA molecules, at least in some cells	380	IN-DEPTH PROBLEMS	406
Telomere length is implicated in cell senescence and cancer	382	FURTHER READING	406
<i>Drosophila</i> has a unique solution to the end-shortening problem	383	CHAPTER 17	
15.5 REGULATION OF EUKARYOTIC GENOME REPLICATION	384	MUTATIONS AND DNA REPAIR	409
Genome replication must be synchronized with the cell cycle	384	17.1 THE CAUSES OF MUTATIONS	409
Origin licensing is the prerequisite for passing the G1-S checkpoint	385	Errors in replication are a source of point mutations	410
Replication origins do not all fire at the same time	386	Replication errors can also lead to insertion and deletion mutations	411
The cell has various options if the genome is damaged	388	Mutations are also caused by chemical and physical mutagens	413
SUMMARY	388	17.2 REPAIR OF MUTATIONS AND OTHER TYPES OF DNA DAMAGE	418
SHORT ANSWER QUESTIONS	389	Direct repair systems fill in nicks and correct some types of nucleotide modification	418
IN-DEPTH PROBLEMS	390	Base excision repairs many types of damaged nucleotide	419
FURTHER READING	390	Nucleotide excision repair is used to correct more extensive types of damage	421
CHAPTER 16		Mismatch repair corrects replication errors	422
RECOMBINATION AND		Single- and double-strand breaks can be repaired	423
TRANSPOSITION	393	Some types of damage can be repaired by homologous recombination	425
16.1 HOMOLOGOUS RECOMBINATION	393	If necessary, DNA damage can be bypassed during genome replication	426
The Holliday and Meselson–Radding models for homologous recombination	394	Defects in DNA repair underlie human diseases, including cancers	427
The double-strand break model for homologous recombination	396	SUMMARY	427
RecBCD is the most important pathway for homologous recombination in bacteria	397	SHORT ANSWER QUESTIONS	428
<i>E. coli</i> has alternative pathways for homologous recombination	398	IN-DEPTH PROBLEMS	429
		FURTHER READING	429

CHAPTER 18			
HOW GENOMES EVOLVE	431	18.3 GENOMES: THE LAST 6 MILLION YEARS	451
8.1 GENOMES: THE FIRST 10 BILLION YEARS	431	The human genome is very similar to that of the chimpanzee	451
The first biochemical systems were centered on RNA	431	Paleogenomics is helping us understand the recent evolution of the human genome	453
The first DNA genomes	433	18.4 GENOMES TODAY: DIVERSITY IN POPULATIONS	455
How unique is life?	434	The origins of HIV/AIDS	455
8.2 THE EVOLUTION OF INCREASINGLY COMPLEX GENOMES	436	The first migrations of humans out of Africa	457
Genome sequences provide extensive evidence of vast gene duplications	436	The diversity of plant genomes is an aid in crop breeding	459
A variety of processes could result in gene duplication	439	SUMMARY	460
Whole-genome duplication is also possible	440	SHORT ANSWER QUESTIONS	462
Smaller duplications can also be identified in the human genome and other genomes	443	IN-DEPTH PROBLEMS	462
Both prokaryotes and eukaryotes acquire genes from other species	444	FURTHER READING	463
Genome evolution also involves rearrangement of existing gene sequences	446	GLOSSARY	465
There are competing hypotheses for the origins of introns	448	INDEX	509
The evolution of the epigenome	450		

CHAPTER 18			
HOW GENOMES EVOLVE	431		
18.1 GENOMES: THE FIRST 10 BILLION YEARS	431		
The first biochemical systems were centered on RNA	431		
The first DNA genomes	433		
How unique is life?	434		
18.2 THE EVOLUTION OF INCREASINGLY COMPLEX GENOMES	436		
Genome sequences provide extensive evidence of past gene duplications	436		
A variety of processes could result in gene duplication	439		
Whole-genome duplication is also possible	440		
Smaller duplications can also be identified in the human genome and other genomes	443		
Both prokaryotes and eukaryotes acquire genes from other species	444		
Genome evolution also involves rearrangement of existing gene sequences	446		
There are competing hypotheses for the origins of introns	448		
The evolution of the epigenome	450		
		18.3 GENOMES: THE LAST 6 MILLION YEARS	451
		The human genome is very similar to that of the chimpanzee	451
		<i>Paleogenomics is helping us understand the recent evolution of the human genome</i>	453
		18.4 GENOMES TODAY: DIVERSITY IN POPULATIONS	455
		The origins of HIV/AIDS	455
		The first migrations of humans out of Africa	457
		The diversity of plant genomes is an aid in crop breeding	459
		SUMMARY	460
		SHORT ANSWER QUESTIONS	462
		IN-DEPTH PROBLEMS	462
		FURTHER READING	463
		GLOSSARY	465
		INDEX	509