GENOMES 4

T. A. BROWN

Garland Science
Taylor & Francis Group
NEW YORK AND LONDON

CONTENTS

CHAPTER 1 GENOMES, TRANSCRIPTOMES, AND PROTEOMES 1

1.1 DNA 2
Genes are made of DNA 3
DNA is a polymer of nucleotides 4
The double helix is stabilized by base pairing and base stacking 8
The double helix has structural flexibility 9

1.2 RNA AND THE TRANSCRIPTOME 11
RNA is a second type of polynucleotide 12
The RNA content of the cell 12
Many RNAs are synthesized as precursor molecules 13
There are different definitions of the transcriptome 15

1.3 PROTEINS AND THE PROTEOME 16
There are four hierarchical levels of protein structure 16
Amino acid diversity underlies protein diversity 17
The link between the transcriptome and the proteome 19
The genetic code is not universal 20
The link between the proteome and the biochemistry of the cell 22

SUMMARY 23

SHORT ANSWER QUESTIONS 24

IN-DEPTH PROBLEMS 24

FURTHER READING 25

CHAPTER 2 STUDYING DNA 27

2.1 ENZYMES FOR DNA MANIPULATION 28
The mode of action of a template-dependent DNA polymerase 28
The types of DNA polymerase used in research 30
Restriction endonucleases enable DNA molecules to be cut at defined positions 32
Gel electrophoresis is used to examine the results of a restriction digest 34
Interesting DNA fragments can be identified by Southern hybridization 35
Ligases join DNA fragments together 37
End-modification enzymes 38

2.2 THE POLYMERASE CHAIN REACTION 38
Carrying out a PCR 39
The rate of product formation can be followed during a PCR 40
PCR has many and diverse applications 41

2.3 DNA CLONING 41
Why is gene cloning important? 41
The simplest cloning vectors are based on *E. coli* plasmids 43
Bacteriophages can also be used as cloning vectors 44
Vectors for longer pieces of DNA 47
DNA can be cloned in organisms other than *E. coli* 48

SUMMARY 50

SHORT ANSWER QUESTIONS 51

IN-DEPTH PROBLEMS 51

FURTHER READING 52

CHAPTER 3 MAPPING GENOMES 55

3.1 WHY A GENOME MAP IS IMPORTANT 55
Genome maps are needed in order to sequence the more complex genomes 55
Genome maps are not just sequencing aids 57

3.2 MARKERS FOR GENETIC MAPPING 58
Genes were the first markers to be used 58
RFLPs and SSLPs are examples of DNA markers 59
Single-nucleotide polymorphisms are the most useful type of DNA marker 61

3.3 THE BASIS TO GENETIC MAPPING 63
The principles of inheritance and the discovery of linkage 63
Partial linkage is explained by the behavior of chromosomes during meiosis 65
From partial linkage to genetic mapping 68

3.4 LINKAGE ANALYSIS WITH DIFFERENT TYPES OF ORGANISMS 69
Linkage analysis when planned breeding experiments are possible 69
Gene mapping by human pedigree analysis 71
Genetic mapping in bacteria 73
The limitations of linkage analysis 74

3.5 PHYSICAL MAPPING BY DIRECT EXAMINATION OF DNA MOLECULES 75
Conventional restriction mapping is applicable only to small DNA molecules 75
Optical mapping can locate restriction sites in longer DNA molecules 77
Optical mapping can be used to map other features in a DNA molecule 79

3.6 PHYSICAL MAPPING BY ASSIGNING MARKERS TO DNA FRAGMENTS 81
Any unique sequence can be used as an STS 81
DNA fragments for STS mapping can be obtained as radiation hybrids 82
A clone library can be used as the mapping reagent 83

SUMMARY 84

SHORT ANSWER QUESTIONS 85

IN-DEPTH PROBLEMS 85

FURTHER READING 86

CHAPTER 4 SEQUENCING GENOMES 87

4.1 CHAIN-TERMINATION SEQUENCING 87
Chain-termination sequencing in outline 87
Not all DNA polymerases can be used for sequencing 89
Chain-termination sequencing with *Taq* polymerase 90
Strengths and limitations of chain-termination sequencing 91

4.2 NEXT-GENERATION SEQUENCING 92
Preparation of a sequencing library is the common feature of next-generation methods 93
Various next-generation sequencing methods have been devised 95
Third- and fourth-generation methods enable sequencing in real time 97

4.3 HOW TO SEQUENCE A GENOME 98
The potential of the shotgun method was proven by the *Haemophilus influenzae* sequence 99
Many prokaryotic genomes have been sequenced by the shotgun method 100
Shotgun sequencing of eukaryotic genomes requires sophisticated assembly programs 102
More complex genomes can be sequenced by a hierarchical shotgun approach 104
What is a genome sequence and do we always need one? 107

4.4 A SURVEY OF EUKARYOTIC GENOME SEQUENCING PROJECTS 109
The Human Genome Project: genome sequencing in the heroic age 109
The Neanderthal genome: assembly of an extinct genome by use of the human sequence as a reference 110
The giant panda genome: shotgun sequencing based entirely on next-generation data 111
The barley genome: the concept of gene space 113

SUMMARY 115

SHORT ANSWER QUESTIONS 115

IN-DEPTH PROBLEMS 116

FURTHER READING 117

CHAPTER 5 GENOME ANNOTATION 119

5.1 GENOME ANNOTATION BY COMPUTER ANALYSIS OF THE DNA SEQUENCE 119
The coding regions of genes are open reading frames 119
Simple ORF scans are less effective with genomes of higher eukaryotes 120
Locating genes for noncoding RNA 122
Homology searches and comparative genomics give an extra dimension to gene prediction 123

5.2 GENOME ANNOTATION BY ANALYSIS OF GENE TRANSCRIPTS 124
Hybridization tests can determine if a fragment contains transcribed sequences 125
Methods are available for precise mapping of the ends of transcripts 126
Exon–intron boundaries can also be located with precision 126

5.3 ANNOTATION BY GENOMEWIDE RNA MAPPING 127
Tiling arrays enable transcripts to be mapped onto chromosomes or entire genomes 128
Transcript sequences can be directly mapped onto a genome 129

5.4 GENOME BROWSERS 131

SUMMARY 132
SHORT ANSWER QUESTIONS 132
IN-DEPTH PROBLEMS 133
FURTHER READING 133

CHAPTER 6 IDENTIFYING GENE FUNCTIONS 135

6.1 COMPUTER ANALYSIS OF GENE FUNCTION 135
Homology reflects evolutionary relationships 135
Homology analysis can provide information on the function of a gene 136
Identification of protein domains can help to assign function to an unknown gene 137
Annotation of gene function requires a common terminology 138

6.2 ASSIGNING FUNCTION BY GENE INACTIVATION AND OVEREXPRESSION 139
Functional analysis by gene inactivation 140
Individual genes can be inactivated by homologous recombination 140
Gene inactivation without homologous recombination 142
Gene overexpression can also be used to assess function 144
The phenotypic effect of gene inactivation or overexpression may be difficult to discern 145

6.3 UNDERSTANDING GENE FUNCTION BY STUDIES OF EXPRESSION PATTERN AND PROTEIN PRODUCT 146
Reporter genes and immunocytochemistry can be used to locate where and when genes are expressed 146
Directed mutagenesis can be used to probe gene function in detail 147

6.4 USING CONVENTIONAL GENETIC ANALYSIS TO IDENTIFY GENE FUNCTION 149
Identification of human genes responsible for inherited diseases 150
Genomewide association studies can also identify genes for diseases and other traits 151

SUMMARY 152
SHORT ANSWER QUESTIONS 153
IN-DEPTH PROBLEMS 153
FURTHER READING 154

CHAPTER 7 EUKARYOTIC NUCLEAR GENOMES 155

7.1 NUCLEAR GENOMES ARE CONTAINED IN CHROMOSOMES 155
Chromosomes are much shorter than the DNA molecules they contain 155
Special features of metaphase chromosomes 157
DNA–protein interactions in centromeres and telomeres 159

7.2 HOW ARE THE GENES ARRANGED IN A NUCLEAR GENOME? 161
Genes are not evenly distributed within a genome 161
A segment of the human genome 162
The yeast genome is very compact 164
Gene organization in other eukaryotes 165

7.3 HOW MANY GENES ARE THERE AND WHAT ARE THEIR FUNCTIONS? 167
Gene numbers can be misleading 168
Gene catalogs reveal the distinctive features of different organisms 169
Families of genes 172
Pseudogenes and other evolutionary relics 174

7.4 THE REPETITIVE DNA CONTENT OF EUKARYOTIC NUCLEAR GENOMES 176
Tandemly repeated DNA is found at centromeres and elsewhere in eukaryotic chromosomes 176
Minisatellites and microsatellites 176
Interspersed repeats 177

SUMMARY 178
SHORT ANSWER QUESTIONS 178
IN-DEPTH PROBLEMS 179
FURTHER READING 179

CHAPTER 8 GENOMES OF PROKARYOTES AND EUKARYOTIC ORGANELLES 181

8.1 PHYSICAL FEATURES OF PROKARYOTIC GENOMES 181
The traditional view of the prokaryotic chromosome 181
Some bacteria have linear or multipartite genomes 183

8.2 GENETIC FEATURES OF PROKARYOTIC GENOMES 186
Gene organization in the *E. coli* K12 genome 186

Operons are characteristic features of prokaryotic genomes 188
Prokaryotic genome sizes and numbers of genes vary according to biological complexity 189
Genome sizes and numbers of genes vary within individual species 190
Distinctions between prokaryotic species are further blurred by lateral gene transfer 192
Metagenomes describe the members of a community 194

8.3 EUKARYOTIC ORGANELLAR GENOMES 195
The endosymbiont theory explains the origin of organellar genomes 195
Most organellar genomes are circular 196
The gene catalogs of organellar genomes 197

SUMMARY 198

SHORT ANSWER QUESTIONS 200

IN-DEPTH PROBLEMS 201

FURTHER READING 201

CHAPTER 9 VIRAL GENOMES AND MOBILE GENETIC ELEMENTS 203

9.1 THE GENOMES OF BACTERIOPHAGES AND EUKARYOTIC VIRUSES 203
Bacteriophage genomes have diverse structures and organizations 203
Replication strategies for bacteriophage genomes 205
Structures and replication strategies for eukaryotic viral genomes 206
Some retroviruses cause cancer 207
Genomes at the edge of life 209

9.2 MOBILE GENETIC ELEMENTS 210
RNA transposons with long terminal repeats are related to viral retroelements 210
Some RNA transposons lack long terminal repeats 212
DNA transposons are common in prokaryotic genomes 213
DNA transposons are less common in eukaryotic genomes 214

SUMMARY 216

SHORT ANSWER QUESTIONS 216

IN-DEPTH PROBLEMS 217

FURTHER READING 217

CHAPTER 10 ACCESSING THE GENOME 219

10.1 INSIDE THE NUCLEUS 219
The nucleus has an ordered internal structure 220
The DNA content of a nondividing nucleus displays different degrees of packaging 221
The nuclear matrix is thought to provide attachment points for chromosomal DNA 222
Each chromosome has its own territory within the nucleus 223
Each chromosome comprises a series of topologically associated domains 224
Insulators mark the boundaries of topologically associated domains 226

10.2 NUCLEOSOME MODIFICATIONS AND GENOME EXPRESSION 228
Acetylation of histones influences many nuclear activities including genome expression 228
Histone deacetylation represses active regions of the genome 229
Acetylation is not the only type of histone modification 230
Nucleosome repositioning also influences gene expression 231

10.3 DNA MODIFICATION AND GENOME EXPRESSION 234
Genome silencing by DNA methylation 234
Methylation is involved in genomic imprinting and X inactivation 235

SUMMARY 236

SHORT ANSWER QUESTIONS 237

IN-DEPTH PROBLEMS 238

FURTHER READING 238

CHAPTER 11 THE ROLE OF DNA-BINDING PROTEINS IN GENOME EXPRESSION 241

11.1 METHODS FOR STUDYING DNA-BINDING PROTEINS AND THEIR ATTACHMENT SITES 241
X-ray crystallography provides structural data for any protein that can be crystallized 241
NMR spectroscopy is used to study the structures of small proteins 243
Gel retardation identifies DNA fragments that bind to proteins 244

Protection assays pinpoint binding sites with greater accuracy 244
Modification interference identifies nucleotides central to protein binding 246
Genomewide scans for protein attachment sites 247

11.2 THE SPECIAL FEATURES OF DNA-BINDING PROTEINS 249
The helix–turn–helix motif is present in prokaryotic and eukaryotic proteins 249
Zinc fingers are common in eukaryotic proteins 250
Other nucleic acid-binding motifs 251

11.3 INTERACTION BETWEEN DNA AND ITS BINDING PROTEINS 252
Direct readout of the nucleotide sequence 252
The nucleotide sequence has a number of indirect effects on helix structure 253
Contacts between DNA and proteins 253

SUMMARY 254

SHORT ANSWER QUESTIONS 255

IN-DEPTH PROBLEMS 256

FURTHER READING 256

CHAPTER 12 TRANSCRIPTOMES 257

12.1 COMPONENTS OF THE TRANSCRIPTOME 257
The mRNA fraction of a transcriptome is small but complex 257
Short noncoding RNAs have diverse functions 259
Long noncoding RNAs are enigmatic transcripts 260
Microarray analysis and RNA sequencing are used to study the contents of transcriptomes 262

12.2 SYNTHESIS OF THE COMPONENTS OF THE TRANSCRIPTOME 263
RNA polymerases are molecular machines for making RNA 264
Transcription start points are indicated by promoter sequences 266
Synthesis of bacterial RNA is regulated by repressor and activator proteins 268
Synthesis of bacterial RNA is also regulated by control over transcription termination 271
Synthesis of eukaryotic RNA is regulated primarily by activator proteins 272

12.3 DEGRADATION OF THE COMPONENTS OF THE TRANSCRIPTOME 275
Several processes are known for nonspecific RNA turnover 275
RNA silencing was first identified as a means of destroying invading viral RNA 276
MicroRNAs regulate genome expression by causing specific target mRNAs to be degraded 278

12.4 INFLUENCE OF RNA PROCESSING ON THE COMPOSITION OF A TRANSCRIPTOME 278
The splicing pathway for eukaryotic pre-mRNA introns 279
The splicing process must have a high degree of precision 280
Enhancer and silencer elements specify alternative splicing pathways 282

12.5 TRANSCRIPTOMES IN RESEARCH 284
Transcriptome analysis as an aid to genome annotation 284
Cancer transcriptomes 286
Transcriptomes and the responses of plants to stress 287

SUMMARY 289

SHORT ANSWER QUESTIONS 289

IN-DEPTH PROBLEMS 290

FURTHER READING 290

CHAPTER 13 PROTEOMES 293

13.1 STUDYING THE COMPOSITION OF A PROTEOME 293
The separation stage of a protein profiling project 294
The identification stage of a protein profiling project 297
Comparing the compositions of two proteomes 299
Analytical protein arrays offer an alternative approach to protein profiling 300

13.2 IDENTIFYING PROTEINS THAT INTERACT WITH ONE ANOTHER 301
Identifying pairs of interacting proteins 301
Identifying the components of multiprotein complexes 304
Identifying proteins with functional interactions 305
Protein interaction maps display the interactions within a proteome 306

13.3 SYNTHESIS AND DEGRADATION OF THE COMPONENTS OF THE PROTEOME 308
Ribosomes are molecular machines for making proteins 308

During stress, bacteria inactivate their ribosomes in order to downsize the proteome 311
Initiation factors mediate large-scale remodeling of eukaryotic proteomes 312
The translation of individual mRNAs can also be regulated 313
Degradation of the components of the proteome 314

13.4 INFLUENCE OF PROTEIN PROCESSING ON THE COMPOSITION OF THE PROTEOME 315
The amino acid sequence contains instructions for protein folding 315
Some proteins are activated by proteolytic cleavage 318
Important changes in protein activity can be brought about by chemical modification 320

13.5 BEYOND THE PROTEOME 322
The metabolome is the complete set of metabolites present in a cell 322
Systems biology provides an integrated description of cellular activity 323

SUMMARY 326
SHORT ANSWER QUESTIONS 326
IN-DEPTH PROBLEMS 327
FURTHER READING 327

CHAPTER 14 GENOME EXPRESSION IN THE CONTEXT OF CELL AND ORGANISM 329

14.1 THE RESPONSE OF THE GENOME TO EXTERNAL SIGNALS 330
Signal transmission by import of the extracellular signaling compound 330
Receptor proteins transmit signals across cell membranes 332
Some signal transduction pathways have few steps between receptor and genome 333
Some signal transduction pathways have many steps between receptor and genome 334
Some signal transduction pathways operate via second messengers 336

14.2 CHANGES IN GENOME ACTIVITY RESULTING IN CELLULAR DIFFERENTIATION 336
Some differentiation processes involve changes to chromatin structure 336
Yeast mating types are determined by gene conversion events 338
Genome rearrangements are responsible for immunoglobulin and T-cell receptor diversity 339

14.3 CHANGES IN GENOME ACTIVITY UNDERLYING DEVELOPMENT 341
Bacteriophage λ: a genetic switch enables a choice to be made between alternative developmental pathways 342
Bacillus sporulation: coordination of activities in two distinct cell types 343
Caenorhabditis elegans: the genetic basis of positional information and the determination of cell fate 346
Fruit flies: conversion of positional information into a segmented body plan 348
Homeotic selector genes are universal features of higher eukaryotic development 350
Homeotic genes also underlie plant development 352

SUMMARY 352
SHORT ANSWER QUESTIONS 353
IN-DEPTH PROBLEMS 354
FURTHER READING 354

CHAPTER 15 GENOME REPLICATION 357

15.1 THE TOPOLOGY OF GENOME REPLICATION 357
The double-helical structure complicates the replication process 358
The Meselson–Stahl experiment proved that replication is semiconservative 359
DNA topoisomerases provide a solution to the topological problem 361
Variations on the semiconservative theme 363

15.2 THE INITIATION PHASE OF GENOME REPLICATION 364
Initiation at the *E. coli* origin of replication 364
Origins of replication have been clearly defined in yeast 365
Origins in higher eukaryotes have been less easy to identify 366

15.3 EVENTS AT THE REPLICATION FORK 367
DNA polymerases are molecular machines for making (and degrading) DNA 367
DNA polymerases have limitations that complicate genome replication 369

Okazaki fragments must be joined together to complete lagging-strand replication 370

15.4 TERMINATION OF GENOME REPLICATION 372
Replication of the *E. coli* genome terminates within a defined region 373
Little is known about termination of replication in eukaryotes 374
Telomerase completes replication of chromosomal DNA molecules, at least in some cells 375
Telomere length is implicated in cell senescence and cancer 378
Drosophila has a unique solution to the end-shortening problem 379

15.5 REGULATION OF EUKARYOTIC GENOME REPLICATION 380
Genome replication must be synchronized with the cell cycle 380
Origin licensing is the prerequisite for passing the G1–S checkpoint 380
Replication origins do not all fire at the same time 382
The cell has various options if the genome is damaged 383

SUMMARY 384

SHORT ANSWER QUESTIONS 385

IN-DEPTH PROBLEMS 385

FURTHER READING 386

CHAPTER 16 MUTATIONS AND DNA REPAIR 389

16.1 THE CAUSES OF MUTATIONS 389
Errors in replication are a source of point mutations 390
Replication errors can also lead to insertion and deletion mutations 391
Mutations are also caused by chemical and physical mutagens 394

16.2 REPAIR OF MUTATIONS AND OTHER TYPES OF DNA DAMAGE 398
Direct repair systems fill in nicks and correct some types of nucleotide modification 398
Base excision repairs many types of damaged nucleotide 399
Nucleotide excision repair is used to correct more extensive types of damage 401
Mismatch repair corrects replication errors 402
Single- and double-strand breaks can be repaired 403
If necessary, DNA damage can be bypassed during genome replication 405
Defects in DNA repair underlie human diseases, including cancers 406

SUMMARY 406

SHORT ANSWER QUESTIONS 407

IN-DEPTH PROBLEMS 407

FURTHER READING 408

CHAPTER 17 RECOMBINATION AND TRANSPOSITION 411

17.1 HOMOLOGOUS RECOMBINATION 412
The Holliday and Meselson–Radding models for homologous recombination 412
The double-strand break model for homologous recombination 414
RecBCD is the most important pathway for homologous recombination in bacteria 415
E. coli can also carry out homologous recombination by the RecFOR pathway 417
Homologous recombination pathways in eukaryotes 417
The primary role of homologous recombination is thought to be DNA repair 418

17.2 SITE-SPECIFIC RECOMBINATION 419
Bacteriophage λ uses site-specific recombination during the lysogenic infection cycle 419
Site-specific recombination is an aid in construction of genetically modified plants 421

17.3 TRANSPOSITION 421
Replicative and conservative transposition of DNA transposons 422
Retroelements transpose replicatively via an RNA intermediate 423

SUMMARY 425

SHORT ANSWER QUESTIONS 426

IN-DEPTH PROBLEMS 427

FURTHER READING 427

CHAPTER 18 HOW GENOMES EVOLVE 429

18.1 GENOMES: THE FIRST 10 BILLION YEARS 429
The first biochemical systems were centered on RNA 429
The first DNA genomes 432
How unique is life? 433

18.2 EVOLUTION OF INCREASINGLY COMPLEX GENOMES 434
Genome sequences provide extensive evidence of past gene duplications 434
A variety of processes could result in gene duplication 438
Whole-genome duplication is also possible 439
Smaller duplications can also be identified in the human genome and other genomes 442
Both prokaryotes and eukaryotes acquire genes from other species 444
Genome evolution also involves rearrangement of existing genes 445
There are competing hypotheses for the origins of introns 448
The evolution of the epigenome 449

18.3 GENOMES: THE LAST 6 MILLION YEARS 450
The human genome is very similar to that of the chimpanzee 451
Paleogenomics is helping us understand the recent evolution of the human genome 452

18.4 GENOMES TODAY: DIVERSITY IN POPULATIONS 453
The origins of HIV/AIDS 454
The first migrations of humans out of Africa 455
The diversity of plant genomes is an aid in crop breeding 457

SUMMARY 458

SHORT ANSWER QUESTIONS 459

IN-DEPTH PROBLEMS 460

FURTHER READING 460

GLOSSARY 463

INDEX 491