

Contents

Preface	v	2.2 DNA Cloning	43
A Note to the Reader	vii	2.2.1 Cloning vectors and the way they are used	44
Contents in Brief	xi	Vectors based on <i>E. coli</i> plasmids	45
Abbreviations	xix	Technical Note 2.3: DNA purification	46
		Cloning vectors based on <i>E. coli</i> bacteriophage genomes	48
		Vectors for longer pieces of DNA	51
		Cloning in organisms other than <i>E. coli</i>	53
PART 1 Studying Genomes	1	2.3 The Polymerase Chain Reaction (PCR)	55
Chapter 1 Genomes, Transcriptomes, and Proteomes	3	2.3.1 Carrying out a PCR	55
1.1 DNA	5	Technical Note 2.4: Working with a clone library	56
1.1.1 Genes are made of DNA	5	2.3.2 The applications of PCR	57
1.1.2 The structure of DNA	8		
Nucleotides and polynucleotides	8	Study Aids	59
The evidence that led to the double helix	9		
The key features of the double helix	11	Chapter 3 Mapping Genomes	63
The double helix has structural flexibility	12	3.1 Genetic and Physical Maps	65
1.2 RNA and the Transcriptome	14	3.2 Genetic Mapping	65
1.2.1 The structure of RNA	15	3.2.1 Genes were the first markers to be used	66
1.2.2 The RNA content of the cell	15	3.2.2 DNA markers for genetic mapping	67
1.2.3 Processing of precursor RNA	17	Restriction fragment length polymorphisms	67
1.2.4 The transcriptome	17	Simple sequence length polymorphisms	68
1.3 Proteins and the Proteome	18	Single nucleotide polymorphisms	69
1.3.1 Protein structure	18	Technical Note 3.1: DNA microarrays and chips	71
The four levels of protein structure	18	3.2.3 Linkage analysis is the basis of genetic mapping	72
Amino acid diversity underlies protein diversity	19	The principles of inheritance and the discovery of linkage	72
1.3.2 The proteome	20	Partial linkage is explained by the behavior of chromosomes during meiosis	74
The link between the transcriptome and the proteome	21	From partial linkage to genetic mapping	77
The genetic code is not universal	22	3.2.4 Linkage analysis with different types of organism	77
The link between the proteome and the biochemistry of the cell	23	Linkage analysis when planned breeding experiments are possible	78
Study Aids	26	Gene mapping by human pedigree analysis	80
		Genetic mapping in bacteria	81
Chapter 2 Studying DNA	31	3.3 Physical Mapping	82
2.1 Enzymes for DNA Manipulation	33	3.3.1 Restriction mapping	84
2.1.1 DNA polymerases	34	The basic methodology for restriction mapping	84
Technical Note 2.1: DNA labeling	34	The scale of restriction mapping is limited by the sizes of the restriction fragments	86
The mode of action of a template-dependent polymerase	35	Direct examination of DNA molecules for restriction sites	87
The types of DNA polymerases used in research	36	3.3.2 Fluorescent <i>in situ</i> hybridization	89
2.1.2 Nucleases	37	<i>In situ</i> hybridization with radioactive or fluorescent probes	89
Restriction endonucleases enable the DNA molecules to be cut at defined positions	38	FISH in action	90
Technical Note 2.2: Agarose gel electrophoresis	40	3.3.3 Sequence tagged site mapping	91
Examining the results of a restriction digest	41	Any unique DNA sequence can be used as an STS	92
2.1.3 DNA ligases	42		
2.1.4 End-modification enzymes	42		

Fragments of DNA for STS mapping	93	Hybridization tests can determine if a	
A clone library can also be used as the		fragment contains transcribed sequences	141
mapping reagent for STS analysis	94	Technical Note 5.1: Techniques for studying RNA	142
Study Aids	97	cDNA sequencing enables genes to be	
		mapped within DNA fragments	142
		Methods are available for precise mapping	
		of the ends of transcripts	143
		Exon–intron boundaries can also be located	
		with precision	144
Chapter 4 Sequencing Genomes	103	5.2 Determining the Functions of Individual	
4.1 The Methodology for DNA Sequencing	104	Genes	144
Technical Note 4.1: Polyacrylamide		5.2.1 Computer analysis of gene function	145
gel electrophoresis	104	Homology reflects evolutionary relationships	145
4.1.1 Chain termination DNA sequencing	105	Homology analysis can provide information	
Chain termination sequencing in outline	105	on the function of an entire gene or of	
Chain termination sequencing requires a		segments within it	145
single-stranded DNA template	107	Using homology searching to assign	
DNA polymerases for chain termination		functions to human disease genes	147
sequencing	108	5.2.2 Assigning gene function by experimental	
The primer determines the region of the		analysis	148
template DNA that will be sequenced	108	Functional analysis by gene inactivation	149
Thermal cycle sequencing offers an		Individual genes can be inactivated by	
alternative to the traditional methodology	109	homologous recombination	149
4.1.2 Alternative methods for DNA sequencing	109	Gene inactivation without homologous	
Chemical degradation sequencing	110	recombination	150
Pyrosequencing is used for rapid		Gene overexpression can also be used to	
determination of very short sequences	111	assess function	151
		The phenotypic effect of gene inactivation or	
4.2 Assembly of a Contiguous DNA Sequence	112	overexpression may be difficult to discern	152
4.2.1 Sequence assembly by the shotgun method	112	5.2.3 More detailed studies of the activity of a	
The potential of the shotgun method was proven		protein coded by an unknown gene	154
by the <i>Haemophilus influenzae</i> sequence	113	Directed mutagenesis can be used to probe	
4.2.2 Sequence assembly by the clone contig		gene function in detail	154
method	115	Reporter genes and immunocytochemistry	
Clone contigs can be built up by chromosome		can be used to locate where and when genes	
walking, but the method is laborious	115	are expressed	155
More rapid methods for clone contig assembly	117	Technical Note 5.2: Site-directed mutagenesis	156
4.2.3 Whole-genome shotgun sequencing	119	5.3 Case Study: Annotation of the <i>Saccharomyces</i>	
Key features of whole-genome shotgun		<i>cerevisiae</i> Genome Sequence	158
sequencing	119	5.3.1 Annotation of the yeast genome sequence	158
4.3 The Human Genome Projects	121	5.3.2 Assigning functions to yeast genes	159
4.3.1 The mapping phase of the Human Genome		Study Aids	162
Project	121		
4.3.2 Sequencing the human genome	122	Chapter 6 Understanding How a Genome	
4.3.3 The future of the human genome projects	123	Functions	167
Study Aids	126	6.1 Studying the Transcriptome	168
		6.1.1 Studying a transcriptome by sequence analysis	168
		6.1.2 Studying a transcriptome by microarray	
		or chip analysis	169
		Using a microarray or chip to study one	
		or more transcriptomes	169
		Studies of the yeast transcriptome	172
		The human transcriptome	173
		6.2 Studying the Proteome	175
		6.2.1 Protein profiling – methodology for	
		identifying the proteins in a proteome	175
Chapter 5 Understanding a Genome			
Sequence	133		
5.1 Locating the Genes in a Genome Sequence	134		
5.1.1 Gene location by sequence inspection	134		
The coding regions of genes are open			
reading frames	134		
Simple ORF scans are less effective with			
DNA of higher eukaryotes	135		
Locating genes for functional RNA	137		
Homology searches and comparative genomics			
give an extra dimension to sequence inspection	138		
Automatic annotation of genome sequences	140		
5.1.2 Experimental techniques for gene location	141		

Separating the proteins in a proteome	175	The traditional view of the prokaryotic chromosome	226
Identifying the proteins in a proteome	177	Some bacteria have linear or multipartite genomes	228
6.2.2 Identifying proteins that interact with one another	179	8.2 The Genetic Features of Prokaryotic Genomes	230
Identifying pairs of interacting proteins by phage display and two-hybrid studies	179	8.2.1 How are the genes organized in a prokaryotic genome?	230
Identifying the components of multiprotein complexes	181	Gene organization in the <i>E. coli</i> genome	231
Identifying proteins with functional interactions	182	Operons are characteristic features of prokaryotic genomes	232
Protein interaction maps	183	8.2.2 How many genes are there and what are their functions?	234
6.3 Beyond the Proteome	184	8.2.3 Prokaryotic genomes and the species concept	236
6.3.1 The metabolome	185	8.3 Eukaryotic Organelle Genomes	238
6.3.2 Understanding biological systems	186	8.3.1 The origins of organelle genomes	238
Study Aids	189	8.3.2 Physical features of organelle genomes	239
		8.3.3 The genetic content of organelle genomes	239
		Study Aids	244
PART 2 Genome Anatomies	195	Chapter 9 Virus Genomes and Mobile Genetic Elements	249
Chapter 7 Eukaryotic Nuclear Genomes	197	9.1 The Genomes of Bacteriophages and Eukaryotic Viruses	250
7.1 Nuclear Genomes are Contained in Chromosomes	198	9.1.1 Bacteriophage genomes	250
7.1.1 Packaging of DNA into chromosomes	198	Bacteriophage genomes have diverse structures and organizations	250
7.1.2 The special features of metaphase chromosomes	199	Replication strategies for bacteriophage genomes	251
DNA–protein interactions in centromeres and telomeres	202	9.1.2 The genomes of eukaryotic viruses	253
		Structures and replication strategies for eukaryotic viral genomes	253
7.2 The Genetic Features of Eukaryotic Nuclear Genomes	203	Genomes at the edge of life	254
7.2.1 Where are the genes in a nuclear genome?	204	9.2 Mobile Genetic Elements	256
Technical Note 7.1: Ultracentrifugation techniques	205	9.2.1 Transposition via an RNA intermediate	257
7.2.2 How are the genes organized in a nuclear genome?	205	RNA transposons with long terminal repeats are related to viral retroelements	257
The genes make up only a small part of the human genome	206	RNA transposons that lack LTRs	259
The yeast genome is very compact	207	9.2.2 DNA transposons	259
Gene organization in other eukaryotes	210	DNA transposons are common in prokaryotic genomes	260
7.2.3 How many genes are there and what are their functions?	211	DNA transposons are less common in eukaryotic genomes	261
The human gene catalog	212	Study Aids	264
Gene catalogs reveal the distinctive features of different organisms	212	PART 3 How Genomes Function	269
Families of genes	215	Chapter 10 Accessing the Genome	271
Pseudogenes and other evolutionary relics	216	10.1 Inside the Nucleus	272
7.2.4 The repetitive DNA content of eukaryotic nuclear genomes	216	10.1.1 The internal architecture of the eukaryotic nucleus	272
Tandemly repeated DNA is found at centromeres and elsewhere in eukaryotic chromosomes	217	The nucleus has a highly ordered internal structure	273
Minisatellites and microsatellites	217	Technical Note 10.1: Fluorescence recovery after photobleaching (FRAP)	274
Interspersed repeats	218		
Study Aids	220		
Chapter 8 Genomes of Prokaryotes and Eukaryotic Organelles	225		
8.1 The Physical Features of Prokaryotic Genomes	226		
8.1.1 The chromosomes of prokaryotes	226		

Each chromosome has its own territory within the nucleus 274

10.1.2 Chromatin domains 275

Functional domains are defined by insulators 276

Some functional domains contain locus control regions 278

10.2 Chromatin Modifications and Genome Expression 279

10.2.1 Chemical modification of histones 280

Acetylation of histones influences many nuclear activities including genome expression 280

Histone deacetylation represses active regions of the genome 282

Acetylation is not the only type of histone modification 282

10.2.2 The influence of nucleosome remodeling on genome expression 284

10.3 DNA Modification and Genome Expression 285

10.3.1 Genome silencing by DNA methylation 285

DNA methyltransferases and the repression of genome activity 286

Methylation is involved in genomic imprinting and X inactivation 287

Study Aids 290

Chapter 11 Assembly of the Transcription Initiation Complex 295

11.1 DNA-binding Proteins and Their Attachment Sites 297

11.1.1 The special features of DNA-binding proteins 297

The helix-turn-helix motif is present in prokaryotic and eukaryotic proteins 297

Technical Note 11.1: X-ray crystallography and nuclear magnetic resonance spectroscopy 298

Zinc fingers are common in eukaryotic proteins 301

Other nucleic acid-binding motifs 301

11.1.2 Locating the positions of DNA-binding sites in a genome 302

Gel retardation identifies DNA fragments that bind to proteins 303

Protection assays pinpoint binding sites with greater accuracy 303

Modification interference identifies nucleotides central to protein binding 304

11.1.3 The interaction between DNA and its binding proteins 305

Direct readout of the nucleotide sequence 306

The nucleotide sequence has a number of indirect effects on helix structure 306

Contacts between DNA and proteins 307

11.2 DNA-Protein Interactions During Transcription Initiation 308

11.2.1 RNA polymerases 308

11.2.2 Recognition sequences for transcription initiation 309

Bacterial RNA polymerases bind to promoter sequences 309

Eukaryotic promoters are more complex 310

11.2.3 Assembly of the transcription initiation complex 312

Transcription initiation in *E. coli* 312

Transcription initiation with RNA polymerase II 312

Transcription initiation with RNA polymerases I and III 315

11.3 Regulation of Transcription Initiation 315

11.3.1 Strategies for controlling transcription initiation in bacteria 316

Promotor structure determines the basal level of transcription initiation 316

Regulatory control over bacterial transcription initiation 317

11.3.2 Control of transcription initiation in eukaryotes 320

Eukaryotic promoters contain regulatory modules 321

Activators and coactivators of eukaryotic transcription initiation 322

The mediator forms the contact between an activator and the RNA polymerase II preinitiation complex 323

Repressors of eukaryotic transcription initiation 324

Controlling the activities of activators and repressors 325

Study Aids 327

Chapter 12 Synthesis and Processing of RNA 333

12.1 Synthesis and Processing of Bacterial RNAs 334

12.1.1 Synthesis of bacterial transcripts 335

Elongation of a transcript by the bacterial RNA polymerase 335

Termination of a bacterial transcript 337

12.1.2 Control over the choice between elongation and termination 338

Antitermination results in termination signals being ignored 338

Attenuation results in premature termination 340

Transcript cleavage proteins can prevent stalling of a backtracked polymerase 341

12.1.3 Processing of bacterial RNAs 343

Cutting events release mature rRNAs and tRNAs from their precursor molecules 343

Nucleotide modifications broaden the chemical properties of tRNAs and rRNAs 346

12.1.4 Degradation of bacterial RNAs 346

Bacterial mRNAs are degraded in the 3'→5' direction 347

12.2 Synthesis and Processing of Eukaryotic RNA 348

12.2.1 Synthesis of eukaryotic mRNAs by RNA polymerase II 348

Capping of RNA polymerase II transcripts occurs immediately after initiation 348

Elongation of eukaryotic mRNAs 350

Termination of synthesis of most mRNAs is combined with polyadenylation	351	13.2.3 Regulation of translation initiation	399
12.2.2 Regulation of mRNA synthesis in eukaryotes	353	13.2.3 The elongation phase of translation	400
12.2.2 Removal of introns from nuclear pre-mRNA	354	Elongation in bacteria and eukaryotes	400
Conserved sequence motifs indicate the key sites in GU-AG introns	355	Peptidyl transferase is a ribozyme	402
Outline of the splicing pathway for GU-AG introns	356	Frameshifting and other unusual events during elongation	403
snRNAs and their associated proteins are the central components of the splicing apparatus	357	13.2.4 Termination of translation	405
Alternative splicing is common in many eukaryotes	359	13.2.5 Translation in the archaea	405
Trans-splicing links exons from different transcription units	362	13.3 Posttranslational Processing of Proteins	406
AU-AC introns are similar to GU-AG introns but require a different splicing apparatus	363	13.3.1 Protein folding	407
12.2.3 Synthesis of functional RNAs in eukaryotes	363	Not all proteins fold spontaneously in the test tube	407
12.2.4 Splicing of eukaryotic pre-rRNA and pre-tRNA	364	In cells, folding is aided by molecular chaperones	409
Introns in eukaryotic pre-rRNAs are autocatalytic	364	13.3.2 Processing by proteolytic cleavage	410
Removal of introns from eukaryotic pre-tRNAs	365	Cleavage of the ends of polypeptides	410
Other types of intron	367	Proteolytic processing of polyproteins	411
12.2.5 Chemical modification of eukaryotic RNAs	368	13.3.3 Processing by chemical modification	412
Small nucleolar RNAs act as guides for chemical modification of eukaryotic rRNAs	369	13.3.4 Inteins	413
RNA editing	369	13.4 Protein Degradation	414
12.2.6 Degradation of eukaryotic RNAs	371	Study Aids	417
Eukaryotes have diverse mechanisms for RNA degradation	371	Chapter 14 Regulation of Genome Activity	423
RNA silencing was first identified as a means of destroying invading viral RNA	373	14.1 Transient Changes in Genome Activity	425
MicroRNAs regulate genome expression by causing specific target mRNAs to be degraded	374	14.1.1 Signal transmission by import of the extracellular signaling compound	427
12.2.7 Transport of RNA within the eukaryotic cell	375	Lactoferrin is an extracellular signaling protein which acts as a transcription activator	428
Study Aids	377	Some imported signaling compounds directly influence the activity of preexisting regulatory proteins	428
Chapter 13 Synthesis and Processing of the Proteome	385	Some imported signaling compounds influence genome activity indirectly	429
13.1 The Role of tRNA in Protein Synthesis	386	14.1.2 Signal transmission mediated by cell surface receptors	432
13.1.1 Aminoacylation: the attachment of amino acids to tRNAs	386	Signal transduction with one step between receptor and genome	433
All tRNAs have a similar structure	386	Signal transduction with many steps between receptor and genome	434
Aminoacyl-tRNA synthetases attach amino acids to tRNAs	388	Signal transduction via second messengers	435
13.1.2 Codon-anticodon interactions: the attachment of tRNAs to mRNA	390	Unraveling a signal transduction pathway	436
13.2 The Role of the Ribosome in Protein Synthesis	392	14.2 Permanent and Semipermanent Changes in Genome Activity	437
13.2.1 Ribosome structure	393	14.2.1 Genome rearrangements	438
Ultracentrifugation was used to measure the sizes of ribosomes and their components	393	Yeast mating types are determined by gene conversion events	438
Probing the fine structure of the ribosome	393	Genome rearrangements are responsible for immunoglobulin and T-cell receptor diversities	439
13.2.2 Initiation of translation	395	14.2.2 Changes in chromatin structure	441
Initiation in bacteria requires an internal ribosome binding site	395	14.2.3 Genome regulation by feedback loops	443
Initiation in eukaryotes is mediated by the cap structure and poly(A) tail	397	14.3 Regulation of Genome Activity During Development	443
Initiation of eukaryotic translation without scanning	398	14.3.1 The lysogenic cycle of bacteriophage λ	444

Chapter 17 Recombination	541	18.4 The Human Genome: the Last Five Million Years	586
17.1 Homologous Recombination	543	Study Aids	589
17.1.1 Models for homologous recombination	543	Chapter 19 Molecular Phylogenetics	595
The Holliday and Meselson–Radding models for homologous recombination	543	19.1 From Classification to Molecular Phylogenetics	596
The double-strand break model for homologous recombination	545	19.1.1 The origin of molecular phylogenetics	596
17.1.2 The biochemistry of homologous recombination	546	Phenetics and cladistics require large datasets	596
The RecBCD pathway of <i>Escherichia coli</i>	546	Large datasets can be obtained by studying molecular characters	597
Other homologous recombination pathways in <i>E. coli</i>	547	19.2 The Reconstruction of DNA-based Phylogenetic Trees	599
Homologous recombination pathways in eukaryotes	548	19.2.1 The key features of DNA-based phylogenetic trees	599
17.1.3 Homologous recombination and DNA repair	549	Gene trees are not the same as species trees	600
17.2 Site-Specific Recombination	550	19.2.2 Tree reconstruction	602
17.2.1 Integration of λ DNA into the <i>E. coli</i> genome	550	Sequence alignment is the essential preliminary to tree reconstruction	602
17.2.2 Site-specific recombination is an aid in genetic engineering	551	Converting alignment data into a phylogenetic tree	603
17.3 Transposition	552	Technical Note 19.1: Phylogenetic analysis	604
17.3.1 Replicative and conservative transposition of DNA transposons	553	Assessing the accuracy of a reconstructed tree	606
17.3.2 Transposition of retroelements	553	Molecular clocks enable the time of divergence of ancestral sequences to be estimated	606
17.3.3 How do cells minimize the harmful effect of transposition?	556	Standard tree reconstruction is not appropriate for all DNA sequence datasets	607
Study Aids	558	19.3 Applications of Molecular Phylogenetics	609
Chapter 18 How Genomes Evolve	563	19.3.1 Examples of the use of phylogenetic trees	609
18.1 Genomes: the First Ten Billion Years	564	DNA phylogenetics has clarified the evolutionary relationships between humans and other primates	609
18.1.1 The origins of genomes	565	The origins of AIDS	610
The first biochemical systems were centered on RNA	565	19.3.2 Molecular phylogenetics as a tool in the study of human prehistory	611
The first DNA genomes	566	Studying genes in populations	611
How unique is life?	568	The origins of modern humans – out of Africa or not?	612
18.2 Acquisition of New Genes	568	Neandertals are not the ancestors of modern Europeans	614
18.2.1 Acquisition of new genes by duplication events	570	The patterns of more recent migrations into Europe are also controversial	616
Genome sequences provide extensive evidence of past gene duplications	571	Prehistoric human migrations into the New World	617
A variety of processes could result in gene duplication	573	Study Aids	621
Whole genome duplication is also possible	574	Appendix	627
Analysis of modern genomes provides evidence for past genome duplications	576	Glossary	653
Smaller duplications can also be identified in the human genome and other genomes	577	Index	683
Genome evolution also involves rearrangement of existing genes	578		
18.2.2 Acquisition of new genes from other species	581		
18.3 Noncoding DNA and Genome Evolution	582		
18.3.1 Transposable elements and genome evolution	583		
18.3.2 The origins of introns	583		
“Introns early” and “introns late”: two competing hypotheses”	584		
The current evidence disproves neither hypothesis	585		