

CONTENTS

Preface	vii
Plan of the third edition	x
Recommended reading	xiii
Introduction to genomics on the web	xiii
Acknowledgements	xiv
List of abbreviations	xxv
1 Introduction and Background	1
Learning goals	1
Genomics: the hub of biology	2
Phenotype = genotype + environment + life history + epigenetics	2
Varieties of genome organization	4
Chromosomes, organelles, and plasmids	4
Genes	6
The scope and applications of genome sequencing projects	9
Variations in genome sequences within species	10
Mutations and disease	10
Single-nucleotide polymorphisms	10
Haplotypes	13
A clinically important haplotype: the major histocompatibility complex	14
Populations	15
Species	17
The biosphere	18
Extinctions	19
The future?	22
Genome projects and our current library of genome information	25
High-throughput sequencing	25
<i>De novo</i> sequencing	26
Resequencing	26
Exome sequencing	26
What's in a genome?	27
Some regions of the genome encode non-protein-coding RNA molecules	28
Some regions of the genome contain pseudogenes	28
Other regions contain binding sites for ligands responsible for regulation of transcription	29
Repetitive elements of unknown function account for surprisingly large fractions of our genomes	29
Dynamic components of genomes	30
Genomics and developmental biology	32
Genes and minds: neurogenomics	35
Genetics of behaviour	36

Proteomics	37
Protein evolution: divergence of sequences and structures within and among species	39
Mechanisms of protein evolution	39
Organization and regulation	42
Some mechanisms of regulation act at the level of transcription	45
Some mechanisms of regulation act at the level of translation	45
Some regulatory mechanisms affect protein activity	47
On the web: genome browsers	47
Genomics and computing	52
Archiving and analysis of genome sequences and related data	52
Databanks in molecular biology	53
Programming	54
Looking forward	55
Recommended reading	55
Exercises and problems	57
2 The Human Genome Project: Achievements and Applications	62
Learning goals	62
'... the end of the beginning'	63
Human genome sequencing	66
What makes us human?	67
Comparative genomics	68
Genomics and language	68
The human genome and medicine	71
Prevention of disease	71
Detection and precise diagnosis	72
Genetic counselling—carrier status	72
Discovery and implementation of effective treatment	72
Tunable healthcare delivery: pharmacogenomics	74
'Pop' applications of genome sequencing	76
Genomics in personal identification	76
DNA 'fingerprinting'	77
Personal identification by amplification of specific regions has superseded the RFLP approach	79
Mitochondrial DNA	79
Analysis of non-human DNA sequences	82
Parentage testing	82
Inference of physical features, and even family name	83
Ethical, legal, and social issues	85
Databases containing human DNA sequence information	85
Use of DNA sequencing in research on human subjects	88
Looking forward	88
Recommended reading	88
Exercises and problems	90

3 Mapping, Sequencing, Annotation, and Databases	97
Learning goals	97
Classical genetics as background	98
What is a gene?	99
Maps and tour guides	99
Genetic maps	100
Linkage	101
Linkage disequilibrium	102
Chromosome banding pattern maps	103
Restriction maps	106
Discovery of the structure of DNA	107
DNA sequencing	110
Frederick Sanger and the development of DNA sequencing	111
DNA sequencing by termination of chain replication	111
The Maxam–Gilbert chemical cleavage method	114
Automation of DNA sequencing	114
Organizing a large-scale sequencing project	116
Bring on the clones: hierarchical—or ‘BAC-to-BAC’—genome sequencing	116
Whole-genome shotgun sequencing	116
Next-generation sequencing	118
Roche 454 Life Sciences	120
Illumina	121
Ion Torrent/Personal Genome Machine (PGM)	124
PacBio	124
Oxford Nanopore	124
10X Genomics	125
The Bionano Irys system	125
Life in the fast lane	126
How much sequencing power is there in the world?	126
Databanks in molecular biology	128
Nucleic acid sequence databases	130
Protein sequence databases	130
Databases of genetic diseases—OMIM and OMIA	130
Databases of structures	131
Specialized or ‘boutique’ databases	132
Expression and proteomics databases	132
Databases of metabolic pathways	133
Bibliographic databases	133
Surveys of molecular biology databases and servers	134
Computer programming in genomics	134
Programming languages	135
How to compute effectively	136
Looking forward	137
Recommended reading	137
Exercises and problems	138

4. Evolution and Genomic Change	143
Learning goals	143
Evolution is exploration	144
Biological systematics	146
Biological nomenclature	146
Measurement of biological similarities and differences	148
Homologues and families	150
Pattern matching—the basic tool of bioinformatics	151
Sequence alignment	151
Defining the optimum alignment	153
Scoring schemes	155
Varieties and extensions	157
Approximate methods for quick screening of databases	158
Pattern matching in three-dimensional structures	160
Evolution of protein sequences, structures, and functions	161
Evolution of protein structure and function	164
Phylogeny	165
Calculation of phylogenetic trees	169
Short-circuiting evolution: genetic engineering	173
Looking forward	175
Recommended reading	176
Exercises and problems	177
 5 Genomes of Prokaryotes and Viruses	 179
Learning goals	179
Evolution and phylogenetic relationships in prokaryotes	180
Major types of prokaryotes	180
Do we know the root of the tree of life?	182
Genome organization in prokaryotes	183
Replication and transcription	184
Gene transfer	184
Archaea	184
The genome of <i>Methanococcus jannaschii</i>	186
Life at extreme temperatures	188
Comparative genomics of hyperthermophilic archaea:	
<i>Thermococcus kodakarensis</i> and pyrococci	189
Bacteria	194
Genomes of pathogenic bacteria	195
Genomics and the development of vaccines	197
Viruses	198
Nucleocytoplasmic large DNA viruses (or giant viruses)	199
Viral genomes	199
Recombinant viruses	199
Viruses and evolution	201
Influenza: a past and current threat	201

'Ome, 'ome, on the range: metagenomics, the genomes in a coherent environmental sample	204
Marine cyanobacteria—an in-depth study	205
Looking forward	208
Recommended reading	208
Exercises and problems	210
6 Genomes of Eukaryotes	211
Learning goals	211
The origin and evolution of eukaryotes	212
Evolution and phylogenetic relationships in eukaryotes	212
The yeast genome	212
The evolution of plants	214
The <i>Arabidopsis thaliana</i> genome	214
Genomes of animals	216
The genome of the sea squirt (<i>Ciona intestinalis</i>)	216
The genome of the pufferfish (<i>Tetraodon nigroviridis</i>)	217
The genome of the chicken (<i>Gallus gallus domesticus</i>)	220
The genome of the platypus (<i>Ornithorhynchus anatinus</i>)	221
The genome of the dog	223
Palaeosequencing—ancient DNA	225
Recovery of DNA from ancient samples	225
DNA from extinct birds	226
High-throughput sequencing of mammoth DNA	227
The phylogeny of elephants	230
Looking forward	230
Recommended reading	231
Exercises and problems	231
7 Comparative Genomics	233
Learning goals	233
Introduction	234
Unity and diversity of life	234
Taxonomy based on sequences	235
Sizes and organization of genomes	238
Genome sizes	238
Genome organization in eukaryotes	241
Photosynthetic sea slugs: endosymbiosis of chloroplasts	242
How genomes differ	243
Variation at the level of individual nucleotides	243
Duplications	243
Duplication of genes	244

Family expansion: G protein-coupled receptors	247
Comparisons at the chromosome level: synteny	252
What makes us human?	252
Comparative genomics	252
Genomes of chimpanzees and humans	252
Genomes of mice and rats	254
Model organisms for study of human diseases	255
The genome of <i>Caenorhabditis elegans</i>	256
The genome of <i>Drosophila melanogaster</i>	256
Homologous genes in humans, worms, and flies	257
Looking forward	258
Recommended reading	260
Exercises and problems	261
 8 The Impact of Genome Sequences on Human Health and Disease	 264
Learning goals	264
Introduction	265
Some diseases are associated with mutations in specific genes	265
Haemoglobinopathies—molecular diseases caused by abnormal haemoglobins	265
Phenylketonuria	266
Alzheimer's disease	268
Identification of genes associated with inherited diseases	268
Genome-wide association studies	271
GWAS of sickle-cell disease	274
GWAS of type 2 diabetes	275
GWAS of schizophrenia	276
The human microbiome	278
Treatment of abnormal microbiome composition	281
Cancer genomics	281
SNPs and cancer	284
Whole-genome sequencing association studies of breast cancer	285
Copy-number alterations in cancer	286
Chromosomal aberrations	288
Epigenetics and cancer	289
microRNAs and cancer	290
Immunotherapy for cancer	291
Looking forward	292
Recommended reading	292
Exercises and problems	294

9 Genomics and Anthropology: Human Evolution, Migration, and Domestication of Plants and Animals	295
Learning goals	295
Ancestry of <i>Homo sapiens</i>	296
The Neanderthal genome	297
The Denisovan genome	298
What do these data tell us?	299
What have Neanderthals and Denisovans done for us lately?	300
Ancient populations and migrations	300
Western civilization? 'I think it would be a good idea'	308
Domestication of the dog	310
Domestication of the horse	311
Domestication of crops	313
Maize (<i>Zea mays</i>)	315
Rice (<i>Oryza sativa</i>)	317
Control of flowering time	318
History of rice domestication	320
Chocolate (<i>Theobroma cacao</i>)	321
The <i>Theobroma cacao</i> genome	322
Looking forward	324
Recommended reading	326
Exercises and problems	327
10 Transcriptomics	329
Learning goals	329
Introduction	330
Microarrays	331
Microarray data are semiquantitative	332
Applications of DNA microarrays	332
Analysis of microarray data	333
RNAseq	336
RNAseq versus microarrays	337
Expression patterns in different physiological states	339
Sleep in rats and fruit flies	340
Expression pattern changes in development	342
Variation of expression patterns during the life cycle of <i>Drosophila melanogaster</i>	342
Flower formation in roses	344
Expression patterns in learning and memory: long-term potentiation	347
Conserved clusters of co-expressing genes	350

Evolutionary changes in expression patterns	351
Applications of transcriptomics in medicine	353
Development of antibiotic resistance in bacteria	353
Childhood leukaemias	357
The Encyclopedia of DNA Elements (ENCODE)	359
Looking forward	359
Recommended reading	360
Exercises and problems	361
11 Proteomics	363
Learning goals	363
Introduction	364
Protein nature and types of proteins	364
Protein structure	365
The chemical structure of proteins	365
Conformation of the polypeptide chain	367
Protein folding patterns	367
Domains	370
Disorder in proteins	370
Post-translational modifications	372
Why is there a common genetic code with 20 canonical amino acids?	374
Separation and analysis of proteins	375
Polyacrylamide gel electrophoresis (PAGE)	375
Two-dimensional PAGE	375
Mass spectrometry	376
Identification of components of a complex mixture	377
Protein sequencing by mass spectrometry	378
Quantitative analysis of relative abundance	378
Measuring deuterium exchange in proteins	379
Experimental methods of protein structure determination	381
X-ray crystallography of proteins	381
Interpretation of the electron density: model building and improvement	383
How accurate are the structures?	384
NMR spectroscopy in structural biology	386
Protein structure determination by NMR	386
Low-temperature electron microscopy (cryoEM)	387
Classifications of protein structures	388
SCOP	389
SCOP2	391
Protein complexes and aggregates	391
Protein aggregation diseases	392
Properties of protein–protein complexes	393
Stoichiometry—what is the composition of the complex?	393
Affinity—how stable is the complex?	394

How are complexes organized in three dimensions?	395
Multisubunit proteins	396
Many proteins change conformation as part of the mechanism of their function	396
Conformational change during enzymatic catalysis	397
Motor proteins	399
Allosteric regulation of protein function	401
Allosteric changes in haemoglobin	402
Conformational states of serine protease inhibitors (serpins)	404
Protein structure prediction and modelling	405
Homology modelling	406
Secondary structure prediction	408
Prediction of novel folds: ROSETTA	408
Available protocols for protein structure prediction	409
Structural genomics	411
Directed evolution and protein design	412
Directed evolution of subtilisin E	412
Looking forward	414
Recommended reading	414
Exercises and problems	415
12 Metabolomics	419
Learning goals	419
Introduction	419
Classification and assignment of protein function	420
The Enzyme Commission	420
The Gene Ontology™ Consortium protein function classification	421
Comparison of EC and GO classifications	423
Metabolic networks	424
Databases of metabolic pathways	425
EcoCyc	425
The Kyoto Encyclopedia of Genes and Genomes	427
The Human Metabolome Database	428
Evolution and phylogeny of metabolic pathways	429
Alignment and comparison of metabolic pathways	431
Comparing linear metabolic pathways	432
Reconstruction of metabolic networks	432
Comparing non-linear metabolic pathways: the pentose phosphate pathway and the Calvin–Benson cycle	434
Metabolomics in ecology	435
Dynamic modelling of metabolic pathways	437
Looking forward	439
Recommended reading	439
Exercises and problems	441

13 Systems Biology	443
Learning goals	443
Introduction	444
Regulatory mechanisms	444
Two parallel networks: physical and logical	445
Networks and graphs	446
Robustness and redundancy	447
Connectivity in networks	448
Dynamics, stability, and robustness	449
Protein complexes and aggregates	451
Protein interaction networks	451
Protein–DNA interactions	456
DNA–protein complexes	456
Structural themes in protein–DNA binding and sequence recognition	457
Bacteriophage T7 DNA polymerase	458
Some protein–DNA complexes that regulate gene transcription	459
Regulatory networks	463
Structures of regulatory networks	464
Structural biology of regulatory networks	465
Gene regulation	466
The transcriptional regulatory network of <i>Escherichia coli</i>	466
The genetic switch of bacteriophage λ	469
Regulation of the lactose operon in <i>Escherichia coli</i>	472
The genetic regulatory network of <i>Saccharomyces cerevisiae</i>	474
Adaptability of the yeast regulatory network	476
Looking forward	479
Recommended reading	479
Exercises and problems	480
 Epilogue	 483
Glossary	484
Index	493