

CONTENTS

Preface to the first edition	vii
Preface to the second edition	ix
Plan of the second edition	x
New to this edition	xii
Recommended reading	xiii
Introduction to genomics on the web	xiii
Acknowledgements	xiii

*Sections marked with * are clinically related*

1 Introduction to Genomics	3
The human genome	4
Phenotype = genotype + environment + life history + epigenetics	4
Contents of the human genome	6
Genes that encode the proteome	8
The leap from the one-dimensional world of sequences to the three-dimensional world we inhabit	9
Varieties of genome organization	10
Chromosomes, organelles, and plasmids	10
Genes	13
Dynamic components of genomes	15
Genome sequencing projects	17
Genome projects and the development of our current information library	18
Variations within and between populations	20
Cancer genome sequencing*	20
Human genome sequencing	20
The human genome and medicine*	21
Prevention of disease	21
Detection and precise diagnosis	22
Discovery and implementation of effective treatment	22
Health care delivery	23
The evolution and development of databases	24
Databank evo-devo	25
Genome browsers	25
Protein evolution: divergence of sequences and structures within and between species	30
Different globins diverged from a common ancestor	30
Ethical, legal, and social issues (*, partly)	32
Databases containing human DNA sequence information	32
Recommended Reading	35
Exercises, Problems, and Weblems	36

2	Genomes are the Hub of Biology	41
	Individuals, populations, the biosphere: past, present, and future	42
	The central dogma, and peripheral ones	43
	Expression patterns	43
	Regulation of gene expression	44
	Proteomics	48
	Genomics and developmental biology	48
	Genes and minds: neurogenomics	50
	Populations	52
	Single-nucleotide polymorphisms (SNPs) and haplotypes	52
	A clinically important haplotype: the major histocompatibility complex*	55
	Mutations and disease*	57
	Genetic diseases – some examples of their causes and treatment*	59
	Haemoglobinopathies – molecular diseases caused by abnormal haemoglobins	59
	Phenylketonuria	60
	Alzheimer's disease	61
	SNPs and cancer	62
	Species	66
	The biosphere	67
	Extinctions	67
	Recommended Reading	70
	Exercises, Problems, and Weblems	71
3	Mapping, Sequencing, Annotation, and Databases	79
	Classical genetics as background	80
	What is a gene?	81
	Maps and tour guides	81
	Genetic maps	82
	Linkage	82
	Linkage disequilibrium	83
	Chromosome banding pattern maps	84
	High-resolution maps, based directly on DNA sequences	88
	Restriction maps	89
	Discovery of the structure of DNA	90
	DNA sequencing	93
	Frederick Sanger and the development of DNA sequencing	94
	The Maxam–Gilbert chemical cleavage method	97
	Automated DNA sequencing	97
	Organizing a large-scale sequencing project	98
	Bring on the clones: hierarchical – or 'BAC-to-BAC' – genome sequencing	98
	Whole-genome shotgun sequencing	99
	High-throughput sequencing	100
	Life in the fast lanes	103
	Databanks in molecular biology	104
	Nucleic acid sequence databases	106

Protein sequence databases	106
Databases of genetic diseases – OMIM and OMIA*	106
Databases of structures	107
Classifications of protein structures	107
Specialized or 'boutique' databases	107
Expression and proteomics databases	107
Databases of metabolic pathways	109
Bibliographic databases	109
Surveys of molecular biology databases and servers	109
Recommended Reading	110
Exercises, Problems, and Weblems	110
4 Comparative Genomics	115
Introduction	116
Unity and diversity of life	116
Taxonomy based on sequences	117
Sizes and organization of genomes	121
Genome sizes	121
Viral genomes	124
Recombinant viruses	124
Influenza: a past and current threat*	126
Genome organization in prokaryotes	129
Replication and transcription	129
Gene transfer	130
Genome organization in eukaryotes	132
Photosynthetic sea slugs: endosymbiosis of chloroplasts	132
How genomes differ	133
Variation at the level of individual nucleotides	133
Duplications	134
Comparisons at the chromosome level: synteny	143
What makes us human?	143
Comparative genomics	143
Combining the approaches: the <i>FOXP2</i> gene	144
Genomes of chimpanzees and humans	144
Genomes of mice and rats	145
Model organisms for study of human diseases*	146
The genome of <i>Caenorhabditis elegans</i>	147
The genome of <i>Drosophila melanogaster</i>	147
Homologous genes in humans, worms, and flies	148
The ENCODE project	151
The modENCODE project	153
Recommended Reading	154
Exercises, Problems, and Weblems	155

5	Evolution and Genomic Change	161
	Evolution is exploration	162
	Biological systematics	164
	Biological nomenclature	164
	Measurement of biological similarities and differences	165
	Homologues and families	166
	Pattern matching – the basic tool of bioinformatics	167
	Sequence alignment	167
	Defining the optimum alignment	171
	Approximate methods for quick screening of databases	173
	Pattern matching in three-dimensional structures	176
	Evolution of protein sequences, structures, and functions	176
	The effects of single-site mutations	177
	Evolution of protein structure and function	178
	Phylogeny	181
	Phylogenetic trees	183
	Clustering methods	184
	Cladistic methods	185
	The problem of varying rates of evolution	185
	Bayesian methods	186
	Short-circuiting evolution: genetic engineering	186
	Recommended Reading	187
	Exercises, Problems, and Weblems	188
6	Genomes of Prokaryotes	191
	Evolution and phylogenetic relationships in prokaryotes	192
	Major types of prokaryotes	192
	Do we know the root of the tree of life?	194
	Archaea	195
	The genome of <i>Methanococcus jannaschii</i>	197
	Life at extreme temperatures	197
	Comparative genomics of hyperthermophilic archaea: <i>Thermococcus kodakarensis</i> and <i>Pyrococci</i>	201
	Bacteria	204
	Genomes of pathogenic bacteria*	204
	Genomics and the development of vaccines*	206
	Metagenomics: the collection of genomes in a coherent environmental sample	208
	Marine cyanobacteria – an in-depth study	208
	Recommended Reading	212
	Exercises, Problems, and Weblems	212

7	Genomes of Eukaryotes	215
	The origin and evolution of eukaryotes	216
	Evolution and phylogenetic relationships in eukaryotes	216
	The yeast genome	216
	The evolution of plants	218
	The genome of the sea squirt (<i>Ciona intestinalis</i>)	220
	The genome of the pufferfish (<i>Tetraodon nigroviridis</i>)	221
	The chicken genome	224
	The platypus genome (<i>Ornithorhynchus anatinus</i>)	225
	The dog genome	226
	Palaeosequencing – ancient DNA	229
	Recovery of DNA from ancient samples	229
	DNA from extinct birds	230
	The moas of New Zealand	230
	The dodo and the solitaire	232
	High-throughput sequencing of mammoth DNA	232
	The mammoth nuclear genome	233
	The phylogeny of elephants	233
	Recommended Reading	234
	Exercises, Problems, and Weblems	235
8	Genomics and Human Biology	237
	Genomics in personal identification	238
	Mitochondrial DNA	239
	Gender identification	240
	Physical characteristics	241
	The domestication of crops	241
	Maize (<i>Zea mays</i>)	243
	Rice (<i>Oryza sativa</i>)	245
	Chocolate (<i>Theobroma cacao</i>)	245
	The <i>T. cacao</i> genome	247
	Genomics in anthropology	250
	The Neanderthal genome	250
	Ancient populations and migrations	251
	Genomics and language	258
	Recommended Reading	260
	Exercises, Problems, and Weblems	261
9	Microarrays and Transcriptomics	265
	Introduction	266
	Applications of DNA microarrays	268
	Analysis of microarray data	269

Expression patterns in different physiological states	272
The diauxic shift in <i>Saccharomyces cerevisiae</i>	272
Sleep in rats and fruit flies	274
Expression pattern changes in development	276
Variation of expression patterns during the life cycle of <i>Drosophila melanogaster</i>	276
Flower formation in roses	278
Expression patterns in learning and memory: long-term potentiation	282
Conserved clusters of co-expressing genes	284
Evolutionary changes in expression patterns	285
Applications of microarrays in medicine*	287
Development of antibiotic resistance in bacteria	287
Childhood leukaemias	290
Whole transcriptome shotgun sequencing: RNA-seq	292
Recommended Reading	293
Exercises, Problems, and Weblems	293
10 Proteomics	297
Introduction	298
Protein nature and types	298
Protein structure	299
The chemical structure of proteins	299
Conformation of the polypeptide chain	300
Protein folding patterns	301
Post-translational modifications	304
Why is there a common genetic code with 20 canonical amino acids?	306
Separation and analysis of proteins	307
Polyacrylamide gel electrophoresis (PAGE)	307
Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE)	307
Mass spectrometry	308
Classification of protein structures	310
SCOP	312
Changes in folding patterns in protein evolution	313
Many proteins change conformation as part of the mechanism of their function	314
Conformational change during enzymatic catalysis	314
Motor proteins	316
Allosteric regulation of protein function	318
Conformational states of serine protease inhibitors (serpins)	320
Protein structure prediction and modelling	322
Homology modelling	323
Available protocols for protein structure prediction	327
Structural genomics	328

Directed evolution and protein design	328
Directed evolution of subtilisin E	329
Enzyme design	330
Protein complexes and aggregates	330
Protein aggregation diseases*	331
Properties of protein–protein complexes	332
Multisubunit proteins	335
Recommended Reading	336
Exercises, Problems, and Weblems	336
11 Systems Biology	341
Introduction to systems biology	342
Two parallel networks: physical and logical	342
Statics and dynamics of networks	343
Pictures of networks as graphs	344
Trees	345
Sources of ideas for systems biology	346
Complexity of sequences	346
Shannon's definition of entropy	347
Randomness of sequences	348
Static and dynamic complexity	349
Computational complexity	350
The metabolome	351
Classification and assignment of protein function	351
Metabolic networks	354
Databases of metabolic pathways	354
Methionine synthesis in <i>Escherichia coli</i>	355
The Kyoto Encyclopedia of Genes and Genomes (KEGG)	356
Evolution and phylogeny of metabolic pathways	357
Carbohydrate metabolism in archaea	358
Reconstruction of metabolic networks	360
Regulatory networks	361
Signal transduction and transcriptional control	362
Structures of regulatory networks	362
Dynamics, stability, and robustness	363
Robustness through redundancy	363
Dynamic modelling	365
Protein interaction networks	366
Structural biology of regulatory networks	370
Protein–DNA interactions	373
Structural themes in protein–DNA binding and sequence recognition	373
An album of transcription regulators	374

Gene regulation	376
The transcriptional regulatory network of <i>Escherichia coli</i>	376
Regulation of the lactose operon in <i>E. coli</i>	380
The genetic regulatory network of <i>Saccharomyces cerevisiae</i>	382
Adaptability of the yeast regulatory network	383
Recommended Reading	384
Exercises, Problems, and Weblems	385
Epilogue	389
Index	390