

Contents

1	Genomes	1
1.1	The Ring of Life	2
1.2	Genome Structure	4
1.3	Genome Size	7
1.4	The Genomes of Modern and Archaic Humans	10
	References	14
2	Organelle Genomes and Endosymbionts	21
2.1	Mitochondria	22
2.1.1	Origin and Evolution of Mitochondria	22
2.1.2	Animal Mitochondrial Genomes	25
2.1.3	Mitochondrial Genomes of Plants and Algae	26
2.1.4	Mitochondrial Genomes of «Other» Eukaryotes	28
2.2	Plastids	29
2.2.1	Origin and Evolution of Plastids	29
2.2.2	Plastid Genomes	31
2.2.3	Plastids in the Amoeba <i>Paulinella chromatophora</i>	32
2.3	Heritable Bacterial Endosymbionts	33
2.3.1	Primary Endosymbionts	33
2.3.2	Secondary Endosymbionts	35
2.4	DNA Barcoding	35
	References	37
3	Sequencing Techniques	43
3.1	Sanger Sequencing	44
3.2	454 Pyrosequencing	45
3.3	Reversible Terminator Sequencing (Illumina)	47
3.4	Ion Semiconductor Sequencing (Ion Torrent)	49
3.5	Single-Molecule Real-Time (SMRT) Sequencing (PacBio)	51
3.6	Nanopore Sequencing	53
3.7	Comparison of Sequencing Platforms	55
	References	57
4	Sequencing Strategies	61
4.1	Shotgun Sequencing	62
4.2	RADseq	67
4.3	Hybrid Enrichment	70
4.4	Expressed Sequence Tags and RNA-Seq	73
4.5	Single-Cell Genomics and Transcriptomics	75
	References	75

5	Assembly and Data Quality	81
5.1	Data Quality and Filtering	82
5.2	Assembly Strategies	84
5.2.1	Greedy Assemblies	87
5.2.2	Overlap-Layout-Consensus (OLC) Assemblies	88
5.2.3	K-mer Assemblies Using de Bruijn Graphs	90
5.3	Comparing Assemblies	94
5.4	De Novo Assembly of Genomes	96
5.4.1	Scaffolding	96
5.4.2	Hybrid Assemblies	97
5.5	De Novo Assembly of Transcriptomes and Metagenomes	97
	References	100
6	Alignment and Mapping	105
6.1	Pairwise Alignment	106
6.2	Local Alignment and BLAST Searches	111
6.3	Multiple Sequence Alignment	114
6.4	Alignment Masking	115
6.5	Mapping Sequence Reads	117
6.6	Whole-Genome Alignments	121
	References	122
7	Finding Genes	127
7.1	What Is a Gene?	128
7.2	Gene Gain and Loss	128
7.3	Homology of Genes	130
7.4	Inferring Orthology	131
7.5	Hidden Markov Profiles	133
7.6	Gene Ontology and the Ortholog Conjecture	136
7.7	Whole-Genome Duplications	138
	References	139
8	Phylogenetic Analyses	143
8.1	Trees	144
8.2	Models of Nucleotide Substitution	147
8.3	Models of Amino Acid Substitutions	152
8.4	Model Selection and Data Partitions	155
8.4.1	Model Selection	155
8.4.2	Partition Finding	157
8.5	Inferring Phylogenies	158
8.5.1	Neighbour Joining	158
8.5.2	Maximum Parsimony	159
8.5.3	Maximum Likelihood	160
8.5.4	Heuristic Methods and Genetic Algorithms	162
8.5.5	Bayesian Inference	163
8.6	Support Measures	165
8.7	Molecular Clocks	166
	References	168

9	Sources of Error and Incongruence in Phylogenomic Analyses	173
9.1	Incongruence in Phylogenomic Analyses	174
9.2	Systematic Errors	177
9.3	Missing Data, Phylogenetic Information Content and Taxon Sampling	180
9.3.1	Missing Data	180
9.3.2	More Genes or More Taxa?	182
9.3.3	Taxon Sampling	182
9.3.4	Gene Sampling	183
9.4	Incongruence Between Gene Trees and Species Trees	186
	References	189
10	Rare Genomic Changes	195
10.1	The Perfect Phylogenetic Marker	196
10.2	Mobile Elements	198
10.3	MicroRNAs	201
10.4	Introns	202
10.5	Gene Order	203
10.6	Changes in the Genetic Code	206
	References	207
	Service Part	
	Glossary	214
	Index	219