
Contents

Part I Basic Processes of Genome Evolution

1 Basic Metabolism Surrounding DNAs	3
1.1 Central Dogma of Molecular Biology	3
1.2 DNA Replication.....	4
1.2.1 Chemical Nature of DNA Molecules	4
1.2.2 DNA Replication Systems	6
1.3 Transcription	9
1.3.1 RNA	9
1.3.2 Messenger RNA (mRNA).....	10
1.3.3 Transcription Machinery and Splicing.....	10
1.3.4 RNA Editing	12
1.4 Translation and Genetic Codes	12
1.4.1 Transfer RNA (tRNA).....	12
1.4.2 Genetic Codes	13
1.4.3 Ribosome and Translation.....	15
1.5 Proteins	16
1.5.1 Amino Acids	16
1.5.2 Four Strata of the Protein Structure	16
1.5.3 Protein Diversity	19
1.5.4 Protein Domains and Motifs	21
1.6 Genes.....	22
1.6.1 Protein Coding Genes	22
1.6.2 Chromosomes	24
References	24
2 Mutation.....	27
2.1 Classification of Mutations	27
2.1.1 What Is Mutation?.....	27
2.1.2 Temporal Unit of Mutation	28
2.1.3 Mutations Affecting Small Regions of DNA Sequences	28

2.1.4	Mutations Affecting Large Regions of DNA Sequences	29
2.2	Nucleotide Substitutions	30
2.2.1	Basic Characteristics of Nucleotide Substitutions	30
2.2.2	Nucleotide Substitutions in Protein Coding Regions.....	33
2.2.3	Methylation Creates “Fifth” Nucleotide	34
2.3	Insertion and Deletion.....	34
2.3.1	Basic Characteristics of Insertions and Deletions.....	34
2.3.2	Insertions and Deletions of Unique Sequences.....	35
2.3.3	Insertions and Deletions of Repeat Sequences	36
2.4	Recombination and Gene Conversion.....	39
2.4.1	Nature of Gene Conversion.....	40
2.4.2	Gene Conversion on Duplog Pairs	41
2.5	Gene Duplication	42
2.5.1	Classification of Duplication Events.....	42
2.5.2	Drift Duplication	43
2.5.3	Genome Duplication	44
2.6	Estimation of Mutation Rates	44
2.6.1	Direct and Indirect Method.....	44
2.6.2	Direct Method	45
2.6.3	Indirect Methods	47
2.6.4	Mutation Rates of Bacteria, Organelle, and Viruses	48
2.7	Mutation Affecting Phenotypes	49
2.7.1	What Is Phenotype?	49
2.7.2	Mutations in Protein Coding Region	49
2.7.3	Mutations in Noncoding Region	51
	References.....	52
3	Phylogeny	55
3.1	DNA Replications Generate Phylogenies	55
3.2	Genealogy of Individuals	57
3.2.1	Genealogy of Haploid Individuals	57
3.2.2	Genealogy of Diploid Individuals.....	58
3.2.3	Paternal or Maternal Genealogy of Diploid Individuals.....	59
3.3	Gene Genealogy.....	61
3.3.1	When a Gene Tree Is Identical to a Genome Tree	61
3.3.2	Gene Phylogenies in Diploids.....	62
3.3.3	Three Layers of Gene Phylogenies	64
3.3.4	Orthology and Paralogy	66
3.3.5	Gene Conversions May Twist the Gene Tree.....	68
3.3.6	Horizontal Gene Transfer and Xenology	70
3.4	Species Phylogeny	71
3.4.1	Two Layers of Species Trees	71
3.4.2	Gene Tree as the Approximation of Species Tree.....	72
3.4.3	Gene Tree May Be Different from Species Tree.....	73
3.5	Basic Concepts of Trees and Networks	74

3.5.1	Mathematical Definition	74
3.5.2	The Number of Possible Tree Topologies	76
3.5.3	How to Describe Trees and Networks.....	78
3.5.4	Static Networks and Dynamic Trees.....	80
3.5.5	Relationship Between Trees.....	81
3.6	Biological Implications of Trees and Networks.....	83
3.6.1	Fission and Fusion of Species and Populations	83
3.6.2	Importance of Branch Length	84
3.6.3	Trees and Taxonomy	85
	References.....	86
4	Neutral Evolution.....	89
4.1	Neutral Evolution as Default Process of the Genome Changes.....	89
4.1.1	Our World Is Finite	90
4.1.2	Unit of Evolution	90
4.2	How to Describe the Random Nature of DNA Propagation.....	91
4.2.1	Gene Genealogy Versus Allele Frequency Change.....	91
4.2.2	Branching Process.....	94
4.2.3	Coalescent Process.....	97
4.2.4	Markov Process.....	102
4.2.5	Diffusion Process	103
4.2.6	A More Realistic Process of Allele Frequency Change of Selectively Neutral Situation	106
4.3	Expected Evolutionary Patterns Under Neutrality	106
4.3.1	Fixation Probability	106
4.3.2	Rate of Evolution	107
4.3.3	Amount of DNA Variation Kept in Population	108
4.4	DNA Polymorphism	109
4.4.1	Single Nucleotide Polymorphism (SNP)	109
4.4.2	Insertions and Deletions (Indel).....	110
4.4.3	Repeat Number Polymorphism.....	110
4.4.4	Copy Number Variation	110
4.5	Mutation Is the Major Player of Evolution	110
4.6	Evolutionary Rate Under the Neutral Evolution.....	111
4.6.1	Molecular Clock.....	112
4.6.2	Heterogeneous Evolutionary Rates Among Proteins	114
4.6.3	Heterogeneous Evolutionary Rates Among Protein Parts	115
4.6.4	Heterogeneous Evolutionary Rates Among Organisms	116
4.6.5	Unit of Evolutionary Rate	118
4.7	Various Features of Neutral Evolution.....	118
4.7.1	Synonymous and Nonsynonymous Substitutions	118
4.7.2	Junk DNA	119
4.7.3	Pseudogenes.....	120
4.7.4	Neutral Evolution at the Macroscopic Level	122
	References.....	122

5	Natural Selection	125
5.1	Basic Concept of Natural Selection	125
5.2	Various Types of Positive Selection	127
5.2.1	Balancing Selection	127
5.2.2	Arms Race Between Hosts and Parasites.....	129
5.3	Natural Selection on Populations with a Large Number of Individuals	129
5.3.1	A Series of Simplifications in Many Theoretical Models.....	129
5.3.2	The Case for Haploids	130
5.3.3	The Case for Diploids	132
5.3.4	Overdominant Selection.....	134
5.4	Natural Selection on Populations with Finite Number of Individuals	135
5.4.1	Gene Genealogy Under Natural Selection.....	135
5.4.2	Diffusion Approximation with Natural Selection	136
5.4.3	Fixation Probability of a Mutant.....	137
5.4.4	Slightly Deleterious and Nearly Neutral Mutations.....	139
5.4.5	Compensatory Mutations	139
5.5	Natural Selection at the Genomic Level	140
5.5.1	Selective Sweep and Background Selection	140
5.5.2	Gain and Loss of Genes	141
5.5.3	Purifying Selection at Synonymous Sites	142
5.5.4	Purifying Selection at Noncoding Regions.....	143
5.5.5	Positive Selection for Ape and Human Genes	144
5.5.6	Detection of Positive Selection Through Genome-Wide Searches.....	146
	References.....	148

Part II Evolving Genomes

6	A Brief History of Life	157
6.1	Origin of Life	157
6.2	Evolution of the Genetic Code.....	160
6.3	Establishment of Cell	162
6.4	Emergence of Eukaryotes	163
6.5	Multicellularization.....	163
6.6	From Origin of Chordates to Emergence of Modern Human	167
	References	171
7	Prokaryote Genomes	175
7.1	Diversity of Prokaryotes and Their Genome Sequencing Efforts.....	175
7.2	The Basic Genome Structure of Prokaryotes.....	177
7.3	GC content and Oligonucleotide Heterogeneity	180
7.4	Horizontal Gene Transfer.....	182

7.5	Codon Usage.....	185
7.6	Metagenomes	187
	References.....	188
8	Eukaryote Genomes.....	193
8.1	Major Differences Between Prokaryote and Eukaryote genomes	193
8.2	Organelle Genomes.....	194
8.2.1	Mitochondria.....	196
8.2.2	Chloroplasts	198
8.2.3	Interaction Between Nuclear and Organelle Genomes	199
8.3	Intron.....	199
8.3.1	Classification of Intron.....	199
8.3.2	Introns Early/Late Controversy.....	200
8.3.3	Functional Regions in Introns.....	201
8.4	Junk DNAs.....	201
8.4.1	Dispersed Repeats.....	201
8.4.2	Tandem Repeats	202
8.4.3	Pseudogenes.....	203
8.4.4	Junk RNAs and Junk Proteins.....	203
8.5	Evolution of Eukaryote Genomes.....	204
8.5.1	Genome Duplication	204
8.5.2	RNA Editing	205
8.5.3	C-Value Paradox	206
8.5.4	Conserved Noncoding Regions.....	206
8.5.5	Mutation Rate and Genome Size	207
8.6	Genome of Multicellular Eukaryotes.....	209
8.6.1	Plant Genomes	209
8.6.2	Fungi Genomes	211
8.6.3	Animal Genomes	212
8.7	Eukaryote Virus Genomes.....	217
	References.....	218
9	Vertebrate Genomes.....	223
9.1	Characteristics of Vertebrate Genomes	223
9.2	Two-Round Genome Duplications.....	225
9.3	Features of Protein Coding Genes	227
9.3.1	Genes Shared with Other Taxonomic Groups.....	227
9.3.2	Genes Specific to Vertebrates.....	228
9.3.3	Olfactory Multigene Families	229
9.3.4	Selective Constraints in the Protein Coding Sequences...	230
9.4	Conserved Noncoding Regions.....	231
9.4.1	Evolutionary Rates of Noncoding Sequences Just Upstream of Protein Coding Regions	232
9.4.2	Highly Conserved Noncoding Sequences.....	232

9.4.3	Conserved SINEs	233
9.4.4	Lineage-Specific Conserved Noncoding Sequences.....	234
9.5	Isochores	234
9.6	Genomes of Specific Vertebrate Lineages.....	238
9.6.1	Teleost Fish	239
9.6.2	Amphibians and Amniotes	239
9.6.3	Mammals.....	240
9.7	Ancient Genomes	243
	References.....	243
10	Human Genome	249
10.1	Overview of the Human Genome	249
10.2	Protein Coding Genes in the Human Genome.....	251
10.3	RNA Coding Genes and Gene Expression Control Regions in the Human Genome	252
10.4	Personal Genomes.....	254
10.5	Genomic Heterogeneity of the Human Genome.....	256
10.6	Genetic Changes That Made Us Human.....	257
10.7	Ancient Human Genomes	259
	References.....	259
 Part III Methods for Evolutionary Genomics		
11	Genome Sequencing.....	265
11.1	DNA Extraction and Purification	265
11.1.1	Special Treatments for Various Samples.....	265
11.1.2	How to Extract Genomic DNA	266
11.2	Construction of Genomic Library	266
11.3	Determination of Nucleotide Sequences: Chemical Tactics	269
11.3.1	General Overview	269
11.3.2	Sanger's Method	270
11.3.3	Sequencing by Synthesis.....	272
11.3.4	Physical Distinguishing of Each Nucleotide.....	273
11.4	Determination of Nucleotide Sequences: Computational Tactics	274
11.4.1	Base Call.....	274
11.4.2	Shotgun Sequencing.....	274
11.4.3	Minimum Tiling Array	275
11.4.4	Haplotype Sequence Determination	276
11.4.5	Resequencing	277
	References.....	278
12	Omic Data Collection.....	281
12.1	Overview of Omic Worlds	281
12.2	Transcriptome	281
12.3	Proteome	283
12.4	Other Omic Data	284

12.4.1	Metabolome	284
12.4.2	Phenome.....	285
12.4.3	Ecome	286
References.....		287
13	Databases	289
13.1	Overview of Databases	289
13.2	Genome Sequence Databases	291
13.3	INSDC Databases	292
13.4	Protein-Related Databases	295
13.5	Literature Databases.....	295
13.6	Other Databases	298
References.....		299
14	Sequence Homology Handling	301
14.1	What Is Homology?	301
14.2	Homology Search.....	302
14.2.1	BLAST Families	302
14.2.2	Examples of Nucleotide Sequence Search Using BLASTN	304
14.2.3	Examples of Amino Acid Sequence Search Using BLAST Families.....	308
14.2.4	Other Homology Search Methods	309
14.2.5	Problem of Homology Search-Dependent Analyses.....	310
14.3	Pairwise Alignment.....	311
14.3.1	Biologically True Alignment.....	311
14.3.2	Mathematically Optimum Alignment	312
14.3.3	Methods Using Dynamic Programming	313
14.3.4	Dot Matrix Method	317
14.4	Multiple Alignment.....	319
14.4.1	Overview	320
14.4.2	CLUSTAL W	320
14.4.3	MAFFT	321
14.4.4	Example of Amino Acid Sequence Alignment	322
14.4.5	MISHIMA.....	322
14.4.6	Problem of Evaluation of Multiple Alignment Results.....	328
14.4.7	Comparison of Major Multiple Alignment Softwares.....	329
14.5	Genome-Wide Homology Viewers	330
References.....		332
15	Evolutionary Distances	335
15.1	Overview of Evolutionary Distances	335
15.2	Nucleotide Substitutions	337
15.2.1	Nucleotide Difference and Nucleotide Substitution	337

15.2.2	Nucleotide Substitution Matrix.....	339
15.2.3	One-Parameter Method.....	340
15.2.4	Two-Parameter Method.....	345
15.2.5	Methods Incorporating Observed Nucleotide Frequencies	349
15.2.6	Other Methods	351
15.2.7	Handling of Heterogeneity Among Sites	353
15.3	Synonymous and Nonsynonymous Substitutions	355
15.3.1	Estimations of the Numbers of Synonymous and Nonsynonymous Sites	355
15.3.2	Weightings for Multiple Paths	357
15.3.3	Correction of Multiple Hits.....	357
15.4	Amino Acid Substitutions.....	360
15.4.1	Poisson Correction	360
15.4.2	Dayhoff Matrix and Its Descendants	361
15.5	Evolutionary Distances Not Based on Substitutions.....	363
	References.....	363
16	Tree and Network Building.....	367
16.1	Classification of Tree-Building Methods	367
16.1.1	Classification by Type of Data	367
16.1.2	Classification by Type of Tree Search Algorithm	370
16.1.3	Character-State Data	370
16.2	Distance Matrix Methods.....	371
16.2.1	UPGMA and WPGMA	371
16.2.2	Minimum Deviation Method and Related Methods	373
16.2.3	Minimum Evolution Method and Related Methods	375
16.2.4	Transformed Distance Methods	377
16.2.5	Use of Quartet OTUs for Tree Construction	378
16.3	Neighbor-Joining Method.....	380
16.3.1	What Are Neighbors?.....	380
16.3.2	Algorithm of the Neighbor-Joining Method	380
16.3.3	A Worked-Out Example of Tree Construction Using the Neighbor-Joining Method.....	384
16.3.4	Methods Related to the Neighbor-Joining Method.....	385
16.4	Phylogenetic Network Construction from Distance Matrix Data	387
16.5	Maximum Parsimony Method	389
16.5.1	The Basic Algorithm of the Maximum Parsimony Method	390
16.5.2	Stepwise Clustering Algorithms for Constructing Maximum Parsimony Trees	393
16.5.3	Compatibility Method.....	396
16.5.4	Theoretical Problems of the Maximum Parsimony Method	396
16.6	The Maximum Likelihood Method.....	397

16.6.1	The Principle of the Maximum Likelihood Method	397
16.6.2	The Basic Algorithm of the Maximum Likelihood Method.....	399
16.6.3	Stepwise Clustering Algorithm for the Maximum Likelihood Method.....	401
16.6.4	Algorithm of the Bayesian Method	403
16.6.5	Theoretical Problems of ML and Bayesian Methods	403
16.7	Phylogenetic Network Construction from Character-State Data	405
16.8	Tree-Searching Algorithms.....	406
16.9	Comparison of Phylogenetic Tree-Making Methods.....	407
16.10	Phylogeny Construction Without Multiple Alignment	409
	References.....	410
17	Population Genomics	417
17.1	Evolutionary Distances Between Populations	417
17.1.1	Distance Between Populations Based on Gene Genealogy	417
17.1.2	Distance Between Species Based on Gene Genealogy.....	420
17.1.3	Distance Between Populations Based on Allele Frequency Differences	421
17.1.4	Evolutionary Distance Between Genomes	424
17.2	Mitochondrial DNA Population Genomics	424
17.2.1	Inference of Gene Genealogy	424
17.2.2	Population Size Fluctuation.....	425
17.2.3	Estimation of Nucleotide Substitution Patterns	427
17.3	Population Genomics of Prokaryotes	428
17.4	Population Genomics of Nuclear Genomes.....	430
17.4.1	Relationship of Individuals and Populations	430
17.4.2	Admixture	433
17.4.3	Introgression	435
17.4.4	Population Size Fluctuation.....	437
17.4.5	Genome-Wide Association Study	438
	References.....	439
	Index.....	443

Introduction to Evolutionary Genomics

Saitou, N.

2013, XXIII, 461 p. 227 illus., Hardcover

ISBN: 978-1-4471-5303-0