

Table of Contents

Introduction 1

CHAPTER 1 / The Molecular Basis of Biology and Evolution 5

Nucleotide Sequences 5

Genomes 7

Genome constituents 7

Somatic genome processing 7

DNA Replication 8

Transcription and Posttranscriptional Modifications of RNA 9

Genes 11

Protein-coding genes 12

RNA-specifying genes 14

Nontranscribed genes 15

Pseudogenes 16

Amino Acids 16

Proteins 19

Translation and Genetic Codes 20

Information Flow among DNA, RNA, and Proteins 24

Mutation 24

Classification of mutations 25

Point Mutations 26

Segmental Mutations 30

Recombination 30

Deletions and insertions 31

Inversions 34

Spatial distribution of mutations 34

Are Mutations Random? 34

CHAPTER 2 / Allele Dynamics in Populations 35

Standing Genetic Variation 35

Gene diversity 36

Nucleotide diversity 37

Structural variation 38

What Is Evolution? 38

Changes in Allele Frequencies 39

Selection 40

Codominance 42

Dominance and recessiveness 43

Overdominance and underdominance 44

Random Genetic Drift 47

Census Population Size and Effective Population Size 49

Short-term effective population size 50

Coalescence and long-term effective population size 51

Factors conspiring to reduce the effective population size relative to the census population size 54

Gene Substitution 55

Fixation probability 55

Fixation time 56

Rate of gene substitution 58

Mutational meltdown: The double jeopardy of small populations 58

Nearly neutral mutations 58

Second-Order Selection 60

The evolution of mutation rates 60

The evolution of mutational robustness 62

Violations of Mendel's Laws of Inheritance 63

Transmission Ratio Distortion 63

Segregation distortion 64

Postsegregation distortion 65

Converting elements 65

Sex allocation distortion 66

Autonomous replicating elements 66

Linkage Equilibrium and Disequilibrium 66

Hitchhiking and Selective Sweep 67

Molecular signatures of selective sweeps 68

The evolution of lactase persistence in Africa and Europe 70

Background Selection 71

Epistasis 71

The Driving Forces in Evolution 72

The neo-Darwinian theory and the neutral mutation hypothesis 72

The distribution of fitness effects 74

A test of neutrality based on genetic polymorphism 75

Consequences of Explosive Population Growth: Single-Nucleotide Variation in Humans 76

CHAPTER 3 / DNA and Amino Acid Sequence Evolution 79

Nucleotide Substitution in a DNA Sequence 79

Jukes and Cantor's one-parameter model 80

Kimura's two-parameter model 82

Number of Nucleotide Substitutions between Two DNA Sequences 83

Number of substitutions between two noncoding sequences 84

Substitution schemes with more than two parameters 86

Violation of assumptions 87

Saturation 88

Number of Substitutions between Two Protein-Coding Genes 88

Number of Amino Acid Replacements between Two Proteins 93

Alignment of Nucleotide and Amino Acid Sequences 93

Pairwise alignment 94

Manual alignment 95

The dot matrix method 95

Scoring matrices and gap penalties 97

Alignment algorithms 100

Multiple-sequence alignment 101

Quality of alignments 104

Alignment of Genomic Sequences 106

CHAPTER 4 / Rates and Patterns of Molecular Evolution 107

Rates of Point Mutation 107

Rates of Segmental Mutations 110

Rates of Nucleotide Substitution 111

Rates of substitution in protein-coding sequences 111

Rates of substitution in noncoding regions 115

Causes of Variation in Substitution Rates 116

The concept of functional constraint 116

Quantifying the degree of protein tolerance toward amino acid replacements 116

Synonymous versus nonsynonymous rates 117

Variation among different gene regions 117

Variation among genes 119

Variables associated with protein evolutionary rates 120

Evolutionary conservation and disease 121

Relaxation of selection 122

Selective intolerance toward indels 123

Identifying positive and purifying selection 123

Estimating the intensity of purifying selection 124

Are slowly evolving regions always important? 125

Male-Driven Evolution: Mutational Input and Slow-X Evolution 126

Rates of Evolution under Positive Selection 128

Prevalence of positive selection 129

Fast-X evolution 130

Rates of Evolution under Balancing Selection 130

Patterns of Substitution and Replacement 130

Patterns of spontaneous mutation 131

Patterns of mutation and strand asymmetry 134

Clustered multinucleotide substitutions: Positive selection or nonrandomness of mutation? 135

Patterns of amino acid replacement 137

What protein properties are conserved in protein evolution? 138

Heterotachy 139

Nonrandom Usage of Synonymous Codons 139

Measures of codon usage bias 140

Species-specific and universal patterns of codon usage 141

Determinants of Codon Usage 142

Interspecific variation in codon usage and amino acid usage 142

Intragenomic variation in codon usage 142

Translational efficiency and translation accuracy 143

The tRNA adaptation index 145

Intragenic variation in codon usage 147

- Indirect selection on codon usage 148
- Why do only some organisms have biased codon usages? 148
- Codon usage in unicellular and multicellular organisms 148
- Codon usage and population size 149

Causes of Variation in Substitution Rates among Evolutionary Lineages 157

- The DNA repair hypothesis 158
- The generation-time effect hypothesis
- The metabolic rate hypothesis 159
- The varying-selection hypothesis 159

Are Living Fossils Molecular Fossils Too? 160

Phyletic Gradualism, Punctuated Equilibria, and Episodic Molecular Evolution 160

Rates of Substitution in Organelle DNA 161

- Mitochondrial rates of evolution 161
- Plastid rates of evolution 162
- Substitution and rearrangement rates 162

Rates of Substitution in Viruses 163

- Human immunodeficiency viruses 163

CHAPTER 5 / Molecular Phylogenetics and Phylogenetic Trees 165

Impacts of Molecular Data on Phylogenetic Studies 165

Advantages of Molecular Data in Phylogenetic Studies 167

Species and Speciation 167

- The species concept 167
- Speciation 168
- Terminology 170

Phylogenetic Trees 170

- Rooted and unrooted trees 171
- Scaled and unscaled trees 172
- The Newick format 173
- Number of possible phylogenetic trees 174
- Tree balance 175
- True and inferred trees 177
- Gene trees and species trees 177
- Taxa and clades 178

Types of Molecular Homology 179

Types of Data 180

- Character data 180
- Assumptions about character evolution 181
- Polarity and taxonomic distribution of character states 182
- Distance data 183

Methods of Tree Reconstruction 184

Distance Matrix Methods 184

- Unweighted pair-group method with arithmetic means (UPGMA) 184
- Sattath and Tversky's neighbors-relation method 186
- Saitou and Nei's neighbor-joining method 187

Maximum Parsimony Methods 187

- Weighted and unweighted parsimony 191
- Searching for the maximum parsimony tree 191

Maximum Likelihood Methods 194

Bayesian Phylogenetics 197

Topological Comparisons 198

- Topological distance 199
- Consensus trees 199
- Supertrees 200

Rooting Unrooted Trees 201

- Outgroup rooting 202
- Midpoint rooting 202

Estimating Branch Lengths 204

Calibrating Phylogenetic Trees and Estimating Divergence Times 205

Assessing Tree Reliability 207

- The bootstrap 208
- Tests for two competing trees 209

Problems Associated with Phylogenetic Reconstruction 211

- Strengths and weaknesses of different methods 211
- Minimizing error in phylogenetic analysis 212

Genome Trees 214

- Genome trees based on shared gene content 214
- Genome trees from BLASTology 214

Molecular Phylogenetic Examples 214

- Phylogeny of apes 215
- The utility of polarized character states: *Cetartiodactyla* and SINE phylogeny 220

Molecular Phylogenetic Archeology 222

The disextinction of the quagga 224

The dusky seaside sparrow: A lesson in conservation biology 225

Molecular Phylogenetics and the Law 227

At the Limits of the Tree Metaphor: The Phylogeny of Eukaryotes and the Origin of Organelles 228

The phylogeny of eukaryotes 228

Origin of organelles 230

Phylogenetic Trees as a Means to an End 232

Parallelism and convergence as signifiers of positive selection 232

Detecting amino acid sites under positive selection 233

Reconstructing ancestral proteins and inferring paleoenvironments 234

Mapping nonmolecular characters onto molecular trees 234

CHAPTER 6 / Reticulate Evolution and Phylogenetic Networks 237

Networks 237

Phylogenetic and Phylogenomic Networks 238

The median network method 239

The conditioned-reconstruction method 240

Inferred reticulations: Are they real? 243

Examples of Real-Life Phylogenetic Networks 243

Reticulate evolution by recombination: A resurrected blood-group allele in humans 244

Speciation by hybridization: The reticulate evolution of woodferns 246

The Tree of Life Hypothesis 247

The Vertical and Horizontal Components of Prokaryote Evolution 249

Prokaryote taxonomy and the meaning of "species" in prokaryotes 250

The Phylogeny of Everything 253

The eukaryote-prokaryote divide and the taxonomic validity of Prokaryota 253

The Eubacteria-Archaeobacteria divide 253

The tripartite tree of life and its inadequacy 255

The Origin of Eukaryotes 257

The gradual origin hypothesis 258

The fateful encounter hypothesis 259

Eukaryotes as an "organizational upgrade" 262

The nonrandom origin of operational and informational genes in eukaryotes 263

Why genes in pieces? The origin of the nuclear membrane 264

All complex life is eukaryotic: The energetics of gene expression 266

The eukaryotic cell as a one-off innovation and a possible solution to the Fermi paradox 268

Archaeobacterial Systematics: Clade-Specific Archaeobacterial Genes and Clade-Specific Horizontal Gene Imports from Eubacteria 269

The Two Primary Domains of Life 271

The Public Goods Hypothesis 271

CHAPTER 7 / Evolution by DNA Duplication 273

Types of DNA Duplication 274

Mechanisms of DNA Duplication 274

Dating Duplications 275

Gene Duplication and Gene Families 276

The Prevalence of Gene Duplication 278

Modes of Evolution of Multigene Families 278

Divergent Evolution of Duplicated Genes 279

Nonfunctionalization and gene loss 280

Nonfunctionalization time 281

Retention of original function following gene duplication 283

Evolution of rRNA-specifying genes 284

Neofunctionalization 285

Multifunctionality and subfunctionalization 287

Neosubfunctionalization 294

Rates of Evolution in Duplicated Genes 295

Rates and patterns of expression divergence between duplicated genes 295

Human Globins 297

Concerted Evolution 299

Unequal crossing over 301

Gene conversion 302

Examples of gene conversion 304

The relative roles of gene conversion and unequal crossing over 306

Factors Affecting Concerted Evolution 308

Number of repeats 308
Arrangement of repeats 308
Structure of the repeat unit 308
Functional requirements and selection 309
Population size 310

Evolutionary Implications of Concerted Evolution 310

Spread of advantageous mutations 311
Retardation of paralogous gene divergence 311
Generation of genic variation 311
Methodological pitfalls due to concerted evolution 311
Positive selection or biased gene conversion? The curious histories of *HAR1* and *FX1* 312

Birth-and-Death Evolution 314

Expansion and contraction of gene families 314
Examples of birth-and-death evolution 315
The death of gene families 325

Mixed Concerted Evolution and Birth-and-Death Evolution 325

Polysomy 326

Polyploidy 326
Diploidization 330
Distinguishing between gene duplication and genome duplication 331

CHAPTER 8 / Evolution by Molecular Tinkering 339

Protein Domains 339

Internal Gene Duplication 340

Properties and prevalence of internal gene duplication 343

Exon-Domain Correspondence 348

Mosaic Proteins 349

Exon Shuffling 351

Phase limitations on exon shuffling 352
Prevalence of domain shuffling and the evolutionary mobility of protein domains 353
Domain shuffling and protein-protein interaction networks 356

Gene Fusion and Fission 356

Domain Accretion 360

Strategies of Multidomain Gene Assembly 361

Evolution by Exonization and Pseudoexonization 362

Evolution of Overlapping Genes 364

Alternative Splicing 368

Sex determination and alternative splicing 369
Evolution of alternative splicing 370
Increasing proteome diversity: Alternative splicing or gene duplication? 372

De Novo Origination of Genes 373

Nested and Interleaved Genes 375

Gene Loss and Unitary Pseudogenes: A Molecular Revisiting of the "Law of Use and Disuse" 376

Functional Convergence 382

Origin and Evolution of Spliceosomal Introns 383

A Grand View of Molecular Tinkering: Suboptimality and Gratuitous Complexity 385

Tinkering in action: The patchwork approach to the evolution of novel metabolic pathways 386
Irremediable complexity by constructive neutral evolution 388

CHAPTER 9 / Mobile Elements in Evolution 391

Mobile Elements, Transposable Elements, and Transposition 391

Classification of Transposable Elements 393

Conservative and replicative transposition 393
DNA- and RNA-mediated transposition 394
Enzymatic classification of transposable elements 394
Autonomous and nonautonomous transposable elements 394
Active and fossil transposable elements 394
Taxonomic, developmental, and target-site specificity of transposition 394

DNA-Mediated Transposable Elements 395

Insertion sequences 395
Transposons 396
Nonautonomous DNA-mediated transposable elements 397

Retroelements 398

Retrons 398
TERT genes 399
Mitochondrial retroplasmids 399
Group II introns and twintrons 400

Retrotransposons 400
Retroviruses 401
Pararetroviruses 402
Evolutionary origin of retroelements 402
Nonautonomous and fossil retrotransposable elements 403

LINEs and SINEs 405

SINEs derived from 7SL RNA 405
SINEs derived from tRNAs and SINEs containing 5S rRNA 407
SINEs containing snRNA 408
Mosaic SINEs 408
Where there's a SINE, there's a LINE 408
Rate of SINEs evolution 410

Retrosequences 410

Retrogenes 411
Semiprocessed retrogenes 413
Retropseudogenes 413
Endogenous non-retroviral fossils 416

The "Ecology" of Transposable Elements 417

Transposable elements and the host genome: An evolutionary tug-of-war 417
Transposable elements and segregation distortion 418
Evolutionary dynamics of transposable-element copy number 419

Genetic and Evolutionary Effects of Transposition 420

Transposable elements as mutagens 420
Transposable elements and somatic mosaicism 424

The molecular domestication of transposable elements 424

Transposition and Speciation 430

Horizontal Gene Transfer 431

Telltale signs of horizontal gene transfer 431
Mechanisms of horizontal gene transfer among prokaryotes 432
Prevalence and limitations of horizontal gene transfer in prokaryotes 435
Genomic consequences of gene transfer among prokaryotes 437
Clinical consequences of gene transfer among prokaryotes 437

Horizontal Gene Transfer Involving Eukaryotes 438

Horizontal gene transfer from eukaryotes to prokaryotes 438
Horizontal gene transfer from prokaryotes to eukaryotes 438
Horizontal transfer among eukaryotes 440
Horizontal gene transfer among plants 441
Horizontal transfer of a functional gene from fungi to aphids 441
Horizontal transfer of transposable elements among animals 442

Promiscuous DNA 446

Transfer of intact functional genes to the nucleus 447
Transfer of nonfunctional DNA segments from organelles to the nucleus: *numts* and *nupts* 447
Rates and evolutionary impacts of norgDNA insertion 448

CHAPTER 10 / Prokaryotic Genome Evolution 451

Genome Size in Prokaryotes 452

The pangenome, the core genome, and the accessory genome 453
Increases and decreases in prokaryotic genome sizes 455

Genome Miniaturization 457

Genome size reduction in intracellular symbionts and parasites 457
The miniaturization of organelle genomes 459
The evolution of mitochondrial genome sizes 460
The evolution of plastid genome sizes 462

The Minimal Genome 463

The comparative genomics approach: Identifying the core genome of all life forms 464
Probabilistic reconstruction of gene content in the last universal ancestor of life 466
The experimental gene inactivation approach: Gene essentiality 466

GC Content in Prokaryotes 467

Possible explanations for variation in GC content 468
Chargaff's parity rules 470

GC Skew and Gene-Density Asymmetries Are Related to DNA Replication Biases 471

Replichores and chirochores 471
The location of genes in leading and lagging strands 474

Chromosomal Evolution in Prokaryotes 477

Evolution of chromosome number in prokaryotes 478
Estimating the number of gene order rearrangement events 480
Gene order evolution 483
Operon evolution 483

The Emergence of Alternative Genetic Codes 486

CHAPTER 11 / Eukaryotic Genome Evolution 491

Functionality and nonfunctionality in eukaryotic genomes 492

- What is “function” in an evolutionary context? 492
- What do genomes do? An evolutionary classification of genomic function 494
- Changes in functional affiliation 496
- Detecting functionality at the genome level 496

Phenotypic validation of positive selection 499

- What proportion of the human genome is functional? 503
- How much garbage DNA is in the human genome? 503

Genome Size, DNA Content, and C Value 505

- Genome size variation and genomic content in eukaryotes 505
- Intraspecific variation in genome size 507

Mutations That Increase or Decrease Genome Size 507

- The contribution of genome duplication to genome size 508
- The contribution of transposable elements to genome size 509
- Deletions and genome size 510

Genomic Paradoxes in Eukaryotes 511

- The C-value paradox 511
- Possible solutions to the C-value paradox 513
- Why so much of the genome is transcribed—or is it? 516

Life History and Cellular Correlates of Genome Size 517

- The nucleocytoplasmic ratio 518
- The coincidence hypothesis 519
- Nucleotypic hypotheses 519
- The nucleoskeletal hypothesis 520
- Is small genome size an adaptation to flight? 521

The C-Value Paradox: The Neutralist Hypothesis 522

- Selfish DNA 523
- The mutational hazard hypothesis 524
- Is it junk DNA or is it indifferent DNA? 526

Trends in Genome Size Evolution 527

- Is there an upper limit to genome size? 527
- Genome miniaturization in eukaryotes 528

Protein-Coding Gene Number Variation and the G-Value Paradox 532

- Possible solutions to the G-value paradox 534
- The I value 535

Gene Number Evolution 536

Methodologies for Studying Gene Repertoire Evolution 537

- Gene-family cluster analysis 538
- Functional clustering of proteins 539
- Supervised machine learning and the subcellular localization of proteins 541
- Gene ontology 542

Chromosome Number and Structure 544

- Chromosome number variation 544
- Chromosome morphology and chromosome types 545
- Chromosome size variation 546
- Euchromatin and heterochromatin 547

Chromosomal Evolution 548

- Chromosome number evolution 548
- Chromosomal rearrangements 551
- Evolutionary patterns of chromosomal rearrangements 555
- Is gene order conserved? 555

Gene Distribution Between and Within Chromosomes 556

- Gene density 556
- Do genes cluster by function? 557

The Repetitive Structure of the Eukaryotic Genome 558

- Tandemly repeated sequences 560
- Mutational processes affecting repeat-unit number in tandemly repeated DNA 562
- The contribution of tandem repeats to genome size 564
- Do tandemly repeated DNA sequences have a function? 564
- Centromeres as examples of indifferent DNA 565

Genome Compositional Architecture 565

- Segmentation algorithms and compositional domains 568
- Compositional architectures of mammalian nuclear genomes 570
- The origin and evolution of compositional domains 572

CHAPTER 12 / The Evolution of Gene Regulation 575

Pretranscriptional Regulation 576

- Regulation by covalent modifications of histones 576
- DNA methylation 576

Regulation at the Transcriptional Level 577

- Promoters 577
- Promoter evolution 580
- Divergent transcription 581
- Enhancers 582

- Shadow enhancers 585
- Insulators 591

Posttranscriptional Regulation 592

- RNA interference 593
- Patterns of evolution of miRNAs 594
- Do miRNAs have a deep evolutionary history? 595
- Does translational regulation contribute to phenotypic evolution? 595

CHAPTER 13 / Experimental Molecular Evolution 597

What Is Experimental Evolution? 598

- The basic design of evolutionary experiments 599
- How to measure fitness and changes in fitness in evolutionary experiments 600

The Contribution of Experimental Evolution to Evolutionary Biology 601

- Population divergence and the adaptive landscape metaphor 602

- Historical contingency 604
- Epistasis 607

Mutation Dynamics 608

- Neutral mutation rates 608
- Non-neutral mutation rates 608

Targets of Selection 610

Literature Cited LC-1

Index I-1