

Contents

Preface xiii

Contributors xvii

Part I Evolution of Life

1. Evolutionary Genomics Leads the Way 3

David Penny and Lesley J. Collins

- 1.1 Introduction 3
- 1.2 Evolution and the Power of Genomes 4
- 1.3 The Problem of Deep Phylogeny and "The Tree" 5
- 1.4 Fred, the Last Common Ancestor of Modern Eukaryotes 7
- 1.5 Eukaryote Origins: Continuity from the RNA World? 10
- 1.6 Minimal Genomes and Reductive Evolution 12
- 1.7 Evolutionary Genomics for the Future 13
- References 14

2. Current Approaches to Phylogenomic Reconstruction 17

Denis Baurain and Hervé Philippe

- 2.1 Phylogenomics and Supermatrices 17
- 2.2 Phylogenetic Signal Versus Nonphylogenetic Signal 19
- 2.3 Probabilistic Models and Nonphylogenetic Signal 22
 - 2.3.1 Homogeneous Models 22
 - 2.3.2 Handling of Rate Signal 23
 - 2.3.3 Handling of Compositional Signal 24

- 2.3.4 Other Model Violations 26
- 2.3.5 Future Developments 27

2.4 Reduction of Nonphylogenetic Signal Under Fixed Models 28

- 2.4.1 Variations in Taxon Sampling 28
- 2.4.2 Recoding and Removal of Offending Data 30

2.5 CAT Model 31

2.6 Case Study: Cambrian Explosion 33

2.7 Conclusion 35

References 36

3. The Universal Tree of Life and the Last Universal Cellular Ancestor: Revolution and Counterrevolutions 43

Patrick Forterre

- 3.1 Introduction 43
- 3.2 The Woesean Revolution 45
- 3.3 A Rampant "Prokaryotic" Counterrevolution 47
- 3.4 How to Polarize Characters Without a Robust Root? 50
- 3.5 The Hidden Root: When the Weather Became Cloudy 51
- 3.6 LUCA and Its Companions 54
- 3.7 The Problem of Horizontal Gene Transfer and Ancient Phylogenies: Trees Versus Gene Webs 54
- 3.8 The Nature of the RNA World 55
- 3.9 The DNA Replication Paradox and the Nature of LUCA 56
- 3.10 When Viruses Find Their Way into the Universal Tree of Life 58
- 3.11 Future Directions 59
- References 60

4. Eukaryote Evolution: The Importance of the Stem Group 63

Anthony M. Poole

- 4.1 Introduction 63
 - 4.1.1 What is Significant About the Origin of the Eukaryote Cell? 63
- 4.2 Interpreting Trees 68
- 4.3 Moving Beyond the Deep Roots of Eukaryotes 70
 - 4.3.1 Origin of the Mitochondrion 71
 - 4.3.2 Stepwise Development of Mitochondria 72
 - 4.3.3 Intron Proliferation and Eukaryote Origins 74
- 4.4 Concluding Remarks 76
- References 77

5. The Role of Information in Evolutionary Genomics of Bacteria 81

Antoine Danchin and Agnieszka Sekowska

- 5.1 Introduction 81
- 5.2 Revisiting Information 83
- 5.3 Ubiquitous Functions for Life 84
- 5.4 The Cenome and the Paleome 87
- 5.5 Functions Corresponding to Nonessential Persistent Genes 89
- 5.6 A Ubiquitous Information-Gaining Process: Making a Young Organism from an Aged One 89
- 5.7 Provisional Conclusion 91
- Acknowledgments 92
- References 92

6. Evolutionary Genomics of Yeasts 95

Bernard Dujon

- 6.1 Introduction 95
- 6.2 A Brief History of Hemiascomycetous Yeast Genomics 96
- 6.3 The Scientific Attractiveness of *S. cerevisiae* 98
 - 6.3.1 Functional Genomics 98
 - 6.3.2 Genome Duplication 102
 - 6.3.3 A Bunch of Fermentative Engines 102

- 6.3.4 Speciation and Species Definition 103

6.4 Evolutionary Genomics of Hemiascomycetes 104

- 6.4.1 Distinct and Specific Genome Organization of Three Major Evolutionary Subdivisions of Hemiascomycetes 104
- 6.4.2 Comparison of Proteins: Pan- and Core-Proteomes 106
- 6.4.3 Genome Redundancy and Paralogues 106
- 6.4.4 Conservation of Synteny 107
- 6.4.5 Genes for Noncoding RNAs, Introns, and Genetic Code Variation 108
- 6.4.6 Sex, Transposons, Plasmids, Inteins, and Horizontal Gene Transfer 109
- 6.4.7 Mitochondrial Genomes and NUMTs 110

6.5 Surprises 111

6.6 What Next? 113

Acknowledgments 115

Epilogue 115

References 115

Part II Evolution of Molecular Repertoires

7. Genotypes and Phenotypes in the Evolution of Molecules 123

Peter Schuster

- 7.1 The Landscape Paradigm 123
- 7.2 Molecular Phenotypes 125
 - 7.2.1 Protein Structures 126
 - 7.2.2 Nucleic Acid Structures 129
- 7.3 The RNA Model 132
 - 7.3.1 RNA Replication and Mutation 133
 - 7.3.2 RNA Secondary Structures 138
 - 7.3.3 Neutrality and Its Consequences 139
 - 7.3.4 Stochastic Effects in RNA Evolution 141
 - 7.3.5 Beyond the One Sequence–One Structure Paradigm 145
- 7.4 Conclusions and Outlook 148
- Acknowledgments 149
- References 149

8. Genome Evolution Studied Through Protein Structure 153

Philip E. Bourne, Kristine Briedis, Christopher Dupont, Ruben Valas, and Song Yang

- 8.1 Introduction 153
- 8.2 Structural Granularity and Its Implications 156
- 8.3 Protein Domains in the Study of Genome Rearrangements 158
- 8.4 Protein Domain Gain and Loss 160
- 8.5 And in the Beginning ... 161
- 8.6 But Let Us Not Forget the Influence of the Environment 161
- 8.7 Conclusions 162
- References 163

9. Chromosomal Rearrangements in Evolution 165

Hao Zhao and Guillaume Bourque

- 9.1 Introduction 165
- 9.2 Genome Representation 166
- 9.3 Constructing Genome Permutations from Sequence Data 167
- 9.4 Genomic Distances 168
 - 9.4.1 Model-Free Distances 169
 - 9.4.2 Rearrangement-Based Distances 170
- 9.5 Reconstruction of Ancestors and Evolutionary Scenarios 174
 - 9.5.1 Model-Free Reconstruction Algorithms 175
 - 9.5.2 Rearrangement-Based Reconstruction Algorithms 176
- 9.6 Recent Applications on Large Genomes 177
- 9.7 Challenges and Promising New Approaches 178
- Acknowledgment 179
- References 179

10. Molecular Structure and Evolution of Genomes 183

Todd A. Castoe, A. P. Jason de Koning, and David D. Pollock

- 10.1 Introduction 183
- 10.2 Overview of Considerations in Studying Protein Evolution 184
- 10.3 Function and Evolutionary Genomics 186
 - 10.3.1 Deciphering Complexities of Protein Evolution 186
 - 10.3.2 The Future of Modeling Protein Evolution: Merging Realism with Tractability 188
 - 10.3.3 The Effect of Increasing Taxon Sampling and Sequence Biodiversity 189
 - 10.3.4 Removing the Mutational Noise and Context-Dependent Biases from Protein Evolution 190
 - 10.3.5 Where is Protein Evolution Going? 191
 - 10.3.6 Detecting Adaptation and Functional Innovation 193
- 10.4 Integrating Inferences to Detect and Interpret Adaptation: An Example with Snake Metabolic Proteins 194
 - 10.4.1 Snake Metabolic Proteins—Integration of Inferences for Adaptation 194
 - 10.4.2 Detection of Accelerated Nonsynonymous Change 195
 - 10.4.3 Changes at Conserved Sites and Coevolutionary Signal 195
 - 10.4.4 Integrating Evolutionary Inferences with Structure and Function Information 197
 - 10.4.5 Further Evidence of Adaptation from Molecular Convergence 198
 - 10.4.6 Integrating Inferences with Possible Causal Factors 199
- 10.5 Conclusion 200
- References 200

11. The Evolution of Protein Material Costs	203	13. Domain Family Analyses to Understand Protein Function Evolution	231
<i>Jason G. Bragg and Andreas Wagner</i>		<i>Adam James Reid, Sarah Addou, Robert Rentzsch, Juan Ranea, and Christine Orengo</i>	
11.1 Introduction	203	13.1 Introduction	231
11.2 Protein Material Costs	204	13.2 Universal Domain Structure Families Identified in the Last Universal Common Ancestor	232
11.3 An Example: Proteomic Sulfur Sparing	205	13.3 Some Domain Families Recur More Frequently and Are Structurally Very Diverse	234
11.4 Episodic Nutrient Scarcity Can Shape Protein Material Costs	205	13.4 Correlation of Structural Diversity in Superfamilies with Functional Diversity	234
11.5 Highly Expressed Gene Products Often Exhibit Reduced Material Costs	206	13.5 To What Extent Does Function Vary Between Homologous?	238
11.6 Material Costs and the Evolution of Genomes	207	13.5.1 Phylogenetic Analysis of Protein Families	238
11.7 Material Costs and Other Costs of Making Proteins	208	13.5.2 Structural Domain Characterization of Clusters of Orthologous Genes Using CATH	240
11.8 Conclusions	209	13.5.3 Evolution of (COGs) Function Within CATH Superfamilies	240
Acknowledgments	209	13.5.4 Resolving Ambiguous Evolutionary Scenarios Between Parent and Child COGs in a CATH Superfamily	241
References	209	13.5.5 Relationship Between Domain Architecture Rearrangement and Functional Divergence Within CATH Superfamilies	242
12. Protein Domains as Evolutionary Units	213	13.6 How Safely Can Function Be Inherited Between Homologues?	245
<i>Andrew D. Moore and Erich Bornberg-Bauer</i>		13.7 How Are Domain Families Distributed in Protein Complexes?	247
12.1 Modular Protein Evolution	213	References	248
12.2 Domain-Based Homology Identification	215	14. Noncoding RNA	251
12.2.1 Domain Architecture Similarity	216	<i>Alexander Donath, Sven Findeiß, Jana Hertel, Manja Marz, Wolfgang Otto, Christine Schulz, Peter F. Stadler, and Stefan Wirth</i>	
12.2.2 Domain Resources and Domain-Based Search	219	14.1 Introduction	251
12.2.3 Deciphering Circular Permutations with Domains	221	14.2 Ancient RNAs	254
12.3 Domains in Genomics and Proteomics	222		
12.3.1 Building Domain Trees	223		
12.4 The Coverage Problem	225		
12.5 Conclusion	227		
References	228		

14.2.1	RNase P and RNase MRP RNA	254	15.2.4	"Exotic" Small RNA Species	297
14.2.2	Signal Recognition Particle RNA	254	15.3	Small RNA Biogenesis	298
14.2.3	snoRNAs	256	15.3.1	Components of the Small RNA Processing Machinery	298
14.3	Domain-Specific RNAs	259	15.3.2	MicroRNA Biogenesis	299
14.3.1	Telomerase RNA	259	15.3.3	Biogenesis of Other Small RNAs	300
14.3.2	Spliceosomal snRNAs	260	15.3.4	Three Main Mechanisms, Same Global Effect on Gene Expression	300
14.3.3	U7 snRNA	262	15.4	Computational microRNA Prediction	302
14.3.4	tmRNA	262	15.5	microRNA Targets	304
14.3.5	6S RNA	265	15.5.1	How Many Targets?	304
14.4	Conserved ncRNAs with Limited Distribution	267	15.5.2	Target Prediction	305
14.4.1	Y RNAs	267	15.5.3	Targets and Polymorphisms	306
14.4.2	Vault RNAs	267	15.6	Evolution of microRNAs	307
14.4.3	7SK RNA	267	15.6.1	Animal microRNAs	307
14.4.4	SmY RNA	270	15.6.2	Plant microRNAs	310
14.4.5	Bacterial RNAs	270	15.6.3	MicroRNAs and Viruses	311
14.4.6	A Zoo of Diverse Examples	274	15.6.4	Mirtrons	313
14.5	ncRNAs from Repeats and Pseudogenes	276	15.7	Origin(s) of microRNA Families	313
14.6	mRNA-like ncRNAs	277	15.7.1	Metazoa	313
14.6.1	Dosage Compensation	279	15.7.2	Mechanisms in Plants	314
14.6.2	Imprinting	280	15.7.3	microRNAs and Transposable Elements	314
14.6.3	Stress Response	280	15.7.4	Are Animal and Plant microRNAs Homologous?	315
14.6.4	Transcriptional Regulators	280	15.8	Genomic Organization	316
14.7	RNAs with Dual Functions	281	15.8.1	Clusters and Families	316
14.7.1	RNAIII	281	15.8.2	Regulation of microRNA Expression	318
14.7.2	SgrS	281	15.9	Summary and Outlook	320
14.7.3	SRA/SRAP	282	References	321	
14.7.4	Enod40	282			
14.8	Concluding Remarks	282			
Acknowledgments	283				
References	283				
15. Evolutionary Genomics of microRNAs and Their Relatives 295					
<i>Andrea Tanzer, Markus Riester, Jana Hertel, Clara Isabel Bermudez-Santana, Jan Gorodkin, Ivo L. Hofacker, and Peter F. Stadler</i>					
15.1	Introduction	295	16. Phylogenetic Utility of RNA Structure: Evolution's Arrow and Emergence of Early Biochemistry and Diversified Life 329		
15.2	The Small RNA Zoo	296	<i>Feng-Jie Sun, Ajith Harish, and Gustavo Caetano-Anollés</i>		
15.2.1	Endogenous siRNAs	296	16.1	Introduction	329
15.2.2	piRNA	297	16.1.1	A Novel Phylogenetic Approach Based on Macromolecular Structure	331
15.2.3	raRNA	297	16.1.2	Broadened Utility of Constraint Analysis	333

16.2	Structural Characters and Derived Phylogenetic Trees	333
16.2.1	Character Coding	334
16.2.2	Character Polarization	335
16.2.3	Phylogenetic Analysis of RNA Structural Characters	338
16.2.4	Character State Change Frequency in RNA Structural Evidence	339
16.2.5	Major Properties of Phylogenetic Trees Derived from RNA Structure	340
16.2.6	Potential Limitations of the Methodology	343
16.3	Applications	344
16.3.1	tRNA	344
16.3.2	5S rRNA	346
16.3.3	RNase P RNA	348
16.3.4	SINE RNA	351
16.3.5	rRNA	352
16.4	Conclusions	353
	Acknowledgments	354
	References	354

Part III Evolution of Biological Networks

17. A Hitchhiker's Guide to Evolving Networks 363

Charles G. Kurland and Otto G. Berg

17.1	Introduction	363
17.1.1	Mere Words?	364
17.1.2	Sideways	365
17.1.3	One Man's Glitch	365
17.1.4	Three Challenges	366
17.2	Phylogenetic Continuities, Biological Coherence	367
17.2.1	The Rosetta Stone	367
17.2.2	Compositional Outliers	368
17.2.3	Phylogeny by Any Other Name	368
17.2.4	The Emperor's Blast Search	370
17.2.5	Between Consenting Adults	371
17.3	Nested Structural Networks	371
17.3.1	Nip and Tuck	371
17.3.2	Fold Selection	372
17.3.3	Higher Order Structural Networks	373

17.4	Optimal Networks	374
17.4.1	Optimality Not Maximality	375
17.4.2	Patchy Environments, Patchy Genomes	377
17.4.3	Fixation of Novel Sequences	378
17.4.4	Selfish Operons	380
17.5	The Emperor's BLAST Search Revisited	381
17.5.1	Ecological Settings	384
17.5.2	The Way Forward	386
17.5.3	Less May be More	387
17.6	Will the Real Missing Link Please Stand Up?	388
17.7	All's Well	389
	Acknowledgments	391
	References	391

18. Evolution of Metabolic Networks 397

Eivind Almaas

18.1	Introduction	397
18.2	Metabolic Network Properties	398
18.3	Network Models For Metabolic Evolution	403
18.4	Dynamic Models Of Genome-Level Metabolic Function	407
	References	410

19. Single-Gene and Whole-Genome Duplications and the Evolution of Protein-Protein Interaction Networks 413

Grigoris Amoutzias and Yves Van de Peer

19.1	Introduction	413
19.2	Evolution of PINs	414
19.3	Single-Gene Duplications	416
19.4	Whole-Genome Duplications	416
19.5	Diploidization Phase	416
19.6	Dosage Balance Hypothesis	417
19.7	Types of Interactions	417
19.8	WGDs, Transient Interactions, and Organismal Complexity	418
19.9	Studies on PPIs of Ohnologues	419

