

# CONTENTS

Preface / [ix](#)

Author / [xi](#)

## 1 • INTRODUCTION / [1](#)

- 1.1 What Is Phylogenomics? / [1](#)
  - 1.1.1 The Early View of Phylogenomics / [1](#)
  - 1.1.2 An Expanded View of Phylogenomics / [2](#)
- 1.2 Anatomy of Gene Trees / [2](#)
- 1.3 Gene Trees versus Species Trees / [3](#)
- 1.4 Phylogenomics and the Tree of Life / [4](#)
- 1.5 Sequencing Workflows to Generate Phylogenomic Data / [6](#)
  - 1.5.1 Sanger Sequencing Workflow / [6](#)
  - 1.5.2 NGS Workflow / [6](#)
  - 1.5.3 Is Sanger Sequencing Still Relevant in Phylogenomics? / [7](#)
- 1.6 The Phylogenomics Laboratory / [8](#)
- References / [10](#)

## 2 • PROPERTIES OF DNA SEQUENCE LOCI: PART I / [13](#)

- 2.1 Genomic Background / [13](#)
  - 2.1.1 Genome Types and Sizes / [13](#)
  - 2.1.2 Composition of Eukaryotic Organellar Genomes / [15](#)
  - 2.1.3 Composition of Eukaryotic Nuclear Genomes / [18](#)
    - 2.1.3.1 Gene Numbers and Densities among Nuclear Genomes / [18](#)
    - 2.1.3.2 Intergenic DNA / [18](#)
- 2.2 DNA Sequence Evolution / [21](#)
  - 2.2.1 Patterns and Processes of Base Substitutions / [21](#)
    - 2.2.1.1 Transition Bias / [21](#)
    - 2.2.1.2 Transition Bias and DNA Replication Errors / [23](#)
    - 2.2.1.3 Saturation of DNA Sites / [24](#)
    - 2.2.1.4 Among-Site Substitution Rate Variation / [26](#)
  - 2.2.2 Tandemly Repeated DNA Sequences / [26](#)
  - 2.2.3 Transposable Elements / [27](#)
  - 2.2.4 Processed Pseudogenes / [30](#)

- 2.2.5 Mitochondrial Pseudogenes (“Numts”) / 30
- 2.2.5.1 Numt Abundance in Eukaryotic Genomes / 30
- 2.2.5.2 Mechanisms of Primary Numt Integration / 31
- 2.2.5.3 Differences between Numts and Mitochondrial DNA / 32

References / 33

### 3 • PROPERTIES OF DNA SEQUENCE LOCI: PART II / 37

- 3.1 Six Assumptions about DNA Sequence Loci in Phylogenomic Studies / 37
  - 3.1.1 Assumption 1: Loci Are Single-Copy in the Genome / 37
  - 3.1.2 Assumption 2: Loci Are Selectively Neutral / 39
    - 3.1.2.1 Does “Junk DNA” Exist? / 39
    - 3.1.2.2 The Neutrality Assumption and the Indirect Effects of Natural Selection / 40
  - 3.1.3 Assumption 3: Sampled Loci Have Independent Gene Trees / 43
    - 3.1.3.1 How Many Independent Loci Exist in Eukaryotic Genomes? / 44
    - 3.1.3.2 Criteria for Delimiting Loci with Independent Gene Trees / 45
  - 3.1.4 Assumption 4: No Historical Recombination within Loci / 47
    - 3.1.4.1 Intralocus Recombination and Gene Trees / 47
    - 3.1.4.2 What Is the Optimal Locus Length? / 48
  - 3.1.5 Assumption 5: Loci Evolved Like a Molecular Clock / 51
  - 3.1.6 Assumption 6: Loci Are Free of Ascertainment Bias / 52
- 3.2 DNA Sequence Loci: Terminology and Types / 52
  - 3.2.1 On Genes, Alleles, and Related Terms / 52
  - 3.2.2 Commonly Used DNA Sequence Loci in Phylogenomic Studies / 53
    - 3.2.2.1 Mitochondrial DNA Loci / 53
    - 3.2.2.2 Nuclear DNA Loci / 54

References / 60

### 4 • DNA EXTRACTION / 67

- 4.1 DNA Extraction Methodology / 67
  - 4.1.1 Summary of the DNA Extraction Process / 67
  - 4.1.2 A Note about DNA Storage Buffers / 69
  - 4.1.3 Extracting DNA from Plants, Fungi, and Invertebrates / 70
  - 4.1.4 Extracting DNA from Formalin-Fixed Museum Specimens / 70
- 4.2 Evaluating the Results of DNA Extractions / 71
  - 4.2.1 Agarose Gel Electrophoresis / 72
    - 4.2.1.1 Troubleshooting / 74
  - 4.2.2 UV Spectrophotometric Evaluation of DNA Samples / 75
    - 4.2.2.1 UV Spectrophotometry to Determine Concentrations of Nucleic Acid Samples / 75
    - 4.2.2.2 UV Spectrophotometry to Determine the Purity of DNA Samples / 76
  - 4.2.3 Fluorometric Quantitation of DNA Samples / 76
- 4.3 The High-Throughput Workflow / 76
  - 4.3.1 High-Throughput DNA Extractions / 77
    - 4.3.1.1 Extracting DNA from 96 Tissue Samples / 77
    - 4.3.1.2 High-Throughput Agarose Gel Electrophoresis / 78
    - 4.3.1.3 High-Throughput UV Spectrophotometry / 78
    - 4.3.1.4 Preparation of Diluted DNA Templates for High-Throughput PCR / 78

References / 79

### 5 • PCR THEORY AND PRACTICE / 81

- 5.1 Historical Overview / 81
- 5.2 DNA Polymerization in Living Cells versus PCR / 83
  - 5.2.1 Brief Review of DNA Polymerization in Living Cells / 83

|         |                                                          |     |
|---------|----------------------------------------------------------|-----|
| 5.2.2   | How the PCR Works /                                      | 85  |
| 5.3     | PCR Procedures /                                         | 89  |
| 5.3.1   | Preparation of PCR Reagents and Reaction Setup /         | 90  |
| 5.3.1.1 | PCR Reagents /                                           | 90  |
| 5.3.1.2 | Importance of Making Reagent Aliquots /                  | 91  |
| 5.3.1.3 | Setting Up PCRs /                                        | 92  |
| 5.3.2   | Thermocycling /                                          | 93  |
| 5.3.3   | Checking PCR Results Using Agarose Gel Electrophoresis / | 94  |
| 5.4     | PCR Troubleshooting /                                    | 94  |
| 5.5     | Reducing PCR Contamination Risk /                        | 97  |
| 5.6     | High-Throughput PCR /                                    | 98  |
| 5.6.1   | Setting Up PCRs in a 96-Sample Microplate Format /       | 98  |
| 5.7     | Other PCR Methods /                                      | 98  |
| 5.7.1   | Hot Start PCR /                                          | 99  |
| 5.7.2   | Long PCR /                                               | 100 |
| 5.7.3   | Reverse Transcriptase-PCR /                              | 101 |
|         | References /                                             | 102 |

## 6 • SANGER SEQUENCING / 105

|         |                                                                                  |     |
|---------|----------------------------------------------------------------------------------|-----|
| 6.1     | Principles of Sanger Sequencing /                                                | 105 |
| 6.1.1   | The Sanger Sequencing Concept /                                                  | 105 |
| 6.1.2   | Modern Sanger Sequencing /                                                       | 107 |
| 6.1.2.1 | Cycle Sequencing Reaction /                                                      | 107 |
| 6.1.2.2 | Gel Electrophoresis of Extension Products /                                      | 108 |
| 6.1.2.3 | Sequence Data Quality /                                                          | 109 |
| 6.2     | Sanger Sequencing Procedures /                                                   | 112 |
| 6.2.1   | Purification of PCR Products /                                                   | 112 |
| 6.2.1.1 | Exo-SAP Treatment of PCR Products /                                              | 112 |
| 6.2.1.2 | Spin Column and Vacuum Manifold Kits<br>for PCR Product Purification /           | 112 |
| 6.2.1.3 | 20% PEG 8000 Precipitation of PCR Products /                                     | 113 |
| 6.2.1.4 | Solid-Phase Reversible Immobilization Beads /                                    | 113 |
| 6.2.1.5 | Gel Purification of PCR Products /                                               | 114 |
| 6.2.1.6 | Which PCR Product Purification Method Is Best? /                                 | 115 |
| 6.2.2   | Setting Up Cycle Sequencing Reactions /                                          | 115 |
| 6.2.3   | Purification of Extension Products /                                             | 115 |
| 6.2.4   | Sequencing in a Capillary Sequencer: Do-It-Yourself or Outsource? /              | 116 |
| 6.3     | High-Throughput Sanger Sequencing /                                              | 116 |
| 6.3.1   | Sequencing 96 Samples on Microplates /                                           | 116 |
| 6.3.2   | Adding Sequencing Primer “Tails” to PCR Primers /                                | 117 |
| 6.3.2.1 | How an M13-Tailed Primer Functions in PCR /                                      | 118 |
| 6.3.2.2 | Cycle Sequencing and M13 Primer Tails /                                          | 118 |
| 6.3.2.3 | On the Importance of Matching Sequencing Primers /                               | 121 |
| 6.3.2.4 | Benefits of Using M13-Tailed Primers /                                           | 123 |
| 6.4     | Haplotype Determination from Sanger Sequence Data /                              | 123 |
| 6.4.1   | PCR Amplification and Sanger Sequencing of Diploid or Polyploid Loci /           | 123 |
| 6.4.2   | Multiple Heterozygous SNP Sites and Haplotype Sequences /                        | 126 |
| 6.4.3   | Methods for Obtaining Nuclear Haplotype Sequences from Sanger<br>Sequence Data / | 127 |
| 6.4.3.1 | Physical Isolation of PCR Haplotypes prior to Sequencing /                       | 128 |
| 6.4.3.2 | Statistical Inference of Haplotypes from Sanger Sequence Data /                  | 128 |
|         | References /                                                                     | 129 |

## **7 • ILLUMINA SEQUENCING / 131**

- 7.1 How Illumina Sequencing Works / 131
  - 7.1.1 Construction of Indexed Sequencing Libraries / 133
  - 7.1.2 Generation of Clusters on a Flow Cell / 133
  - 7.1.3 Sequencing of Clusters / 135
- 7.2 Methods for Obtaining Multiplexed Hybrid Selection Libraries / 143
  - 7.2.1 Library Preparation Approaches / 145
    - 7.2.1.1 Traditional Illumina Library Approach / 145
    - 7.2.1.2 Meyer and Kircher Library Approach / 156
    - 7.2.1.3 Rohland and Reich Library Approach / 163
    - 7.2.1.4 Nextera Library Approach / 165
  - 7.2.2 In-Solution Hybrid Selection / 175
  - 7.2.3 Indexing, Pooling, and Hybrid Selection Efficiency Revisited / 185
- 7.3 Cost-Effective Methods for Obtaining Multiplexed Targeted-Loci Libraries / 187
  - 7.3.1 Sequence Capture Using PCR-Generated Probes (SCPP) / 187
  - 7.3.2 Parallel Tagged Amplicon Sequencing / 190
- References / 191

## **8 • DEVELOPING DNA SEQUENCE LOCI / 195**

- 8.1 Primer Design Theory / 196
  - 8.1.1 Rules of Primer Design / 196
  - 8.1.2 Final Comments about Primer Design Rules / 204
  - 8.1.3 Testing New Primers in the Lab / 205
- 8.2 Primer and Probe Design Approaches / 205
  - 8.2.1 Single Template Approaches for Developing PCR-Based Loci / 206
    - 8.2.1.1 Single Template Methods Using Genomic Cloning Methods / 206
    - 8.2.1.2 Single Template Methods Using Available Genomics Resources / 210
    - 8.2.1.3 Single Template Methods Using NGS Partial Genome Data / 210
    - 8.2.1.4 Single Template Methods Using Whole Genome Sequences / 211
  - 8.2.2 Multiple Homologous Template Approaches for Designing PCR-Based and Anchor Loci / 211
    - 8.2.2.1 Designing Universal Primers by Comparative Sequence Analysis / 212
    - 8.2.2.2 Multiple Homologous Template Approaches Using Whole Genome Sequences / 215
    - 8.2.2.3 Designing Anchor Loci Probes Using Whole Genome Sequences / 216
- References / 217

## **9 • FUTURE OF PHYLOGENOMIC DATA ACQUISITION / 221**

- 9.1 The Impending Flood of Genomes / 221
- 9.2 In Silico Acquisition of Phylogenomic Datasets / 222
- References / 224
- Index / 225