## CONTENTS

Preface........................................ page ix

0 Meet Python.................................. 1
  0.1 Getting Started......................... 1
  0.2 Big Numbers............................ 3
  0.3 Strange Division....................... 3
  0.4 Naming Things......................... 4
  0.5 What’s in a Name?...................... 6
  0.6 From Numbers to Strings.............. 6
  0.7 Slicing.................................. 8
  0.8 Adding Strings......................... 9
  0.9 Negative Indices....................... 10
  0.10 Fancy Slicing......................... 10
  0.11 And Now to Lists..................... 11
  0.12 Lists for Free!....................... 13
  0.13 Changing Values...................... 14
  0.14 More on Mutability................... 17
  0.15 Booleans................................ 18
  Putting it All Together................... 21

Part I Python versus Pathogens........... 23

1 Computing GC Content.................... 25
  1.1 Representing DNA on a Computer..... 26
  1.2 Python Functions...................... 26
  1.3 A Short Comment on Docstrings..... 28
  1.4 Bigger and Better Functions....... 28
  1.5 Why Write Functions?............... 29
  1.6 Making Decisions..................... 29
  1.7 A Potential Pitfall................... 32
  1.8 GC Content of Strings of Length 1, 2, 3, or 4: if, elif, else 34
  1.9 Oops!.................................. 36
  1.10 The “Perfect” GC Content Function: for Loops 38
## CONTENTS

1.11 Another Example of for Loops: Converting DNA to RNA  
Putting it All Together  

2 Pathogenicity Islands  
2.1 How *Salmonella* Enters Host Cells  
2.2 Investigating Pathogenicity Islands  
2.3 Looping Over Lists  
2.4 Looping Over Lists with `range`  
2.5 From Gum to CAT Boxes  
2.6 Functions Can Call Other Functions!  
Putting it All Together  

3 Open Reading Frames and Genes  
3.1 Open Reading Frames and the Central Dogma  
3.2 GC Content and ORFs  
3.3 The `countStarts` Function  
3.4 The `genString` Function  
3.5 `while` Loops and Population Genetics  
Putting it All Together  

4 Finding Genes (at last!)  
4.1 From ORFs to Genes  
4.2 Genes Occur on Both Strands  
4.3 Determining the Function of a Protein  
Putting it All Together  

Part II Sequence Alignment and Sex Determination  

5 Recursion  
5.1 A Brief Diversion Before Recursion  
5.2 And Now For Recursion!  
5.3 The Factorial Function  
5.4 How to Write a Recursive Function  
5.5 Another Example: Recursive Reverse  
Putting it All Together  

6 The Use-It-Or-Lose-It Principle  
6.1 Peptide Fragments  
6.2 Making Change
7 Dictionaries, Memoization, and Speed 100
7.1 Dictionaries 101
7.2 Using a Dictionary 102
7.3 What Kinds of Things Can be Keys and Values? 103
7.4 Optional Section: How Dictionaries Work (and why you should care) 105
7.5 Memoization 109
7.6 Memoizing LCS 113
Putting it All Together 115

8 Sequence Alignment and the Evolution of Sex Chromosomes 116
8.1 The Sequence Alignment Score 116
8.2 From Scores to Alignments 121
8.3 Sequence Alignment Scoring with Variable Rewards and Penalties 123
8.4 Optional Section: How Substitution Matrices are Computed 126
8.5 Optional Section: Getting the Actual Sequence Alignment 129
8.6 Identifying Orthologs 135
8.7 Comparing Chromosomes 136
Putting it All Together 138

Part III Phylogenetic Reconstruction and the Origin of Modern Humans 141

9 Representing and Working with Trees 145
9.1 Representing Trees 146
9.2 Computing with Trees 148
Putting it All Together 152

10 Drawing Trees 154
10.1 Drawing Fractal Trees 158
10.2 Drawing Phylogenetic Trees 159
Putting it All Together 162

11 The UPGMA Algorithm 163
11.1 The Algorithm 164
11.2 Implementing UPGMA in Python 169
11.3 Calibrating Trees 171
Putting it All Together 172
## CONTENTS

**Part IV Additional Topics** 175

12 RNA Secondary Structure Prediction 177

13 Gene Regulatory Networks and the Maximum Likelihood Method 186
   Putting it All Together 195

14 Birds, Bees, and Genetic Algorithms 196
   14.1 Fast Algorithms 198
   14.2 Slow Algorithms 199
   14.3 Genetic Algorithms 201

Where to go from here 205

Index 206