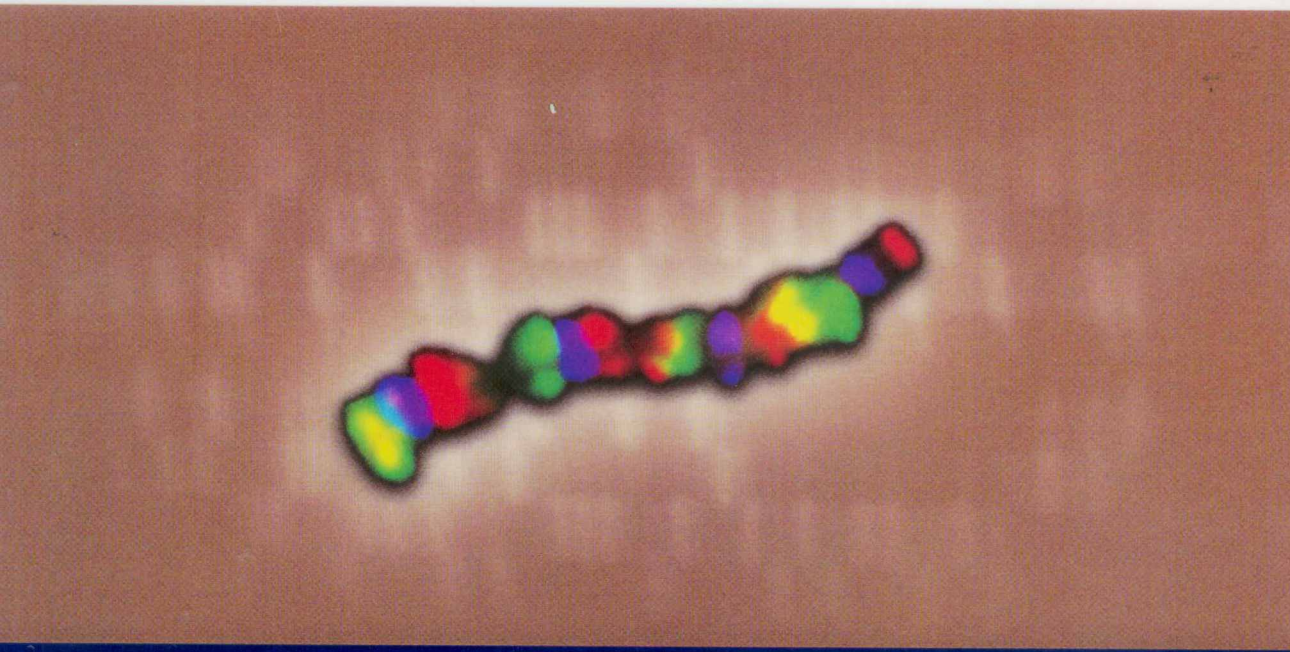
Perl

Programming for Biologists



D. Curtis Jamison

WILEY

Contents

Part I. The Basics 1

Introduction 3

Chapter 1. An Introduction to Perl 7
- 1.1 The Perl Interpreter 7
- 1.2 Your First Perl Program 8
- 1.3 How the Perl Interpreter Works 9
- Chapter Summary 10
- For More Information 11
- Exercises 11

Chapter 2. Variables and Data Types 13
- 2.1 Perl Variables 13
- 2.2 Scalar Values 14
- 2.3 Calculations 17
- 2.4 Interpolation and Escapes 19
- 2.5 Variable Definition 22
- 2.6 Special Variables 23
- Chapter Summary 23
- For More Information 24
- Exercises 24
- Programming Challenges 24

Chapter 3. Arrays and Hashes 27
- 3.1 Arrays 27
- 3.3 Array Manipulation 30
  - 3.3.1 Push and Pop, Shift and Unshift 30
  - 3.3.2 Splice 31
  - 3.3.3 Other Useful Array Functions 33
  - 3.3.4 List and Scalar Context 34
- 3.4 Hashes 37
- 3.5 Maintaining a Hash 38

Perl Programming for Biologists, D. Curtis Jamison
ISBN 0-471-43059-5

Copyright © 2003 Wiley-Liss, Inc.

Chapter Summary 40
For More Information 40
Exercises 40
Programming Challenge 41

Chapter 4. Control Structures 43
4.1 Comparisons 44
4.2 Choices 45
4.2.1 If 45
4.2.2 Boolean Operators 46
4.2.3 Else 47
4.3 Loops 49
4.3.1 For Loops 50
4.3.2 Foreach Loops 52
4.4 Indeterminate Loops 54
4.4.1 While 54
4.4.2 Repeat Until 56
4.5 Loop Exits 57
4.5.1 Last 57
4.5.2 Next and Continue 57
Chapter Summary 59
Exercises 59
Programming Challenges 60

Part II. Intermediate Perl 61

Chapter 5. Subroutines 63
5.1 Creating a Subroutine 63
5.2 Arguments 64
5.3 Return 65
5.3.1 Wantarray 66
5.4 Scope 67
5.4.1 My 67
5.5 Passing Arguments with References 70
5.6 Sort Subroutines 71
Chapter Summary 73
For More Information 74
Exercises 74
Programming Challenges 74

Chapter 6. String Manipulation 75
6.1 Array-Based Character Manipulation 75
6.2 Regular Expressions 78
6.2.1 Match 79
6.2.2 Substitute 81
6.2.3 Translate 81

6.3 Patterns 82
6.3.1 Atoms 83
6.3.2 Special Atoms 83
6.3.3 Quantifiers 84
6.3.4 Assertions 85
6.3.5 Alternatives 85
Chapter Summary 86
For More Information 87
Exercises 87
Programming Challenges 87

Chapter 7. Input and Output 89
7.1 Program Parameters 89
7.2 File I/O 90
7.2.1 Filehandles 90
7.2.2 Working with Files 91
7.2.3 Built-in File Handles 92
7.2.4 File Safety 93
7.2.5 The Input Operator 94
7.2.6 Binary I/O 97
7.3 Interprocess Communications 97
7.3.1 Processes 98
7.3.2 Process Pipes 98
7.3.3 Creating Processes 99
7.3.4 Monitoring Processes 100
7.3.5 Implicit Forks 101
Chapter Summary 102
For More Information 102
Exercises 102
Programming Challenges 103

Chapter 8. Perl Modules and Packages 105
8.1 Modules 105
8.2 Packages 107
8.3 Combining Packages and Modules 109
8.4 Included Modules 110
8.4.1 CGI 110
8.4.2 Getopt 110
8.4.3 Io 112
8.4.4 File::Path 112
8.4.5 Strict 113
8.5 The CPAN 114
8.5.1 Setting Up the CPAN Module 114
8.5.2 Finding Modules 115

8.5.3 Installing Modules 117
8.5.4 Managing Installed Modules 119
Chapter Summary 121
For More Information 121
Exercises 121
Programming Challenges 122

Part III. Advanced Perl 123

Chapter 9. References 125
9.1 Creating References 125
9.2 ref() 126
9.3 Anonymous Referents 127
9.4 Tables 128
Chapter Summary 130
Exercises 130
Programming Challenge 130

Chapter 10. Object-Oriented Programming 133
10.1 Introduction to Objects 133
10.1.1 The OOP Approach 134
10.1.2 Class Design 135
10.1.3 Inheritance 136
10.2 Perl Objects 136
10.2.1 Rule Number One 137
10.2.2 Rule Number Two 137
10.2.3 Rule Number Three 138
10.2.4 Methods 139
10.2.5 Constructors 141
10.2.6 Accessors 143
10.2.7 OOP Versus Procedural 143
Chapter Summary 145
For More Information 146
Exercises 146
Programming Challenges 146

Chapter 11. Bioperl 147
11.1 Sequences 147
11.2 SeqFeature 149
11.3 Annotation 150
11.4 Sequence I/O 151
11.5 Cool Tools 152
11.6 Example Bioperl Programs 154
11.6.1 Primer.pl 154
11.6.2 Primer3.pm 156
Chapter Summary 161

For More Information 161
Exercises 161
Programming Challenges 162

Appendix A. Partial Perl Reference 163
Chapter 3 163
Chapter 4 163
Chapter 5 164
Chapter 6 164
Chapter 7 164
Chapter 8 165
Chapter 9 165

Appendix B. Bioinformatics File Formats 167
GenBank 167
ASN.1 170
EMBL 175
PDB 177
Fasta 181
BLAST 182
ACEDB 183

Index 185