

---

# Contents

---

## Part I Whetting Your Appetite

---

<b>1</b>	<b>Introduction</b>	3
1.1	Information	3
1.2	What Is Linux?	3
1.3	What Is Shell Programming?	4
1.4	What Is <code>sed</code> ?	4
1.5	What Is <code>awk</code> ?	4
1.6	What Is <code>perl</code> ?	5
1.7	Prerequisites	5
1.8	Conventions	6

---

## Part II Computer and Operating Systems

---

<b>2</b>	<b>Unix/Linux</b>	9
2.1	What Is a Computer?	9
2.2	Some History	10
2.2.1	Versions of Unix	11
2.2.2	The Rise of Linux	12
2.2.3	Why a Penguin?	14
2.2.4	Linux Distributions	14
2.3	X-Windows	14
2.3.1	How Does It Work?	15
2.4	Unix/Linux Architecture	16
2.5	What Is the Difference Between Unix and Linux?	17
2.6	What Is the Difference Between Unix/Linux and Windows?	17
2.7	What Is the Difference Between Unix/Linux and Mac OS X?	18
2.8	One Computer, Two Operating Systems	18
2.8.1	VMware	19
2.8.2	CygWin	19

2.8.3	Wine	19
2.8.4	Others	20
2.9	Knoppix	20
2.10	Software Running Under Linux	21
2.10.1	Biosciences	21
2.10.2	Office and Co.	22

---

## Part III Working with Unix/Linux

---

<b>3</b>	<b>The First Touch</b>	27
3.1	Login	27
3.1.1	Working with CygWin or Mac OS X	27
3.1.2	Working Directly on a Unix/Linux Computer	28
3.1.3	Working on a Remote Unix/Linux Computer	28
3.2	Using Commands	30
3.2.1	Syntax of Commands	31
3.2.2	Editing the Command Line	32
3.2.3	Change Password	32
3.2.4	How to Find out More	33
3.3	Logout	33
	Exercises	34
<b>4</b>	<b>Working with Files</b>	35
4.1	Browsing Files and Directories	36
4.2	File Attributes	38
4.3	Special Files: . and	39
4.4	Protecting Files and Directories	41
4.4.1	Directories	41
4.4.2	Files	41
4.4.3	Examples	41
4.4.4	Changing File Attributes	42
4.4.5	Extended File Attributes	44
4.5	File Archives	44
4.6	File Compression	47
4.7	Searching for Files	49
	Exercises	50
<b>5</b>	<b>Installing BLAST and ClustalW</b>	53
5.1	Downloading the Programs via FTP	53
5.1.1	Downloading BLAST	54
5.1.2	Downloading ClustalW	55
5.2	Installing BLAST	57
5.3	Running BLAST	57
5.4	Installing ClustalW	59

5.5	Running ClustalW .....	59
	Exercises .....	60
<b>6</b>	<b>Working with Text .....</b>	<b>63</b>
6.1	A Quick Start: <code>cat</code> .....	64
6.1.1	Text Sorting .....	65
6.1.2	View File Beginning or End .....	67
6.1.3	Scrolling Through Files .....	67
6.1.4	Character, Word and Line Counting .....	67
6.1.5	Finding Text .....	68
6.1.6	Text File Comparisons .....	69
6.2	<code>pico</code> .....	70
6.3	<code>vi</code> and <code>vim</code> .....	72
6.3.1	Immediate Takeoff .....	72
6.3.2	Starting <code>vi</code> .....	73
6.3.3	Modes .....	73
6.3.4	Moving the Cursor .....	74
6.3.5	Doing Corrections .....	75
6.3.6	Save and Quit .....	76
6.3.7	Copy and Paste .....	76
6.3.8	Search and Replace .....	78
	Exercises .....	78
<b>7</b>	<b>Using the Shell .....</b>	<b>81</b>
7.1	What Is the Shell? .....	81
7.2	Different Shells .....	82
7.3	Setting the Default Shell .....	83
7.4	Useful Shortcuts .....	84
7.5	Redirections .....	85
7.6	Pipes .....	87
7.7	Lists of Commands .....	88
7.8	Aliases .....	89
7.9	Scheduling Commands .....	89
7.10	Wildcards .....	91
7.11	Processes .....	92
7.11.1	Checking Processes .....	93
7.11.2	Background Processes .....	94
7.11.3	Killing Processes .....	95
	Exercises .....	96
<b>8</b>	<b>Shell Programming .....</b>	<b>97</b>
8.1	Script Files .....	97
8.2	Modifying the Path .....	99
8.3	Variables .....	100
8.4	Input and Output .....	102

8.4.1	<code>echo</code> .....	103
8.4.2	Here Documents: <code>&lt;&lt;</code> .....	103
8.4.3	<code>read</code> and <code>line</code> .....	104
8.4.4	Script Parameters .....	105
8.5	Substitutions and Expansions .....	106
8.5.1	Variable Substitution .....	107
8.5.2	Command Expansion .....	108
8.6	Quoting .....	109
8.6.1	Escape Character .....	109
8.6.2	Single Quotes .....	109
8.6.3	Double Quotes .....	109
8.7	Decisions – Flow Control .....	110
8.7.1	<code>if...then...elif...else...fi</code> .....	110
8.7.2	<code>test</code> .....	111
8.7.3	<code>while...do...done</code> .....	113
8.7.4	<code>until...do...done</code> .....	115
8.7.5	<code>for...in...do...done</code> .....	116
8.7.6	<code>case...in...esac</code> .....	117
8.7.7	<code>select...in...do</code> .....	118
8.8	Debugging .....	119
8.8.1	<code>bash -xv</code> .....	120
8.8.2	<code>trap</code> .....	121
8.9	Examples .....	122
8.9.1	Check for DNA as File Content .....	122
8.9.2	Time Signal .....	123
8.9.3	Select Files to Archive .....	123
8.9.4	Remove Spaces .....	125
	Exercises .....	125
<b>9</b>	<b>Regular Expressions</b> .....	<b>127</b>
9.1	Using Regular Expressions .....	128
9.2	Search Pattern and Examples .....	130
9.2.1	Single-Character Meta Characters .....	130
9.2.2	Quantifiers .....	133
9.2.3	Grouping .....	134
9.2.4	Anchors .....	134
9.2.5	Escape Sequences .....	135
9.2.6	Alternation .....	136
9.2.7	Back References .....	137
9.2.8	Character Classes .....	138
9.2.9	Priorities .....	138
9.2.10	<code>egrep</code> Options .....	139
9.3	Regular Expressions and <code>vim</code> .....	139
	Exercises .....	140

<b>10 Sed</b> .....	141
10.1 When to Use <b>sed</b> ? .....	142
10.2 Getting Started .....	142
10.3 How <b>sed</b> Works .....	144
10.3.1 Pattern Space .....	144
10.3.2 Hold Space .....	145
10.4 <b>sed</b> Syntax .....	145
10.4.1 Addresses .....	146
10.4.2 <b>sed</b> and Regular Expressions .....	147
10.5 Commands .....	148
10.5.1 Substitutions .....	149
10.5.2 Transliterations .....	151
10.5.3 Deletions .....	152
10.5.4 Insertions and Changes .....	152
10.5.5 <b>sed</b> Script Files .....	154
10.5.6 Printing .....	155
10.5.7 Reading and Writing Files .....	156
10.5.8 Advanced <b>sed</b> .....	157
10.6 Examples .....	157
10.6.1 Gene Tree .....	157
10.6.2 File Tree .....	158
10.6.3 Reversing Line Order .....	159
Exercises .....	160

---

## Part IV Programming

---

<b>11 Awk</b> .....	163
11.1 Getting Started .....	164
11.2 <b>awk</b> 's Syntax .....	165
11.3 Example File .....	166
11.4 Patterns .....	167
11.4.1 Regular Expressions .....	167
11.4.2 Pattern-Matching Expressions .....	168
11.4.3 Relational Character Expressions .....	169
11.4.4 Relational Number Expressions .....	171
11.4.5 Mixing and Conversion of Numbers and Characters ..	171
11.4.6 Ranges .....	172
11.4.7 <b>BEGIN</b> and <b>END</b> .....	173
11.5 Variables .....	174
11.5.1 Assignment Operators .....	174
11.5.2 Increment and Decrement .....	176
11.5.3 Predefined Variables .....	177
11.5.4 Arrays .....	181
11.5.5 Shell Versus <b>awk</b> Variables .....	185

11.6	Scripts and Executables . . . . .	185
11.7	Decisions – Flow Control . . . . .	186
11.7.1	if . . . else . . . . .	187
11.7.2	while . . . . .	188
11.7.3	do . . . while . . . . .	189
11.7.4	for . . . . .	189
11.7.5	Leaving Loops . . . . .	190
11.8	Actions . . . . .	191
11.8.1	Printing . . . . .	191
11.8.2	Numerical Calculations . . . . .	195
11.8.3	String Manipulation . . . . .	196
11.8.4	System Commands . . . . .	200
11.8.5	User-Defined Functions . . . . .	201
11.9	Input with <code>getline</code> . . . . .	206
11.9.1	Reading from a File . . . . .	206
11.9.2	Reading from the Shell . . . . .	209
11.10	Examples . . . . .	210
11.10.1	Sort an Array by Indices . . . . .	210
11.10.2	Sum up Atom Positions . . . . .	210
11.10.3	Translate DNA to Protein . . . . .	211
11.10.4	Calculate Atomic Composition of Proteins . . . . .	212
11.10.5	Calculate Distance Between Cysteines . . . . .	215
	Exercises . . . . .	220
<b>12</b>	<b>Perl</b> . . . . .	<b>221</b>
12.1	Intention of this Chapter . . . . .	221
12.2	Running <code>perl</code> Scripts . . . . .	222
12.3	Variables . . . . .	222
12.3.1	Scalars . . . . .	223
12.3.2	Arrays . . . . .	224
12.3.3	Hashes . . . . .	229
12.3.4	Built-in Variables . . . . .	233
12.4	Decisions – Flow Control . . . . .	234
12.4.1	if . . . elsif . . . else . . . . .	234
12.4.2	unless, die and warn . . . . .	234
12.4.3	while . . . . .	235
12.4.4	do . . . while . . . . .	235
12.4.5	until . . . . .	235
12.4.6	do . . . until . . . . .	235
12.4.7	for . . . . .	236
12.4.8	foreach . . . . .	236
12.4.9	Controlling Loops . . . . .	237
12.5	Data Input . . . . .	239
12.5.1	Command Line . . . . .	239
12.5.2	Internal . . . . .	239

12.5.3	Files	240
12.6	Data Output	242
12.6.1	<code>print</code> , <code>printf</code> and <code>sprintf</code>	242
12.6.2	Here Documents: <code>&lt;&lt;</code>	242
12.6.3	Files	243
12.7	Hash Databases	244
12.8	Regular Expressions	245
12.8.1	Special Escape Sequences	245
12.8.2	Matching: <code>m/.../</code>	246
12.9	String Manipulations	247
12.9.1	Substitute: <code>s/.../.../</code>	247
12.9.2	Transliterate: <code>tr/.../.../</code>	247
12.9.3	Common Commands	248
12.10	Calculations	249
12.11	Subroutines	250
12.12	Packages and Modules	251
12.13	Bioperl	254
12.14	You Want More?	254
12.15	Examples	254
12.15.1	Reverse Complement DNA	255
12.15.2	Calculate GC Content	255
12.15.3	Restriction Enzyme Digestion	256
12.15.4	Levenshtein Distance of Sequences	259
	Exercises	263
<b>A</b>	<b>Appendix</b>	265
A.1	Keyboard Shortcuts	265
A.2	Boot Problems	266
A.3	Optimizing the Bash Shell	266
A.3.1	Startup Files	266
A.3.2	A Helpful Shell Prompt	267
A.4	Text File Conversion (Unix ↔ DOS)	267
A.5	Devices	267
A.6	Mounting Filesystems	268
A.6.1	Floppy Disk	269
A.6.2	CD-ROM	270
A.6.3	PCMCIA and Compact Flash	270
A.7	Nucleotide and Protein Codes	271
A.8	Special Characters	272
	<b>Solutions</b>	273
	<b>References</b>	279
	<b>Index</b>	281