Unix and Perl to the Rescue!

Your research has generated gigabytes of data and now you need to analyze it. You hate using spreadsheets, but it’s all you know, so what else can you do? This book will transform how you work with large and complex data sets, teaching you powerful programming tools for slicing and dicing data to suit your needs.

Written in a fun and accessible style, this step-by-step guide will inspire and inform non-programmers about the essential aspects of Unix and Perl. It shows how, with just a little programming knowledge, you can write programs that could save you hours, or even days. No prior experience is required, and new concepts are introduced using numerous code examples that you can try out for yourself. Going beyond the basics, the authors touch upon many broader topics that will help those new to programming, including debugging and how to write in a good programming style.

KEITH BRADNAM is a project scientist in the Genome Center at the University of California, Davis. He has extensive experience working with model organism databases and spent four years as a project leader at WormBase, helping to develop this important bioinformatics resource.

IAN KORF is an Associate Professor in Molecular and Cellular Biology at the University of California, Davis. His research seeks to understand structure and function in genomic DNA. He has developed new tools for gene prediction, co-authored the only book devoted to BLAST and helped in the development of BioPerl.
Unix and Perl to the Rescue!
A field guide for the life sciences (and other data-rich pursuits)

KEITH BRADNAM
University of California, Davis

IAN KORF
University of California, Davis
Cambridge University Press
978-1-107-00068-1 — UNIX and Perl to the Rescue!
Keith Bradnam, Ian Korf
Frontmatter
More Information

CAMBRIDGE UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India
103 Penang Road, #05-06/07, Visioncres Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University’s mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107000681

© Keith Bradnam and Ian Korf 2012

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2012
Reprinted 2014

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging in Publication data
B radnam, Keith.
UNIX and Perl to the rescue! : a field guide for the life sciences (and other data-rich pursuits) / Keith Bradnam, Ian Korf.
p. cm.
Includes bibliographical references and index.
1. Science – Data processing. 2. UNIX (Computer file) 3. Perl (Computer program language)
I. Korf, Ian. II. Title.
Q183.9.B73 2012
005.4 '32–dc23 2011047927

ISBN 978-1-107-00068-1 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.
## Contents

**Part 1**  Introduction and background  
1.1 Introduction  
1.2 How to use this book  

**Part 2**  Installing Unix and Perl  
2.1 What do I need in order to learn Unix and Perl?  
2.2 Installing Linux on a PC  
2.3 Installing a code editor  

**Part 3**  Essential Unix  
3.1 Introduction to Unix  
3.2 The Unix terminal  
3.3 The Unix command prompt  
3.4 Your first Unix command  
3.5 The hierarchy of a Unix filesystem  
3.6 Finding out where you are in the filesystem  
3.7 How to navigate a Unix filesystem  
3.8 Absolute and relative paths  
3.9 Working with your home directory  
3.10 The Unix shell  
3.11 Environment variables  
3.12 Introduction to command-line options  
3.13 Man pages  
3.14 Working with directories  
3.15 The importance of saving keystrokes  
3.16 Moving and renaming files  
3.17 Moving and renaming directories  
3.18 How to remove files  
3.19 How to copy files and directories  
3.20 Working with text files  
3.21 Introduction to aliases  
3.22 Editing text files  
3.23 Automating Unix commands  
3.24 How to hide files and find hidden files  
3.25 Creating a configuration file  
3.26 Programming with Unix  
3.27 Unix file permissions  
3.28 How to specify which directories contain programs  
3.29 Creating useful shell scripts  
3.30 Unix summary
### Part 4 Essential Perl

1. **Hello world**  
2. **Scalar variables**  
3. **Use warnings**  
4. **Maths and functions**  
5. **Perl vs. perl**  
6. **Conditional statements**  
7. **Use strict**  
8. **Stopping programs**  
9. **Working with strings**  
10. **Dealing with special characters**  
11. **Matching operators**  
12. **The transliteration operator**  
13. **List context**  
14. **Introduction to Arrays**  
15. **Array manipulation**  
16. **The @ARGV array**  
17. **Defined and undefined variables**  
18. **Sorting**  
19. **Introduction to loops**  
20. **More loops**  
21. **Loop control**  
22. **Data input and output**  
23. **Reading and writing files**  
24. **Introduction to hashes**  
25. **Working with hashes**  
26. **Introduction to regular expressions**  
27. **Regular expression metacharacters**  
28. **Working with regular expressions**  
29. **Interacting with other programs**  
30. **Using functions and subroutines**  
31. **Returning data from a subroutine**  
32. **Part 4 summary**

### Part 5 Advanced Unix

1. **Introduction to advanced Unix**  
2. **Introduction to process control**  
3. **The grep command**  
4. **Viewing and controlling program output**  
5. **Redirecting input and output**  
6. **Standard error**  
7. **Connecting commands with pipelines**  
8. **Advanced text manipulation**
## Contents

**Part 6 Advanced Perl**  
6.1 Regular expressions revisited  
6.2 Function libraries  
6.3 References and two-dimensional arrays  
6.4 Records and other hash references  
6.5 Using references with subroutines  
6.6 Complex data structures  
6.7 Adding command-line options  
6.8 OOP basics  
6.9 CPAN  

**Part 7 Programming topics**  
7.1 Debugging strategies  
7.2 Common error messages  
7.3 Code beautification  
7.4 Abstraction  
7.5 Data management  
7.6 Documentation  
7.7 Revision control  
7.8 Working with other people’s data  
7.9 Getting help  

*Appendix*  
*Index*