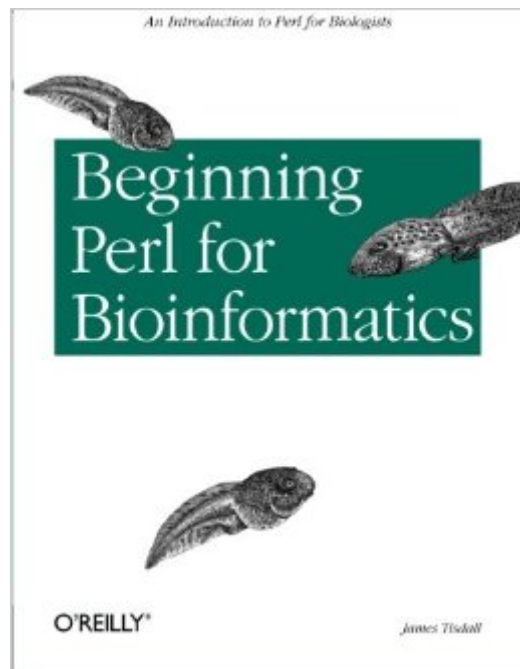


The book was found

# Beginning Perl For Bioinformatics



## Synopsis

With its highly developed capacity to detect patterns in data, Perl has become one of the most popular languages for biological data analysis. But if you're a biologist with little or no programming experience, starting out in Perl can be a challenge. Many biologists have a difficult time learning how to apply the language to bioinformatics. The most popular Perl programming books are often too theoretical and too focused on computer science for a non-programming biologist who needs to solve very specific problems. *Beginning Perl for Bioinformatics* is designed to get you quickly over the Perl language barrier by approaching programming as an important new laboratory skill, revealing Perl programs and techniques that are immediately useful in the lab. Each chapter focuses on solving a particular bioinformatics problem or class of problems, starting with the simplest and increasing in complexity as the book progresses. Each chapter includes programming exercises and teaches bioinformatics by showing and modifying programs that deal with various kinds of practical biological problems. By the end of the book you'll have a solid understanding of Perl basics, a collection of programs for such tasks as parsing BLAST and GenBank, and the skills to take on more advanced bioinformatics programming. Some of the later chapters focus in greater detail on specific bioinformatics topics. This book is suitable for use as a classroom textbook, for self-study, and as a reference.

The book covers:

- Programming basics and working with DNA sequences and strings
- Debugging your code
- Simulating gene mutations using random number generators
- Regular expressions and finding motifs in data
- Arrays, hashes, and relational databases
- Regular expressions and restriction maps
- Using Perl to parse PDB records, annotations in GenBank, and BLAST output

## Book Information

Paperback: 400 pages

Publisher: O'Reilly Media; 1 edition (November 1, 2001)

Language: English

ISBN-10: 0596000804

ISBN-13: 978-0596000806

Product Dimensions: 7 x 0.8 x 9.2 inches

Shipping Weight: 1.2 pounds (View shipping rates and policies)

Average Customer Review: 4.4 out of 5 stars [See all reviews](#) (46 customer reviews)

Best Sellers Rank: #212,523 in Books (See Top 100 in Books) #15 in [Books > Computers & Technology > Programming > Languages & Tools > Perl](#) #51 in [Books > Computers &](#)

## Customer Reviews

Finally someone has written a beginning book on PERL for biologists, and has also done an excellent job of doing so. This book assumes no prior programming experience, and therefore suits the biologist who needs to concentrate on using computers to solve biological problems, and not have to become a computer scientist in the process. PERL can be a very cryptic language, but it is also extremely concise, and PERL programmers frequently and rightfully boast about their "one-liners" that accomplish complicated tasks with only one line of code. Since it is addressed to readers with no programming experience, the author introduces some elementary concepts of programming in the first three chapters. These include what text editor to use, how to install PERL, how run PERL programs, and other relevant elementary topics. The author then gets down to writing a program to store a DNA sequence in chapter 4. Very basic, it merely reads in a string and prints it out, but serves to start readers on their way to developing more useful programs. Later a program for the transcription of DNA to RNA is given, which illustrates nicely the binding, substitution and trace operators. Block diagrams are used here, and throughout the book, to illustrate basic PERL operators. The author shows in detail how to read protein sequence data from a file and how to use it in a PERL program. The reader is also introduced to the most ubiquitous data structure in all of computing: the array. Already the reader gets a taste of the power of PERL to manipulate arrays, using operations such as 'unshift', 'push', 'splice', etc. The next chapter introduces conditional statements in PERL, as a warm-up for the discussion on finding motifs in sequences. The reader can see why PERL is the language of choice in bioinformatics, with its ability to find substrings or patterns in strings. Things do become more cryptic in the discussion of regular expressions, but the reader can get through it with some effort. Interesting programs are given for determining the frequency of nucleotides. Since the programs have become more complicated to this point, a discussion of subroutines follows in the next chapter. And, for the same reason, the reader is introduced to debugging in PERL in this chapter also. The greater the complexity of the program, the harder it becomes to avoid making mistakes, and even more difficult to find them. The very important concepts of pass by value versus pass by reference are discussed briefly in this chapter. Random number generators, so important in any consideration of mutations, are discussed in chapter 7. It is shown, via some straightforward programs, how to select a random location in DNA and mutate it with some other nucleotide. In addition, the author shows how to use random

numbers to generate DNA sequences and mutate them in order to study the effect of mutations over time. The next chapter is the most interesting in the book, for it shows how PERL can be used to simulate how the genetic code directs the translation of DNA into protein, the hash data structure being used extensively for this purpose. The author shows how to read DNA from files in FASTA format, and discusses in detail reading frames. He gives a useful subroutine to translate reading frames. The author returns to regular expressions in chapter 9, wherein they are used as 'wildcards' to search for a particular string in a collection of strings. In addition, the range operator is used to find restriction sites. Regular expressions are also used in the next chapter to manipulate GenBank 'flat files'. The author does however give URLs for more sophisticated bioinformatics software. This is followed in chapter 11 by a discussion of the use of PERL to work with files in the Protein Data Bank. Recursion, one of the most powerful techniques in programming, is introduced here. Chapter 12 covers the Basic Local Alignment Search Tool (BLAST), wherein readers get a taste of the field of computational biology. This extremely popular software package is used to find similarity between a given sequence and a library of known sequences. The author does discuss some of the basic rudiments of string matching and homology, and encourages the reader to consult the BLAST documentation for further details. In addition, the author briefly discusses the Bioperl project in this chapter, and shows the reader how to run some elementary computations using it. This book definitely is a timely one and it will serve the needs of biologists who need to obtain some programming expertise in PERL. There are helpful exercises at the end of each chapter that serve to solidify the understanding of the concepts introduced in the chapter. After a thorough study of it, readers will be well-equipped to use PERL in bioinformatics. With more mathematical background, readers after finishing it will be able to enter the exciting field of computational biology, a field that is exploding, and one in which will require imaginative programming skill in the future.

"Bioinformatics" is the new sexy term for what used to be called simply "computational biology". Simply put, it involves pretty much any application of computation techniques to biological problems. The reason for the new nomenclature and the greatly increased interest in the topic is, like much in modern biology, a more-or-less direct consequence of the many genome sequencing projects of the last decade. The consensus in the field seems to be that it's more productive (and certainly easier) to teach biologists how to program, rather than try to get programmers up to speed on the intricacies of molecular biology. For similar reasons, Perl is a popular language to learn: it's easy to get off the ground and be productive with it, without requiring a heavy computer science background. (This, of course, has downsides as well...) Never one to miss out on a trend, I'm going to be teaching a

course on Bioperl and advanced Perl programming, starting next fall, which means I'm doing a lot of reading in this topic area, trying to develop lectures and find good background reading material. One of the first books I grabbed was Beginning Perl for Bioinformatics, which has been sitting on my "to read" shelf since O'Reilly sent me a review copy in December of 2001. It's a typical O'Reilly "animal" book (the cover bears three tadpoles), which does a decent job of introducing the basic features of the Perl language, and it should enable a dedicated student to get to the point where she can produce small useful programs. However, I'm not completely happy about the book's organization, and I think the occasional "if you're not a biologist, here's some background" interjections could have been cut without hurting anything. The initial chapters in the book cover "meta" information, such as theoretical limits to computation, installing (or finding) the Perl interpreter on your computer, picking a text editor, and locating on-line documentation. Some general programming theory stuff is covered as well -- the code-run-debug cycle, top-down versus bottom-up design, the use of pseudocode. There's also some biology background, but it's very introductory level stuff -- DNA has four bases, proteins are made of 20 amino acids, and so on. In chapter four, the book begins to get into actual Perl, with some coverage of string manipulation. Examples deal with simulating the transcription of DNA into RNA. Chapters five and six continue to flesh out the language, covering loops, basic file I/O, and subroutines. Chapter seven introduces the `rand()` function, in the context of simulating mutations in DNA. Subsequent chapters introduce the hash data type (using a RNA->protein translation simulation), regular expressions (as a way to store the recognition patterns of restriction endonucleases), and parsing database flat files and BLAST program output. I'm clearly out of the target audience of the book, as I already have a strong working knowledge of Perl. Perhaps that's why I found the order that concepts were presented in to be a bit strange -- for example, hashes, which are a fundamental data type, aren't introduced until halfway through the book, and regular expressions (one of the key features of Perl) first appear even later. As I said above, I also found the biological background sections to be more distracting than anything, but I've also got a strong biology background, so perhaps I'm off base here too. That said, I think a person with a CS background would be better served with a copy of Learning Perl and an introductory molecular biology text than with this particular book. One of the things I did enjoy about the book were the frequent coding examples, all of which presented realistic computational biology sorts of problems and then demonstrated how to solve them. I'm sure that when I get around to writing lectures, I'll be leafing through this book looking for problems I can use in class. Overall, recommended for biologists without programming experience who would like to get started using Perl for simple programming. Not recommended for people with computer science

backgrounds looking to get into bioinformatics.

[Download to continue reading...](#)

Beginning Perl for Bioinformatics Building Bioinformatics Solutions: with Perl, R and MySQL  
Mastering Perl for Bioinformatics Beginning Perl Web Development: From Novice to Professional  
(Beginning: From Novice to Professional) Perl: Crash Course - The Ultimate Beginner's Course to  
Learning Perl Programming in Under 12 Hours Effective Perl Programming: Ways to Write Better,  
More Idiomatic Perl (2nd Edition) (Effective Software Development Series) Mastering Perl/Tk:  
Graphical User Interfaces in Perl Perl Programming Success in a Day: Beginners Guide to Fast,  
Easy, and Efficient Learning of Perl Programming Programming the Perl DBI: Database  
programming with Perl Beginning Perl Beginning Perl (Expert's Voice in Open Source) Beginning  
Perl (Programmer to Programmer) Building Bioinformatics Solutions 2nd edition Structural  
Bioinformatics Bioinformatics: The Machine Learning Approach, Second Edition (Adaptive  
Computation and Machine Learning) Bioinformatics and Functional Genomics Understanding  
Bioinformatics Bioinformatics and Computational Biology in Drug Discovery and Development  
Bioinformatics For Dummies Discovering Genomics, Proteomics and Bioinformatics (2nd Edition)

[Dmca](#)