

# Tutorial: Beginning Perl for Bioinformatics

James Tisdall  
Biocomputing Associates  
P.O. Box 413  
Kimberton, PA 19442  
610 933-1200

April 11, 2002

## 1 Description

This tutorial is designed to introduce programming to biologists. It presents the basics of Perl programming using biological data. Attendees should possess at least a basic knowledge of elementary molecular biology. The tutorial does not presuppose any programming experience, although it is also suitable for biologists with some programming experience in another language. The goal is to make non-programming biologists capable of writing useful Perl programs for use in their own labs.

After the tutorial the attendees will be prepared to attempt to:

- Write and run Perl programs on their computers
- Write programs that perform useful work in the research lab
- Store biological data in the basic Perl data structures: scalars, arrays, and hashes
- Search for motifs in biological sequence data using regular expressions
- Read and write fasta files, and other formats such as GenBank, PDB, and BLAST
- Implement the genetic code
- Employ subroutines and modules
- Use the Bioperl modules
- Control other bioinformatics programs

They will also be prepared for a more complete study of Perl, and for the use of collections of modules such as Bioperl (the subject of another tutorial at ISMB2002) in their programs.

The tutorial handout will be a copy of my book *Beginning Perl for Bioinformatics* from O'Reilly and Associates. The tutorial will cover some of the material from the first six chapters. You may examine the book at the web site <http://www.oreilly.com/catalog/begperlbio/>, which includes the table of contents and a sample chapter, as well as solutions to most of the exercises, and downloadable versions of all the example programs from the book.

## 2 Biographical Description

I have been working in bioinformatics since 1991, when I joined the Computational Biology and Informatics Laboratory at the University of Pennsylvania, where I was a programmer and systems manager in the Human Genome Project, and became one of the first to use the Perl computer language in bioinformatics, in my program *DNA WorkBench* and other programs. In 1994-1995 I joined Mercator Genetics in Menlo Park, California, as computational biologist, and then went to Fox Chase Cancer Center in Philadelphia, Pennsylvania, as the manager of bioinformatics, from 1995-2000. Since the end of 2000 I've been working as a consultant for Biocomputing Associates of Kimberton, Pennsylvania (<http://www.biocomputingassociates.com>), consulting on bioinformatics in medical research.

In October 2001 O'Reilly and Associates Publishers released my book *Beginning Perl for Bioinformatics*. I am currently writing a follow-up volume tentatively titled *Mastering Perl for Bioinformatics* for release in 2002. I have also recently coauthored scientific papers, and published a few tutorial articles: *Why Biologists Want to Program Computers*, *Parsing Protein Domains With Perl*, *Beginning Bioinformatics for Perl Programmers*. In January 2002 I presented a similar tutorial at the O'Reilly Bioinformatics Conference in Tucson, Arizona. I have also delivered training courses on this topic at Fox Chase Cancer Center and at Mercator Genetics.

I have a B.A. in mathematics from City College of New York, a M.S. in computer science from Columbia University, and have completed the course work and qualifying exams for the PhD in computer science at the University of Pennsylvania. I have publications in the fields of bioinformatics, complexity theory, and linguistics. I also worked for six years as a programmer in the Speech Research Department of Bell Laboratories in Murray Hill, New Jersey, and have been a fellow at the University of Michigan and at the University of Pennsylvania.