

Introduction to Bioinformatic Computation. BIPG 610/810

Summary: Bioinformatics is a fundamental component of modern biomedical sciences. Only computers have the capability to collect, organize, annotate, and process the enormous amounts of information about the organization and structure of the biosphere. Familiarity with a computer language is essential to those who seek expertise in bioinformatics. The main goal of this course is to provide basic programming skills to biological and medical students who may lack a background in computer sciences. Programming will be taught using important biological examples. This course focuses in particular on PERL because it is easy to acquire and is the most commonly-used language used in genomics and database analysis. Since UNIX is the dominant platform in present-day computational biology, students will be made familiar with UNIX environments, including Linux and OSX. Each registered student will be given an account on the MUO Linux computer cluster, the most powerful computer within the College. Hands-on programming experience will be obtained in the Bioinformatics Computer Lab at UT. Students are encouraged to bring their own problems and projects to work on during this course. In addition to learning programming and other bioinformatics skills, the students of this course acquire knowledge in how to present the final product of bioinformatics research and how to write a scientific paper on the subject. This work resulted in the publication of four articles listed below.*

Instruction: The course is taught only by Alexei Fedorov, and may be assisted by his graduate students.

Course director: Dr. Alexei Fedorov, Dept. of Medicine, Vice Director BPG Program, Director of Bioinformatics Lab, (419) 383-5270, alexei.fedorov@utoledo.edu.

Format: Conventional face-to-face meetings only (no video lectures/labs. Those alternatives are available on U-tube for Linux and Perl from multiple sources.)

Office hours: Every Monday and Wednesday from 9 AM to 10 AM in the office of Dr. Fedorov (Room 0012, Ruppert Center, HSC; tel: 419-383-5270). Also students may contact remotely via Skype (Afedorov_lab) every Monday from 8 PM to 9 PM.

Homework time policy: Each homework assignment must be returned in seven days by noon. (For example, for a Wednesday class this homework must be returned via e-mail next Wednesday by 12 pm.) Absolutely NO excuse for a late homework return (automatic F).

Topics covered:

- We start first two weeks with learning Linux. Students must learn the following 30 commands with their main options (link to the “magic words”). After one month, there will

be an exam on Linux. Here is an example of the questions from the 2009 course exam on Linux: (link to file)

- From the second week of the course we start learning Perl. In the first half of the course (till Spring-break, ~March 03-10) the students must learn the following aspects in PERL:

- 1) Variables (numeric and strings)
- 2) Loops (while, for, redo block)
- 3) If-else control statements
- 4) Arrays
- 5) Hashes
- 6) Multiple arrays, anonymous arrays
- 7) Array of hashes, hash of arrays
- 8) Regular expressions
- 9) System calls to invoke various programs inside Perl scripts
- 10) Subroutines
- 11) Packages and modules
- 12) References

From the beginning of the second half of the course we start a real bioinformatics project. All students will work as a single team. Our goal is to make a real research and attack one of the important problems/challenges in genomics. In two months we should try our best and collect data for a publication. The most active students will be among the co-authors. Below* is the list of papers that have been published by students in this course. Here are the links to several programs that have been written by the students from this course:

snoRNA.pl
NTcomposition.pl
AMI_bbackoff2011.pl (intron, home)

Grading principles:

- Homework/project 40%
- LABs+ activity 20%
- Mid-term Exam 15%
- Final Exam 25%

* **Published Articles:** (click on links to view each article)

1. Computer Identification of snoRNA Genes Using a Mammalian Orthologous Intron Database, Nucleic Acids Research 2005, 33:4578-4583 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=16093549>

2. Alternative Splicing Mutation Database: A Hub for Investigations of Alternative Splicing Using Mutational Evidence, BMC Research Notes 2008, 1:3 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=18>

[611286](#)

3. Calculation of Splicing Potential from the Alternative Splicing Mutation Database, BMC Research Notes 2008,

1:4 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=18611287>

4. Genomic Mid-Range Inhomogeneity Correlates with an Abundance of RNA Secondary Structures, BMC Genomics 2008,

9:284 <http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pubmed&pubmedid=18549495>