

**Syllabus: CS 5984**  
**Computation for the Life Sciences**  
**Fall, 2002**

<http://courses.cs.vt.edu/~cs4life/>

**Instructors:**

	<b>Lenwood S. Heath</b>	<b>Naren Ramakrishnan</b>
<b>Office</b>	638 McBryde Hall	629 McBryde Hall
<b>Office Hours</b>	MW 8-10	TTh 10-12
<b>EMAIL</b>	heath@vt.edu	naren@cs.vt.edu

**Class Meets: McBryde 210, MWF 11:15–12:05**

**Index Number: 96701**

**Prerequisites:**

- Graduate standing in a life sciences department or permission of the instructors.
- Previous coursework in genetics, molecular biology, and cell biology.
- Not open to graduate students in computer science.

**Textbook and other materials:**

- *Beginning Perl for Bioinformatics*, James Tisdall, O'Reilly, 2001.
- *A First Course in Database Systems*, Jeffrey D. Ullman and Jennifer Widom, Prentice Hall, Second Edition, 2001.
- Lecture notes and other materials available through the course web site.

**On Reserve:**

For current list, see class web site.

**Description**

This course develops the background needed by graduate students pursuing the bioinformatics option or the PhD in Genetics, Bioinformatics, and Computational Biology but not having a computer science background. Fundamentals of computer science discussed and practiced include Linux, data, and the importance of the command line interface; the Perl programming language; program design, implementation, and testing; field-structured, semi-structured, and relational databases; basic data structures, algorithms, and algorithm analysis; and data mining.

## Grading Policy

Grading for the course is on a 1000-point scale, with the points distributed as follows:

<b>Homework assignments:</b> 8 at about 50 points each	400
<b>Programming and database assignments:</b> 6 at about 50 points each	300
<b>Exams:</b> 3 at about 100 points each (evenings)	300
<b>Total:</b>	1000

See the course calendar for details.

A typical homework assignment consists of 2 or 3 problems. All homework must be prepared with L<sup>A</sup>T<sub>E</sub>X or other word processing system. Homework may be submitted as a stapled printout in class on the date due or may be emailed to the instructor making the assignment as a PDF file prepared by Adobe Distiller.

A typical programming or database assignment consists of 2 or 3 programs or database queries to write and test. A programming or database assignment must be prepared with a text editor (*vi*, *pico*, or equivalent) and submitted by email to the instructor making the assignment, together with the input and output files from testing.

Each assignment is due at 5:00 PM on the due date. **No late assignments will be accepted.**

## Ethics

The Honor Code applies. All work submitted must be the student's own work. Students may solicit help only from the instructor.

## Announcement

If any student needs special accommodations because of a disability, please contact the instructor during the first week of classes.

**Topics**

<b>WEEK</b>	<b>TOPIC</b>	<b>INSTRUCTOR</b>
1	Introduction to computation for bioinformatics	LSH
1–2	Introduction to Linux: logging in; the command line; pipes; text editing — <code>pico</code> and <code>vi</code>	LSH
3	Introduction to Perl: the Perl interpreter; running Perl; <code>make</code> ; scalar data and variables	LSH
4	Perl: standard input, standard output, and standard error; control structures; programming strategies	LSH
5	Perl: subroutines, arrays, and random mutations	LSH
6	Perl: hashes; regular expressions; parsing web pages (HTML)	LSH
7	Field-oriented databases: GenBank <i>et al.</i> ; parsing with regular expressions	NR
7–8	Semi-structured databases: PubMed <i>et al.</i> ; parsing with regular expressions	NR
8–11	Relational databases: schemas; queries; queries via Linux; queries via Perl	NR
12–13	Data structures, algorithms, and algorithm analysis: examples using Perl	LSH
14	Data mining, inductive logic programming, and related topics	NR

END OF SYLLABUS