Emerging HPC Technologies

Computer Information Systems Division
Wake Technical Community College

HPC 272 (3-0-3)	Chris James, Instructor
Course Syllabus Addendum	Fall Semester, 2003
Bioinformatics Certificate	Office: LeMay #21
Phone: 662-3416 (Secretary)	Office Hours: M,T,TH 5-6 PM
Email: ctjames@waketech.edu	Office Phone: 662-3656

Enrollment Information: This is a restricted enrollment course. The course is open to individuals possessing course work in the field of biology (baccalaureate degree recommended) or related sciences who wish to pursue the Bioinformatics certificate offered through Wake Technical Community College's High Performance Computing curriculum.

The Bioinformatics certificate is aimed at individuals who:

- are working in the field of biology and require bioinformatics skills.
- are enrolled in an undergraduate biology curriculum who wish to gain knowledge in the field of bioinformatics for later employment opportunities.
- wish to obtain academic preparation for the information technology portion of the Certified Bioinformatics Specialist certification offered by the National Bioinformatics Institute.

Other students may be admitted to the course by prior arrangement with the course instructor. Email the course instructor (ctjames@waketech.edu) for additional information and registration procedures regarding this course or for more general information regarding the Bioinformatics certificate program.

Course Duration: This course meets for the first 8 weeks of the Fall semester (August 19, 2003 - October 13, 2003).

Class Meeting Time: This is a hybrid web-based course combining distance education with on campus class meetings. The course contact hours are equally divided between on campus meetings an online access. The online portion of the coursed is accessed through a web browser based interface known as Blackboard. The on-campus meetings will be on Monday nights from 6:00 PM to 9:30 PM. The first on campus meeting will be **August 25, 2003**.

Course Catalog Description: This course introduces students to emerging technologies in the field of High Performance Computing (HPC). Emphasis is placed on the new technologies in the HPC field and a review of HPC and cluster systems already implemented. Upon completion, students should be able to discuss, in written and oral form emerging technologies in the HPC field.

Course Overview: This course is a survey course introducing the emerging field of bioinformatics. Bioinformatics is an interdisciplinary field that uses computer algorithms to interpret biological data, such as DNA or protein sequences or structures. The course assumes a background knowledge in biology.

Topics include the use of computers in biology, computational approaches to biological problems, the use of a UNIX workstation in bioinformatics, biological tools available to bioinformaticians, the Perl programming language, and the use of databases and visualization in bioinformatics.

Course Textbook: "Developing Bioinformatics Computer Skills", by Cynthia Gibas and Per Jambeck, O'Reilly publishers, ISBN: 1565926641. From the publisher's website: "*Developing Bioinformatics Computer Skills will help biologists, researchers, and students develop a structured approach to biological data and the computer tools they'll need to analyze it. The book covers the Unix file system, building tools and databases for bioinformatics, computational approaches to biological problems, an introduction to Perl for bioinformatics, data mining, data visualization, and tips for tailoring data analysis software to individual research needs*."

Course Goals: The principal goals for the course are as follows:

1. Understand how computers are used in the field of bioinformatics.
2. Examine the design philosophy of the UNIX operating system
3. Overview the major public biological databases available to researchers.
4. Introduce the use of the Perl programming language in bioinformatics.
5. Examine the use of databases for data mining and programs for data visualization in bioinformatics.

Course Objectives: Upon successful completion of the course, students will be able to:

1. Cite examples of how computers are used in the field of bioinformatics.
2. Execute simple commands using the UNIX command line.
3. Differentiate among the content of the major publicly available biological databases.
4. Execute simple commands in the structured query language (SQL), the standardized language used to manipulate databases.
5. List examples of common graphical visualization tools available to researchers.

Grade Evaluation: Two tests plus a comprehensive final, all of equal value (10 point grading scale).

Testing: All tests will be closed-book take home tests. The format of the tests will vary, but expect multiple choice, matching, fill in the blank type, and short answer type questions.

Course Outline:

Part I. **The Bioinformatics Workstation**

UNIX is the preferred operating system for running bioinformatics applications, both in academic and industrial organizations. This section will provide a general overview of UNIX operating system, the strengths of UNIX, and working on the UNIX command line.

Text chapters:

3. Setting Up Your Workstation
4. Files and Directories in Unix
5. Working on a Unix System

Part II. **The Perl Programming Language**

The Perl programming language is the most commonly used programming language used in association with bioinformatics. This section will describe the features of Perl which make this programming language so valuable to bioinformatic computing.

Text chapters:

12. Automating Data Analysis with Perl

Part III. **Computers and Biology**

The first part of the course introduces the fundamentals concepts of informatics to the field of biology. This section will also include a brief review of important biological concepts that are the focus of bioinformatics. Review topics will include nucleic acid structure, the genetic code, structure and function of genes, organization of the genome, regulation of gene expression, transcription, translation and protein structure and function.

Text chapters:

1. Biology in the Computer Age
2. Computational Approaches to Biological Questions

Part IV. **Tools for Bioinformatics**

This section will provide an introduction to the most commonly used nucleic acid and protein databases and analysis tools on the Internet.

Text chapters:

6. Biological Research on the Web
7. Sequence Analysis, Pairwise Alignment, and Database Searching
8. Multiple Sequence Alignments, Trees, and Profiles
9. Visualizing Protein Structures and Computing Structural Properties
10. Predicting Protein Structure and Function from Sequence
11. Tools for Genomics and Proteomics

Part V. **Databases and Data Visualization**

The final section of the course will introduce data storage and data visualization concepts. Powerful data management tools and computational techniques are required to store, share, analyze and compare the exponentially increasing store of biological information. Bioinformatics combines the tools of mathematics, computer science and biology with the aim of uncovering patterns and associations within and between sets of biological data.

Text chapters:

13. Building Biological Databases
14. Visualization and Data Mining

Instructional Methods: This is a hybrid web-based course combining distance education using Blackboard software with on-campus class meetings.

Course Webpage: All instructional materials for this course are available from the following location:

`http://dist-ed.waketech.edu/`

Course Orientation: There is no orientation meeting for this course outside of the regular on-campus meetings.

Course Access: Students will be able to access the Blackboard interface to the course on the first day of the semester (**August, 19, 2003**).

Getting Started: You should visit the course webpage after the first on-campus meeting.