Advanced HPC Data Mining
Computer Information Systems Division
Wake Technical Community College

HPC 270 (2-2-3)	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

Course Catalog Description: This course provides an introduction to data intensive computing on HPC machines. Topics include distributed mass storage, efficient retrieval techniques, data management tools, appropriate data structures and case studies. Upon completion, students should be able to define and discuss performance evaluation of a database in a HPC environment.

Course Overview: This course involves a survey introduction to major public biological databases. Introduction to public domain DNA sequence databases, use of software and Internet resources for database searching, use of database information in sequence comparisons, and sequence alignment. Continuation of the use of Perl programming in the field of bioinformatics.

Course Textbook: "Bioinformatics for Dummies", by Jean-Michel Claverie and Cedric Notredame, John Wiley & Sons, ISBN#: 0764516965. From the publisher's description: "*A practical introduction to bioinformatics and computer technologies that biochemical and pharmaceutical researchers use to analyze genetic and biological data.*"

Course Goals: The design goals of the course include the following:

1. Provide an introduction to biological databases as a means to interpret the rapidly expanding amount of biological information.
2. Compare the various public domain biological databases.
3. Become knowledgeable about the storage, retrieval, sharing and use of biological data, information, and tools.
4. Submit DNA sequence information to the databases.
5. Retrieve information from the databases.

Course Objectives: Upon successful completion of this course, the student will be able demonstrate competency in the following:

1. Retrieve gene sequence information from GenBank.
2. Use BLAST program to conduct gene similarity searches.
3. Perform sequence analysis.
4. Align multiple sequences with Clustal W program.
5. Write Perl programs to process sequence data.

Course Outline:

Part I.   Biological Databases Overview
Part II.   Nucleotide Sequence Databases
Part III.   Nucleotide Sequence Analysis
Part IV.   Protein Sequence Databases
Part V.   Protein Sequence Analysis

Grade Evaluation:

Two tests, a comprehensive final, and laboratory, all of equal value (ten-point grading scale).

Course Duration: This course meets for the second 8 weeks of the fall semester (October 20, 2003 - December 16, 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. A Blackboard tutorial will be given on the night of the first class meeting.

Room Location: On campus meetings will be in room TB-219.

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 second fall semester session (**October 20, 2003**).

Login ID: Your login name will be your first initial, middle initial and last name, plus the last two digits of your student ID. Your password will be your student ID. Additional information and trouble shooting your login ID is available here:

http://www.wake.tec.nc.us/dist_ed/internet/help.html

Getting Started: You should visit the course webpage after the first on campus meeting. Students who register after classes start will not have immediate access to the class.