

BME 295/382 & CSE 298/478: Computational Genomics Spring 2008

Time/place: TuTh 2-3:15pm / ITE 119

Instructor: Ion Mandoiu

Office: ITE 261

Phone: 486-3784

E-mail: ion@engr.uconn.edu

Office hours: MW 12-1pm or by appointment

Course outline: Started in 1995 by the completion of the first genome sequence of a free-living organism, *H. influenzae*, the genomic era has led to hundreds of complete genome sequences deposited in public databases and many more genome projects at various stages of completion. The large-scale availability of genome data is revolutionizing biological and medical research, with data-driven computational approaches taking a central role. This course covers fundamental computational methods for genomic data analysis, with a main emphasis on statistical methods and current applications in genomics and genetic epidemiology.

Tentative list of topics to be covered: Basic probability theory and statistics; introduction to R; statistical modeling of biological sequences; EM and Gibbs sampling algorithms for DNA motif discovery; motif site scoring and p-values; Markov chains; profile HMMs for representing sequence families; models of DNA and protein evolution; likelihood methods in phylogenetics, bootstrapping; basic principles population genetics; genotype phasing and haplotype frequency estimation; computation of Mendelian likelihoods, Elston-Stewart and Lander-Green algorithms; case-control association tests. The topics may change according to progress.

Textbooks: There is no required textbook for this course. Most of the covered material appears in the following books:

- R.C. Deonier, S. Tavaré, M.S. Waterman, *Computational genome analysis: an introduction*, Springer Verlag, 2005.
- R. Durbin, S. Eddy, A. Krogh, G. Mitchison, *Biological sequence analysis: probabilistic models of protein and nucleic acids*, Cambridge University Press, 1998.
- K. Lange. *Mathematical and Statistical Methods for Genetic Analysis*, 2nd ed., Springer Verlag, 2002.

Prerequisites: Desired pre-requisites are an introductory course in bioinformatics or undergraduate-level courses in biology, programming, and statistics. However, required background in these areas will be provided as needed via lectures and supplementary readings.

Grading: Grades will be assigned based on bi-weekly homework assignments (50%) and a final project (50%). Homework assignments will consist of both theoretical problems reinforcing the material covered in lectures and practical assignments using the R statistical package. The subject of the final project can be either a topic covered in the course or related topics approved by the instructor. The final project will give you the opportunity to study a computational genomics problem in more depth. Suitable final project topics include surveys of computational genomics topics not covered in the lectures, design and implementation of novel algorithms, and empirical evaluation of existing methods. A list of potential topics will be provided towards the middle of the semester, although you are encouraged to devise your own. You will be required to submit a written final report of 15-20 pages and give a short presentation on your project at the end of the semester. When appropriate, projects can be done in teams.

HuskyCT: We have a HuskyCT (formerly Vista/WebCT) site for the class; you can access it by logging in with your NetID and password at <http://vista.uconn.edu/>. Please check this site regularly for class materials, grades, assignment clarifications, changes in class schedule, and other class announcements.

Academic honesty: You are expected to adhere to the highest standards of academic honesty. All submitted solutions must be your own work. You may discuss ideas and concepts with other people, but *must not share written solutions or computer code*. Use of published materials is allowed, but the sources should be explicitly stated in your submission. Violations will be reviewed and sanctioned according to the University Policy on Academic Integrity.

Students with disabilities: If you have a documented disability for which you are or may be requesting an accommodation, you are encouraged to contact the instructor and the Center for Students with Disabilities or the University Program for College Students with Learning Disabilities as soon as possible to better ensure that such accommodations are implemented in a timely fashion.