Bioinformatics Symposium

Ottawa-Carleton Joint Collaborative Program in Bioinformatics

September 5, 2008, at the University of Ottawa

The Ottawa-Carleton Institute of Biology, the Ottawa-Carleton Institute of Computer Science, the Ottawa-Carleton Institute of Mathematics and Statistics, and the Biochemistry (BCH) graduate program and the Microbiology & Immunology graduate (MIC) program of the Department of Biochemistry, Immunology and Microbiology (BMI) at the University of Ottawa are organizing a one-day symposium to mark the beginning of the joint collaborative MSc program in Bioinformatics, and the start of the academic year.

The idea behind this symposium is essentially to create a graduate and post-graduate community of people who work in bioinformatics and related fields. Consequently, this symposium is open to anybody and is not limited to students enrolled in the Bioinformatics program!

The symposium will be held on **Friday September 5, 2008, from 9-4:30 in Gendron 080** at the University of Ottawa (see map on last page). There will be an onsite BBQ for lunch, organized by the Biology Graduate Students' Association of the University of Ottawa.

Registration is free but mandatory . Graduate students (MSc, PhD) and PDFs with interests in Bioinformatics are strongly encouraged to register as soon as possible, and to give a presentation. each presentations will be 30 minutes long (*including* discussion).

Inquiries and registration should be directed to Stephane Aris-Brosou (`sarisbro@uottawa.ca`). Details are also available online at `http://aix1.uottawa.ca/~sarisbro/`.

1 Program

Time	Presenter	Affiliation	Title
09:00-09:30	Stéphane Aris-Brosou	BIO	Opening remarks.
09:30-10:00	Yubing Liu	BMI	Genomic characterization of six1 function during skeletal muscle differentiation.
10:00-10:30	Zahra Montazeri	BMI	Statistical identification of regulatory relationships between genes.
10:30-11:00	**refreshments**		
11:00-11:30	Pei-Chun Hsieh	BMI	Gene network inference applied to yeast data.
11:30-12:00	Nicolas Rodrigue	BIO	Bayesian models of protein-coding nucleotide sequence evolution with heterogeneous preferences at the amino acid level.
12:00-01:30	**lunch break**	(BBQ)	*In the courtyard between Marion and CAREG*
01:30-02:00	Predrag Mizdrak	CS	Iterative multiple sequence alignment (MSA) and phylogeny inference using a maximum likelihood (ML) method.
02:00-02:30	Xiaoquan Yao	BIO	The nucleotide following the initiation codon does not influence translation efficiency in yeast – from a regression analysis.
02:30-03:00	Rob Carter	BIO	The extent of concerted evolution of eukaryotic genes affects the evolutionary rates of its transcribing polymerase and associated general transcription factors.
03:00-03:30	**refreshments**		
03:30-04:00	Robert Davies	MATH	Eliminating redundant SNPs improves power in genome wide association studies
04:00-04:30	Michel Dumontier	BIO	Facilitating biological knowledge discovery with semantic web technologies.
04:30-04:45	Stéphane Aris-Brosou	BIO	Closing remarks.

2 Abstracts

1. **Rob Carter**, *The extent of concerted evolution of eukaryotic genes affects the evolutionary rates of its transcribing polymerase and associated general transcription factors*: The genes in eukaryotic genomes are partitioned into three classes, each transcribed by one of three DNA-dependent RNA polymerases. The genes in each of these classes have different propensities for evolving in a concerted fashion. Here we show that the transcriptional machineries evolve at a rate proportional to the extent of concerted evolution affecting its target genes. The consequences of rapidly evolving transcriptional machineries on identifying the origin and evolutionary relationships between subunits of the three RNA polymerases will also be discussed.

2. **Robert Davies**, *Eliminating redundant SNPs improves power in genome wide association studies*: Genome Wide Association analyses involve testing hundreds of thousands of Single Nucleotide Polymoprhisms (SNPs) for association with disease; as such, meaningful analysis requires corrections for multiple testing. However, as many SNPs show similar characteristics, applying corrections which assume independence is overly harsh. Here, we aim to block together and eliminate redundant SNPs using population specific linkage measures from control genotype data.

3. **Michel Dumontier**, *Facilitating biological knowledge discovery with semantic web technologies*: Bioinformatics will forever be burdened with the challenges of managing highly dynamic biological data. Over the past 20 years, bioinformaticians have struggled with developing less than satisfactory solutions using a wide variety of tools and technologies. Recently, standardization of the Web Ontology Language (OWL), as part of the W3C's Semantic Web effort, offers exciting new opportunities that will facilitate biological knowledge management. In this talk, I will discuss the basic tools of the trade and highlight their use in creating a sophisticated knowledge base to store and query knowledge regarding the pharmacogenomics of depression.

4. **Pei-Chun Hsieh**, *Gene network inference applied to yeast data*: Since yeast gene networks have been investigated for many years, yeast serves as an ideal organism for the validation of network reconstruction algorithms. On the basis of gene expression measurements over time, our statistical model predicts which genes regulate cell-cycle-regulated genes of interest. For validation of the method, such predictions were compared to information in the literature..

5. **Yubing Liu**, *Genomic characterization of six1 function during skeletal muscle differentiation*: The Six gene family of transcription factors counts six members in vertebrates. Six1 is important for skeletal muscle development. Using C2C12 myoblast cell line as a model, we performed chromatin immunoprecipitation (ChIP) and then ChIP-on-Chip to illuminate the gene regulation network of Six1 in a genome wide fashion.

6. **Predrag Mizdrak**, *Iterative multiple sequence alignment (MSA) and phylogeny inference using a maximum likelihood (ML) method*: Most of the MSA methods today require a

guide tree, which represents evolutionary relationships between the input sequences. So, to solve an MSA problem, we need a guide tree; but to get a guide tree, we need a good MSA. We present a method for iteratively improving an MSA and its corresponding guide tree using a maximum likelihood procedure.

7. **Zahra Montazeri**, *Statistical identification of regulatory relationships between genes*: Discovering causal relationships between genes on the basis of observed data is often of interest. The researcher has a number of gene expression measurements over time and wishes to predict which genes regulate genes of interest. For this task, I constructed a statistical model that can predict the regulating gene that dominates the expression dynamics of each regulated gene of interest. I will present a number of regression models and prior distribution used to infer the model parameters representing gene-gene influences. These models are modified to deal with missing data case that commonly occurs in microarray studies. The proposed methods are applied to a set of data from plant cell cultures and also to yeast data.

8. **Nicolas Rodrigue**, *Bayesian models of protein-coding nucleotide sequence evolution with heterogeneous preferences at the amino acid level*: Using a nonparametric device known as the Dirichlet process prior, we propose a Markovian model of codon substitution that recognizes the heterogeneity of amino acid preferences across the coding positions of a gene. Bayesian methods of model assessment and ranking show its improved performance over traditional models, and emphasize its usefulness for inferring site-specific selective patterns operating at the amino acid level.

9. **Xiaoquan Yao**, *The nucleotide following the initiation codon does not influence translation efficiency in yeast – from a regression analysis*: In yeast, whether the nucleotide following the initiation codon (+4 nucleotide) influences translation efficiency is not clear. This study took advantage of yeast transcriptomic and proteomic data and applied a regression analysis to explore this question. The result shows that the +4 nucleotide does not influence translation efficiency in yeast.