# INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS

**University of Minnesota**
400 Lind Hall
207 Church Street S.E.
Minneapolis, Minnesota 55455
FAX (612) 626-7370 telephone (612) 624-6066 e-mail: `ima-staff@ima.umn.edu`
IMA Schedules via finger: `finger seminar@ima.umn.edu`
Newsletters, Updates and preprints are available via
anonymous ftp: `ftp.ima.umn.edu`, www: `http://www.ima.umn.edu/`

The IMA was founded by and receives major support from the National Science Foundation.

---

# IMA NEWSLETTER # 324

**6–31 October 2003**

2003–2004 Program

## PROBABILITY AND STATISTICS IN COMPLEX SYSTEMS

See `http://www.ima.umn.edu/complex/` for a full description of the 2003–2004 program on Probability and Statistics in Complex Systems: Genomics, Networks, and Financial Engineering

IMA schedules are subject to revision, particularly during workshops. See `http://www.ima.umn.edu/~seminar/sched` and `http://www.ima.umn.edu/newsltrs/` for the latest scheduling information.

*********************************************************************************

## PART I: NEWS AND NOTES

**IMA Workshop:**

### Comparative Genomics

20-24 October 2003

Organizers: Jens Lagergren (Royal Institute of Technology in Stockholm), Bernard Moret (University of New Mexico ), David Sankoff (University of Ottawa),

See http://www.ima.umn.edu/complex/fall/c2.html

**IMA Website**

Comments or suggestions concerning the IMA website may be addressed to

`webmaster@ima.umn.edu`.

In particular, we appreciate any information about World-Wide Web links appropriate to current and upcoming IMA programs.

---

PARTICIPATING INSTITUTIONS: Centrum voor Wiskunde en Informatica (CWI), Consiglio Nazionale delle Ricerche, Georgia Institute of Technology, Indiana University, Iowa State University, Kent State University, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Michigan State University, Mississippi State University, Northern Illinois University, Ohio State University, Pennsylvania State University, Purdue University, Sandia National Laboratories, Seoul National University (BK21 Math-SNU), Seoul National University (SRCCS), Texas A&M University, University of Chicago, University of Cincinnati, University of Delaware, University of Houston, University of Illinois (Urbana), University of Iowa, University of Kentucky, University of Maryland, University of Michigan, University of Minnesota, University of Notre Dame, University of Pittsburgh, University of Wisconsin, University of Wyoming, Wayne State University.

PARTICIPATING CORPORATIONS: Boeing, ExxonMobil, Ford, General Motors, Honeywell, IBM, Lockheed Martin, Lucent, Motorola, Schlumberger, Siemens, Telcordia Technologies, 3M.

Version of November 5, 2003

************************************************************************************

## PART II: Schedule for 1–31 OCTOBER 2003

### Monday, October 6

The 10:30 IMA break will be in Lind Hall 400.

### Tuesday, October 7

The 10:30 IMA break will be in Lind Hall 400.

**IMA POSTDOC SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 11:15-12:15 | **Dr. Taerim Lee**<br>Korea National Open University | Tree Structured Prognostic Model for Hepatocellular Carcinoma |

*Abstract*: Recent progress in both diagnostic and therapeutic technique of hepatocellular carcinoma appears to improve the prognosis. The purpose of this study was designed to evaluate the prognosis of HCC in relation to treatment methods and their affecting factors by the tree- structured model. This paper attempt to identify and quantify the effect of prognostic factors namely, patient characteristics that related to the prognosis of Hepatocellular Carcinoma. Our proposed survival tree model uses statistical tests for the modified Cox-Snell residuals derived from fitting a Cox proportional hazards model, and therefore it can detect curvature of covariates and interactions among them. As a result, split covariate selection bias is negligible. Furthermore, our proposed piecewise constant modeling can be generalized with piecewise multiple linear modeling without much computational burden.

The IMA Postdoc Seminar is organized by Balaji Gopalakrishnan and Antar Bandyopadhyay.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

### Wednesday, October 8

The 10:30 IMA break will be in Lind Hall 400.

**BROWN BAG SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 12:00 | **Lili Ju**<br>IMA | Numerical Simulations of the Quantized Vortices in a Thin Superconducting Hollow Sphere |

*Abstract*: In this talk, we investigate the vortex nucleation in a thin superconducting hollow sphere. The problem is studied using a simplified system of Ginzburg-Landau equations. We present finite volume methods which preserve the discrete gauge invariance for the time dependent simulation. The spatial discretization is based on a spherical centroidal Voronoi tessellation which offers a very effective high resolution mesh on the sphere for the order parameter as well as other physically interesting variables such as the super-current and the induced magnetic field. Various vortex configurations and energy diagrams are computed.

**COMPLEX SYSTEMS SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 3:30 pm | **Michael Newton**<br>U of Wisconsin, Madison and IMA | Randomization Test for Array-Based Comparative Genomic Hybridization |

*Abstract*: After low-level preprocessing, data from an array-based comparative genomic hybridization (aCGH) experiment may be viewed as a set of independent and identically distributed (iid) copies of the random vector $X = (X_1, X_2, ..., X_n)$ where $X_i \ \mathrm{Bernoulli}(p_i)$, but in which there is dependence among the components owing to aspects of the tumor growth process that is being measured by $X$. Here $i$ denotes a position in the genome and $X_i$ indicates whether or not genomic damage occurs at position $i$ in the sampled tumor. One statistical problem is to test the null hypotheses $H_0 : p_i = p$ for all $i$; that is, there are no aberration hot-spots in the genome. I will describe several attempts to implement such a test using conditional inference and three permutation methods. I conjecture that an elementary shuffling procedure provides a conservative hypothesis test. The approach is tested on a set of 60 aCGH profiles from cancer cell lines. Issues with hidden-Markov model-based inferences will also be discussed.

**Thursday, October 9**

The 10:30 IMA break will be in Lind Hall 400.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

**Friday, October 10**

The 10:30 IMA break will be in Lind Hall 400.

**IMA/MCIM INDUSTRIAL PROBLEM SEMINAR, 570 Vincent Hall:**

| | | |
|---|---|---|
| 1:25pm | **Indraneel Das**<br>United Technologies Research Center, Hartford, CT | Challenges in Industrial Optimization . . . Technical or Otherwise |

*Abstract*: Numerous industrial problems have been approached using optimization and many corporations have benefited from the resulting solutions. Yet, why has it not grown into a core competency in the corporate world? The process of producing value for the industry using Optimization involves

- (a) identifying/getting the problem
- (b) analyzing the technical aspects and “solving” the problem
- (c) explaining why “the answer” is the answer, and what is good about it
- (d) setting oneself up for the next problem.

The speaker will discuss the challenges in each of these steps. In particular, he will point out the underlying technical challenges in optimization problems ranging from petroleum and telecommunications to manufacturing and energy solutions, all drawn from his diverse industrial background, and emphasize relevant areas of research.

**Monday, October 13**

The 10:30 IMA break will be in Lind Hall 400.

**Tuesday, October 14**

The 10:30 IMA break will be in Lind Hall 400.

**IMA POSTDOC SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 11:15-12:15 | **Dr. Simon Morgan**<br>University of Minnesota | Applications of Geometric Measure Theory to Variational Problems with Surfaces |

*Abstract*: Geometric measure theory offers compactness theorems for candidate objects in variational problems, giving existence results in variational problems. Additionally it offers regularity information about limits of regular objects.

I will introduce the objects of geometric measure theory; rectifiable sets, currents and varifolds. Key examples of each and their associated topologies will be given.

The primary application of geometric measure theory was minimal surface theory, but I will also explain my own research into geometric measure theory techniques and my motivating applications.

The IMA Postdoc Seminar is organized by Balaji Gopalakrishnan and Antar Bandyopadhyay.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

**Wednesday, October 15**

The 10:30 IMA break will be in Lind Hall 400.

**BROWN BAG SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 12:00 | **Sabera Kazi**<br>Honeywell | QOS-enabled MAC protocol for Wireless Networks of UAVS |

**COMPLEX SYSTEMS SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 2:30 pm | **Stephen Willson**<br>Iowa State University and IMA | Finding Supertrees Using Distance Methods |

*Abstract*: Suppose that a tree $T$ has the set $S$ of leaves and $S'$ is a subset of $S$. Then $T|S'$ denotes the tree obtained from $T$ with the set $S'$ of leaves, showing only the relationships among $S'$. Suppose that for $i = 1, \ldots, k$, $T_i$ is a tree with the set $S_i$ of leaves. A "supertree" is a single tree $T$ with the set $S$ of leaves, where S is the union of the $S_i$, such that $T|S_i = T_i$ for each $i$. A major problem in phylogeny is, given such a collection of trees $T_i$, to identify a supertree $T$ if one exists. A method for solving this problem could be used to combine different phylogenetic trees from different researchers into a larger tree.

The most common method currently used to find supertrees is MRP (matrix representation with parsimony). It is computationally intensive, it often leads to a great many possible solution trees, and it is not founded on theory.

This talk proposes a supertree method in which additional input information is required. We require in addition that for each tree $T_i$ there is given an additive distance function $d_i$ on $S_i$ which estimates the amount of evolution between any two taxa. We then try to create a supertree $T$ with additive distance function $d$ which equals $d_i$ when restricted to $S_i$.

Some results of simulations will be presented.

**Thursday, October 16**

The 10:30 IMA break will be in Lind Hall 400.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

**Friday, October 17**

The 10:30 IMA break will be in Lind Hall 400.

**IMA/MCIM INDUSTRIAL PROBLEM SEMINAR, 570 Vincent Hall:**

| | | |
|---|---|---|
| 1:25pm | **D. Subramanian**<br>Honeywell Labs, Minneapolis | Mathematical Programming and Multiaircraft Conflict Resolution |

*Abstract*: Free flight is an emerging paradigm in Air Traffic Management (ATM). Conflict detection and resolution is the heart of any free flight concept, and is the focus of this presentation. We address the problem of optimal cooperative three-dimensional (3D) conflict resolution involving multiple aircraft using rigorous numerical trajectory optimization methods. The conflict problem is posed as an optimal control problem of finding trajectories that minimize a certain objective function while maintaining the safe separation between each aircraft pair. We assume the origin and destination of the aircrafts are known and consider aircrafts models with simplified kinematics as well as detailed nonlinear point-mass dynamics. The protection zone around the aircraft is modeled to be cylindrical in shape. We propose a novel formulation of the cylindrical protection zone using continuous variables. The optimal control is converted to a finite dimensional Non Linear Program (NLP) using collocation on finite elements. We solve the NLP using an Interior Point algorithm that incorporates a novel line search method. Lastly, we also discuss some open problems of research interest in the above context.

**Monday, October 20**

**IMA Workshop:**

## Comparative Genomics

20-24 October 2003

Organizers: Jens Lagergren (Royal Institute of Technology in Stockholm), Bernard Moret (University of New Mexico ), David Sankoff (University of Ottawa),

See http://www.ima.umn.edu/complex/fall/c2.html

The increasing availability of complete genomes from diverse organisms offers unprecedented opportunities. Exploitation of the full power intergenomic comparative maps for all types of genomic events will be central in biological, medical and bioinformatics research in the post-genomic era. Several areas are crucial to the success of this enterprise, for instance: understanding patterns and processes of genome evolutionary change, mapping genomic mutational events, and the utilization of such maps as bioinformatics tools. Genomic data also facilitates phylogeny reconstruction based on genomic mutational events rather than nucleotide substitution.

The kernel of comparative genome analysis is the establishment of the correspondence (orthology analysis) between genes in different genomes. It is such intergenomic maps that make it possible to translate information from one organism to another. Genome evolution is shaped by a multitude of evolutionary events acting at various organizational levels. On a low level point mutations affect individual nucleotides. On a higher level genome segments are affected by processes such as duplication, lateral transfer, inversion, transposition, deletion and insertion. Finally, the whole genome is influenced by speciation and hybridization of organism lineages. The complexity of genome evolution poses a serious challenge in developing mathematical models and algorithms. Fortunately, there is a spectra of algorithmic techniques that can be applied to problems from this domain, ranging from exact, heuristic, fixed parameter and approximation algorithms for problems based on parsimony models to Monte Carlo Markov Chain algorithms for Bayesian analysis of problems based on probabilistic models.

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

| | | |
|---|---|---|
| 8:30 | **Coffee and Registration** | Reception Room EE/CS 3-176 |
| 9:15 | **Douglas N. Arnold, Scot Adams, and Organizers** | Welcome and Introduction |
| 9:30 | **Michael Lynch**<br>Indiana University | The Origins of Genome Complexity |

*Abstract*: Complete genomic sequences from diverse phylogenetic lineages reveal striking increases in genome complexity across the prokaryote to unicellular eukaryote to multicellular eukaryote boundaries. The changes include gradual growth in gene number resulting from the retention of duplicate genes, more abrupt increases in the abundance of spliceosomal introns and mobile genetic elements, and enhanced modularity of gene regulation. A case can be made that many of these changes emerged passively in response to substantial long-term population-size reductions that accompanied increases in organism size and magnified the power of random genetic drift. Under this model, much of the restructuring of eukaryotic genome organization and the roots of many aspects of organismal complexity were initiated by nonadaptive processes. Although the mutational changes necessary for genomic modification are initiated by molecular processes, the population-genetic environment ultimately defines the permissible paths of evolution. The simple genomes of most microbial species can be understood in this context, without invoking direct selection for streamlined genomes, and direct selection for complexity need not be invoked to explain genomic expansion in multicellular species.

| | | |
|---|---|---|
| 10:20 | **Discussion** | |
| 10:30 | **Coffee** | Reception Room EE/CS 3-176 |
| 11:00 | **Nadia El-Mabrouk**<br>University of Montreal | Reconstructing the Ancestor of a Modern Genome with Multigene Families |

*Abstract*: Given a particular model of evolution and an optimization criterion, the problem is to recover an ancestor of a modern genome modeled as an ordered sequence of signed genes. One direct application is to infer gene orders at the ancestral nodes of a phylogenetic tree. Implicit in the rearrangement literature is that each gene is present exactly once in each genome. This hypothesis is clearly unguaranteed for divergent species containing several copies of highly paralogous and orthologous genes. In this presentation, we consider models of genome evolution that take multigene families into account.

We first present a genome-wide doubling event. Genome duplication is an important source of new gene functions and novel physiological pathways. Originally (ancestrally), a duplicated genome contains two identical copies of each chromosome, but through genomic rearrangements, this simple doubled structure is disrupted. At the time of observation, each of the chromosomes resulting from the accumulation of rearrangements can be decomposed into a succession of conserved segments, such that each segment appears exactly twice in the genome. We present exact algorithms for reconstructing the ancestral doubled genome in linear time, minimizing the number of inversions and/or translocations required to derive the observed order of genes along the present-day chromosomes.

The second part of the presentation will concern a model of duplications at a regional level. In this model, chromosomal regions (one or more genes) are duplicated from one location of the genome to another. Studies from human genomic sequence indicate that many of these segments have been duplicatively transposed in very recent evolutionary time. The implicit hypothesis is that a genome with multigene families has an ancestor containing exactly one copy of each gene that has evolved through a series of duplication transpositions and substring inversions. We present an algorithm for reconstructing an ancestral genome giving rise to the minimal number of duplication transpositions and reversals. We then show how to use this algorithm to recover gene orders at the ancestral nodes of a phylogenetic tree.

| | | |
|---|---|---|
| 11:50 | **Discussion** | |
| 12:00 | **Lunch Break** | |
| 1:30 | **Roderic Guigö**<br>Institut Municipal d'Investigacio Medica (IMIM/UPF/CRG) | Comparative Gene Prediction |

*Abstract*: Comparative genomics is emerging as a powerful tool to characterize complex genomes. Gene prediction, in particular, has benefited from the availability of genome sequences from organisms across the whole eukaryotic spectrum. The comparison of the human and mouse genome sequences, for instance, has contributed substantially to refine the gene content of the human (and mouse) genomes. In my talk, I will stress how comparative genomes may be particularly useful to identify genes which deviated from the standard characteristics, and that, for this reason, may escape identification by other means.

| | | |
|---|---|---|
| 2:20 | **Discussion** | |
| 2:30 | **Coffee** | Reception Room EE/CS 3-176 |
| 3:00 | **Guillaume Bourque**<br>Universite de Montreal | A Comparative Approach for Multiple Gene Network Inference Using Time-Series Gene Expression Data |

*Abstract*: We present a method for gene network inference and revision based on time-series data. Gene networks are modeled using linear differential equations and a generalized stepwise multiple linear regression algorithm is used to recover the interaction coefficients. Our system was design for the recovery of gene interactions concurrently in many gene regulatory networks related by a graph or a tree. Suppose we are studying a certain regulatory network in different species of known phylogeny. We can think of the different networks as being related to each other in that way and use this information. Alternatively, we might be interested in the development stages of this network or we could be studying the same system but in different tissues related at a different level. The idea is that, given gene expression data for each species, or each stage of development, or each tissue, we seek to recover each individual network while minimizing a cost based on the differences along the edges of the graph or the tree. We show how this comparative framework allows new insights and facilitates the gene network inference process.

| | | |
|---|---|---|
| 3:20 | **Discussion** | |
| 3:30 | **Avril Coghlan**<br>Trinity College Dublin, Ireland | Origins of Recently Gained Introns in Caenorhabditis |

*Abstract*: The genomes of the nematodes Caenorhabditis elegans and C. briggsae both contain about 100,000 introns, of which about 6000 are unique to one species or the other. To study the origins of new introns, we used a rigorous method involving phylogenetic comparisons to animal orthologs and other nematode paralogs to identify cases where an intron content difference between C. elegans and C. briggsae was almost certainly caused by intron insertion rather than deletion. We identified 57 putative recently gained introns in C. briggsae and 112 in C. elegans. Novel introns have a stronger exon splice site consensus sequence than the general population of introns, and they show the same preference for phase 0 sites in codons over phases 1 and 2 as seen in the general population. More of the novel introns are inserted in genes that are expressed in the germline than expected by chance. As compared to matched control sets of C. briggsae introns, the novel introns in C. briggsae are more likely to contain an annotated repeat element (1.7-fold; P = 0.011), and the ends of the intron are more likely to be close to the ends of the repeat element (1.5-fold; P = 0.029). Similar but weaker trends are also seen in C. elegans novel introns. One family of C briggsae repeat elements, which is related to the Helitron class of

putative nonautonomous transposons, is found in significantly more novel introns than reference introns (P ¡ 1e-05). These results support the hypothesis that novel introns originate as a result of transposable element insertions into proto-splice site consensus sites in germline-expressed genes.

Joint work with Kenneth H. Wolfe (Trinity College Dublin, Ireland).

| | | |
|---|---|---|
| 3:50 | **Discussion** | |
| 4:00 | **Igor V. Sharakhov**<br>University of Notre Dame | High Rates of Genome Rearrangements in Malaria Mosquitoes, Anopheles gambiae and A. funestus |

*Abstract*: The rates of chromosomal evolution vary among different genomic segments and eukaryotic lineages [1]. A comparative genomic study between Drosophila melanogaster and Anopheles gambiae shows extensive reshuffling of gene order within chromosomes [2]. Genus Drosophila has a very high rate of paracentric inversions [3]. Our study determines rates of chromosomal rearrangement in genus Anopheles. Anopheles gambiae and A. funestus, important vectors of malaria in tropical Africa, are in the same subgenus and diverged about as recently as humans and chimpanzees ($\tilde{5}$ million years ago) [4]. Using fluorescence in situ hybridization (FISH), we mapped A. funestus cDNA clones on the five arms of the polytene chromosome complement. Of 157 cDNAs used as probes, 116 mapped to single chromosomal locations on the A. funestus cytogenetic map, and the remainder hybridized in multiple locations. Those 116 cDNAs were mapped in silico to the completely sequenced A. gambiae genome. The relative positions of sequences with unique map locations in both species support the hypothesized chromosome arm homologies and the reciprocal whole arm translocation between 2L and 3R, postulated previously on the basis of relative length and banding pattern [5]. Correspondence between chromosome arms was contradicted by only two of the cDNAs examined in this study. Within corresponding arms, paracentric inversions have had a major impact on genome architecture since the divergence of these species. Gene order has not been preserved along the length of any chromosome arm, although there are conserved segments in some regions near centromeres where the rate of meiotic recombination may be reduced. Inversions have involved large as well as relatively small chromosomal segments. One of three small inversions at the distal end of 2R includes a rearrangement involving the 8C region in A. gambiae that contains the major Plasmodium-refractoriness locus Pen1 [6]. What has been the extent of rearrangement of gene order between these species? The number of inversion events can be estimated by considering the mean length of conserved segments, because this length decreases with each inversion fixed since the divergence of A. gambiae and A. funestus from a common ancestor. The method of Nadeau and Taylor [7] was applied to estimate mean lengths of all conserved segments in the genome, based on the nucleotide distance in A. gambiae between the outermost markers that defined the segments observed in our sample. An assumption of the method, that rearrangements fixed during evolution are randomly distributed in the genome, seems unlikely given the extraordinary concentration of polymorphic inversions on 2R in both lineages. Of eight polymorphic inversions described in A. gambiae, seven occur on chromosome 2R [8]. Similarly, 11 of 15 polymorphic inversions found in A. funestus involve 2R [9]. Accordingly, we assessed each arm independently. The estimated mean lengths of all conserved segments on each arm, defined with respect to A. gambiae, were X, $2.0 \pm 0.2$ megabases (Mb); 2R, $0.9 \pm 0.2$ Mb; 2L, $2.2 \pm 0.4$ Mb; 3R, $2.2 \pm 1.0$ Mb; and 3L, $1.1 \pm 0.4$ Mb. In a slight departure from Nadeau and Taylor [7], each rearrangement was assumed to be an inversion requiring two disruption events. Therefore, n inversions result in 2n + 1 conserved segments. The number of inversions on each arm was $5 \pm 1$, $36 \pm 9$, $11 \pm 3$, $11 \pm 3$, and $19 \pm 5$, respectively. Assuming a divergence time of 5 million years [4], the rate of fixation per My for each chromosome arm can be estimated as 0.5, 3.6, 1.1, 1.1, and 1.9, respectively (or 7 when estimated across the genome). When normalized to account for differences in chromosome length, the number of inversions per Mb per My for X, 2R, 2L, 3R, and 3L was estimated as 0.023, 0.057, 0.022, 0.021, and 0.044, respectively (0.031 genome-wide). This rate is even more extreme than the genome-wide estimate for Drosophila [3]. Moreover, our results indicate that 2R has a higher rate of rearrangement than other arms. It is clear that tightly linked genes in A. gambiae are unlikely to be similarly linked in A. funestus, particularly on 2R. The estimate of mean conserved segment length derived for each arm can be used to predict the probability of linkage in A. funestus, given the known distance between genes in A. gambiae and the assumption of random distribution of breakpoints [7]. As an example, the probability that genes 1 Mb apart on 2R in A. gambiae are linked on 2R in A. funestus is only 0.31. Polymorphic inversions on chromosome 2R are widespread within the A. gambiae and A. funestus and are believed to indicate adaptations to different environmental niches [8, 9]. Identification of genes encoded within these inversions could provide clues to factors determining mosquito behavior and vectorial capacity. Thus, the main features of genome rearrangements in malaria mosquitoes, A. gambiae and A. funestus, can be summarized as following: (1) the reciprocal whole arm translocation has preserved a synteny (the occurrence of genes) at the whole-arm level; (2) high rate of paracentric inversions, especially on 2R, have had a major impact on extensive gene order reshuffling.

Our results suggest that the success of positional cloning or interspecific microarray experiments may be limited to either very closely related anopheline species or small genomic fragments. Further comparative studies of these two genomes will provide valuable insights into the mechanism and effects of chromosomal rearrangements. This study was supported by grants from NIH (AI48842) to N.J.B. and from the Indiana 21st Century Research & Technology Fund to F.H.C.

References:

1. E. Eichler, D. Sankoff, Science 301, 5634 (2003).
2. E. M. Zdobnov et al., Science 298, 149 (2002).

3. J. González, J. M. Ranz, A. Ruiz, Genetics 161, 1137 (2002)
4. I. V. Sharakhov et al., Science 298, 182 (2002).
5. I. V. Sharakhov, M. V. Sharakhova, C. M. Mbogo, L. L. Koekemoer, G. Yan, Genetics 159, 211 (2001)
6. L. Zheng, et al., Science 276, 425 (1997)
7. J. H. Nadeau and B. A. Taylor, Proc. Natl. Acad. Sci. U.S.A. 81, 814 (1984)
8. M. Coluzzi, A. Sabatini, V. Petrarca, M. A. Di Deco, Trans. R. Soc. Trop. Med. Hyg. 73, 483 (1979)
9. I. Dia, D. Boccolini, C. Antonio-Nkondjio, C. Costantini, D. Fontenille, Parassitologia 42, 227 (2000)

Joint work with Andrew C. Serazin (1), Olga G. Grushko (1), Ali Dana (1), Neil Lobo (1), Maureen E. Hillenmeyer (1), Richard Westerman (2), Jeanne Romero-Severson (3), Carlo Costantini (4), N'Fale Sagnon (4) Frank H. Collins (1), Nora J. Besansky (1)

(1) Center for Tropical Disease Research and Training, University of Notre Dame, Notre Dame, IN 46556-0369, USA.

(2) Horticulture Department, Purdue University, West Lafayette, IN 47907-1159, USA.

(3) Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN 47907-1165, USA.

(4) Centre National de Recherche et de Formation sur le Paludisme, Ouagadougou, Burkina Faso.

| | | |
|---|---|---|
| 4:20 | **Discussion** | |
| 4:30 | **Second Chances** | Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions. |
| 5:00 | **Group Photo** | |
| 5:10 | **IMA Tea and more (with Poster Session)** | IMA East, 400 Lind Hall |
| poster session | **Max Alekseyev**<br>University of California, San Diego | Genome Halving Problem |

*Abstract*: Genome Halving Problem is motivated by an evolution mechanism that duplicates the entire genome. The result of such duplication, so-called perfectly duplicated genome, contains two identical copies of each chromosome. The genome then is a subject to reversal and/or translocation rearrangement operations. For given rearranged duplicated genome, Genome Halving Problem attempts to recover its closest perfectly duplicated ancestor. Solution to this problem is used as a building block for more sophisticated genome rearrangement algorithms.

Genome Halving Problem was first introduced and solved in a series of papers by Nadia El-Mabrouk and David Sankoff. Their algorithm is rather complex and, to the best of our knowledge, it was never implemented as a computer program. In our work we present a new simpler and more general algorithm for Genome Halving Problem as well as its implementation in C++.

Joint with Pavel Pevzner, University of California, San Diego.

poster session

**Lars Arvestad**
Royal Institute of Technology (KTH)

New Methods for Estimating Amino Acid Replacement Rates

*Abstract*: Two new methods for estimating replacement rate matrices from protein sequence alignments are presented and shown to perform better than another recent method, Müller-Vingron's resolvent method, in a variety of settings. Furthermore, the best method is demonstrated to be robust on small datasets and practical also on very large datasets of real data. Neither short nor divergent sequence pairs have to be discarded, making the method economical with data.

poster session

**Anne Bergeron**
LaCIM, Université du Québec à Montréal, Canada

*Abstract*: See http://www.ima.umn.edu/complex/abstracts/bergeron/Poster.pdf.

poster session

**Steven B. Cannon**
University of Minnesota

Distinguishing Orthologs from Paralogs by Integrating Comparative Genome Data and Gene Phylogenies

*Abstract*: Background: In eukaryotic genomes, most genes are members of gene families. When comparing genes from two species, therefore, most genes in one species will be homologous to multiple genes in the second. This often makes it difficult to distinguish orthologs (separated through speciation) from paralogs (separated by other types of gene duplication). Combining phylogenetic relationships and genomic position in both genomes helps to distinguish between these scenarios. This kind of comparison can also help to describe how gene families have evolved within a single genome that has undergone polyploidy or other large-scale duplications, as in the case of Arabidopsis thaliana and probably most plant genomes.

Results: We describe a suite of programs called OrthoParaMap that makes genomic comparisons, identifies syntenic regions, determines whether sets of genes in a gene family are related through speciation or internal chromosomal duplications, maps this information onto phylogenetic trees, and infers internal nodes within the phylogenetic tree that may represent local as opposed to speciation or segmental duplication. We describe the application of the software using three examples: the melanoma-associated antigen (MAGE) gene family on the X chromosomes of mouse and human; the 20S proteasome subunit gene family in Arabidopsis, and the major latex protein gene family in Arabidopsis.

Conclusion: OrthoParaMap combines comparative genomic positional information and phylogenetic reconstructions to identify which gene duplications are likely to have arisen through internal genomic duplications (such as polyploidy), through speciation, or through local duplications (such as unequal crossing-over). The software is freely available at http://www.tc.umn.edu/c̃ann0010/Software.html.

Joint work with Georgiana May (Plant Biology Department and Ecology, Evolution, and Behavior Department, University of Minnesota, St. Paul, MN 55108, USA) and Nevin D. Young (Plant Biology Department and Plant Pathology Department, University of Minnesota, St. Paul, MN 55108, USA).

poster session

**Dimitra Chalkia**
The Pennsylvania State University

Phylogenetic Analysis of Formin Homology Proteins in Arabidopsis Thaliana and Oryza Sativa

*Abstract*: The plant cell cytoskeleton plays an important role in many cellular processes, including cell polarity establishment and cytokinesis. Proteins that regulate cytoskeletal assembly are likely to be a part of the signaling cascade that governs plant cell morphogenesis. Formins are members of a large protein family that is defined by the presence of the highly conserved Formin Homology II (FH2) domain. In a wide range of organisms, including vertebrates, arthropods, nematodes and fungi, formins have been implicated in the regulation of cytoskeletal assembly and in the control of cytokinesis and cell polarity establishment and maintenance. The genomes of Arabidopsis thaliana and Oryza sativa contain putative formin-like proteins based on the presence of an FH2 domain. Arabidopsis thaliana formins have been tentatively sub-divided into two clades: Type I and Type II, based on the FH2 domain alignment. We have extended this analysis to cover both Arabidopsis and rice and have provided an evolutionary context for these plant formin families.Our phylogenetic analysis shows that formins are divided in two distinct clades in plants. This phylogenetic clustering is also supported by the stuctrural features of these proteins. This division of plant formins in two distinctive groups seems to predate the split of monocots/eudicots. The detailed evolutionary relationships of plant formins remain unclear. The placement of fungi formins at the basal position of the tree is in accordance with the most recent proposed phylogenetic scheme for eukaryotes. Animal and plant formins cluster together, and split into two major groups. This clustering may suggest that

their last common ancestor had already at least two different types of formins.

Joint work with Tatiana Bibikova, Simon Gilroy, Wojciech Makalowski.

| | | |
|---|---|---|
| poster session | **Jack R. Collins**<br>National Cancer Institute/SAIC-Frederick | Comparative Genomic Analysis as a Tool to Elucidate Eukaryotic Chromosomal Structure and Function |

*Abstract*: The release of several nearly-complete eukaryotic genomes has provided a massive data set for analysis in terms of elucidation of the sets of genes which they encode and the mechanisms by which those genes are regulated. However, it has become apparent that programs that predict gene structure, organization and regulation can be aided appreciably with the use of information from multiple organisms. We have applied cross-species comparative analysis between human, mouse, and rat to the identify and characterize relatively closely spaced chromosomally duplicated regions (segmental duplications) as well as putative regulatory regions. In the case of the chromosomally duplicated regions, we have taken a gene centric view of these regions, specifically providing insights into genes likely to become aberrantly expressed by either duplication or deletion events in each of the species. These regions have been analyzed in terms of their correlation with other genomic features such as gene density, common repeat family content and several other features. For the putative regulatory region analysis, a database has been constructed in which transcription factor binding sites, stem-loops, overly-represented oligonucleotides, SNPs and other genomic features can be compared amongst any of the genes mapped in common between the three genomes. This analysis has identified both conserved transcription factor binding site patterns as well as additional conserved regions which may be involved in gene regulatory events important in coordinating the accurate temporal and spatial expression patterns observed during the course of development. The implications of including more genomes (as they become available) in these types of analyses in the understanding of biological mechanisms involved in disease, especially cancer, will also be discussed.

Joint work with Robert M. Stephen, Advanced Biomedical Computing Center, National Cancer Institute/SAIC-Frederick, Frederick, MD 21702 (bobs@ncifcrf.gov, collinsj@ncifcrf.gov).

| | | |
|---|---|---|
| poster session | **Ramana V. Davuluri**<br>The Ohio State University | Mammalian Promoter Database: A Computational Platform for Comparative Genomics of Mammalian Transcriptional Regulation |

*Abstract*: Transcription in mammalian cells is a highly complex process that involves multiple layers of general and gene-specific transcription factors. Although extensive molecular research has been providing important details about several transcription factors and their binding sites in the target gene promoters, the information generated over the years is highly fragmented. In order to better integrate this vast amount of information with the genome sequences, we have developed a new database called MPromDb (Mammalian Promoter Database), an information resource of mammalian gene regulatory regions. MPromDb (Version 1.0) contains 28,306 experimentally supported and 32,121 computationally annotated promoters, and mapping of 4,231 experimentally known binding sites, with links to published literature. Each promoter sequence in MPromDb is presented in the form of an image map with annotations of first exon, cis-regulatory elements and plots of CpG scores, with interactive contextual menus for easy navigation. MPromDb provides a platform for comparative genomics of transcriptional regulation, since promoters of orthologous genes are linked with each other and displayed in the same record. The current version contains 9,331 human-mouse orthologous pairs. The database can be searched for promoter sequences, transcription factors, and their direct target genes, through a user-friendly web interface at http://bioinformatics.med.ohio-state.edu/MPromDb.

Joint work with Hao Sun, Saranyan K. Palaniswamy, Twyla T. Pohar, and Victor Jin ( Human Cancer Genetics Program, Comprehensive Cancer Center, Department of Molecular Virology, Immunology & Medical Genetics, The Ohio State University, 420 W 12th Avenue, TMRF 524, Columbus, OH 43210, USA).

| | | |
|---|---|---|
| poster session | **Anant Godbole**<br>East Tennessee State University | Distributional Approximations in Genome Reconstruction |

*Abstract*: See http://www.ima.umn.edu/complex/abstracts/godbole/godbole-abstract.pdf

| | | |
|---|---|---|
| poster session | **Steve Goldstein**<br>University of Wisconsin and OpGen | Graph Compression Algorithms for Efficiently Comparing Genomes |

*Abstract*: Optical Mapping is a system capable of producing genome-wide ordered restriction maps. Such a restriction map provides a description of an organism's genome, a description not unlike the sequence of the genome, albeit at a coarser resolution. Just as comparisons of whole genome sequences are leading to an exciting array of biological advances, comparisons of optical maps will provide a wealth of valuable information.

Now that optical mapping has entered the high-throughput era, there is a need for software to compare restriction maps of closely related organisms. We present an algorithmic framework for this task, closely modeled after DNA sequence comparison algorithms. The major challenge lies in adapting the exact matching phase of the sequence algorithms to handle the imprecision inherent in determining restriction fragment lengths. Our graph-based approach not only overcomes this challenge, but also can be applied to sequence algorithms, providing advantages over suffix-tree approaches.

| | | |
|---|---|---|
| poster session | **Josefa González**<br>Universitat Autónoma de Barcelona | Duplicative and Conservative Transpositions of the Larval Serum Protein 1 Genes in the Genus Drosophila |

*Abstract*: In the genus Drosophila, homologous chromosomal elements show a remarkable conservation of gene content but not of gene order, indicating that paracentric inversions are the most common kind of genomic change. Detailed physical maps of chromosomes X, 2 and 4 of Drosophila repleta and D. buzzatii, both belonging to the Drosophila subgenus, were constructed and their gene rearrangements compared with the homologous chromosomes in D. melanogaster. We estimated that 393 paracentric inversions have been fixed in the whole genome since the divergence between D. repleta and D. melanogaster, that amounts to an average rate of 0.053 disruptions/Mb/myr. Only two exceptions to the chromosomal homologies were found and we have further analyzed one of them: the transposition of the Larval serum protein 1 (Lsp1) genes. Comparative molecular analysis of the transposed genes and their flanking regions can help to elucidate the time, direction and mechanism of gene transposition. In the D. melanogaster genome, three Lsp1 ge es, alpha, beta and gamma, are present and each is located on a different chromosome. We have characterized the molecular organization of Lsp1 genes in D. buzzatii and in D. pseudoobscura, a species of the Sophophora subgenus. Our results show that only two Lsp1 genes (beta and gamma) exist in these two species suggesting that the duplicative transposition generating Lsp1alpha, took place ¡30 myr ago in the D. melanogaster lineage. D. buzzatii and D. pseudoobscura show the same chromosomal localization and genomic organization, different from that of D. melanogaster for the Lsp1beta and Lsp1gamma genes. Thus we conclude that this is likely to be the ancestral organization and both genes must have conservatively transposed in the D. melanogaster lineage ¡30 myr ago. Finally, the duplicative transposition which gave rise to Lsp1beta and Lsp1gamma must have occurred before the divergence of the three Drosophila species (40-62 myr ago). Overall, at least two duplicative and two conservative transpositions are necessary to explain the present chromosomal distribution of Lsp1 genes in the three Drosophila species. In D. buzzatii and D. pseudoobscura, Lsp1beta and Lsp1gamma are localized close to snRNA or tRNA genes. RNA genes have been implied in the origin of chromosomal rearrangements in prokaryotes and yeasts and we find clear evidence for a role of snRNA genes in the transposition of Lsp1beta genes in Drosophila. Analysis of the 5' non coding regions of the Lsp1beta and Lsp1gamma genes has led to identify the putative cis-acting regulatory regions of these genes which seemingly transposed along with the coding sequences.

Joint work with Ferran Casals and Alfredo Ruiz Departament de Genética i Microbiologia. Universitat Autónoma de Barcelona, 08193, Bellaterra (Barcelona), Spain.

| | | |
|---|---|---|
| poster session | **Tzvika Hartman**<br>Weizmann Institute of Science, Rehovot, Israel | A Simpler 1.5-Approximation Algorithm for Sorting by Transpositions |

*Abstract*: In this work we study the problem of sorting by transpositions. First, we prove that the problem of sorting circular permutations by transpositions is equivalent to the problem of sorting linear ones. Hence, all algorithms for sorting linear permutations by transpositions can be used to sort circular permutations. Then, we derive our main result: A new quadratic 1.5-approximation algorithm, which is considerably simpler than the extant algorithms of Bafna and Pevzner (1998) and Christie (1999). Thus, the algorithm achieves running time which is equal to the best known, with the advantage of being much simpler. Moreover, the analysis of the algorithm is significantly less involved, and provides a good starting point for studying related open problems.

Joint work with Ron Shamir School of Computer Science, Tel-Aviv University, Tel-Aviv, Israel.

| | | |
|---|---|---|
| poster session | **Robert (Bob) Mau**<br>University of Wisconsin-Madison | Inferring Orthologous Regions via a Pseudo-Gibbs Sampler: Finding the Pieces of the Rearrangement Puzzle |

*Abstract*: See http://www.ima.umn.edu/complex/abstracts/mau/poster.pdf.

| | | |
|---|---|---|
| poster session | **Daniel P. Miranker**<br>University of Texas - Austin | Application of MoBIoS for Conserved Primer Pair Discovery |

*Abstract*: MoBIoS, a Molecular Biological Information System is a next generation database management system focused on scalable retrieval and mining of unorthodox biological data types that are poorly supported by relational database systems. MoBIoS comprises built-in data types for biological sequences and Mass Spectra. The MoBIoS storage manager extends traditional database systems by including built-in support for hierarchical clustering and nearest-neighbor and range search in metric spaces. In addition to built-in metrics to support sequence homology and protein identification, users may add their own metrics.

We report on the first biological application of MoBIoS; a comparative study of the entire genomes of the plants rice and Arabidopsis to determine conserved pairs of strings of DNA that could be used to prime polymerase chain reactions (PCRs). Identification of such set of paired conserved primers would allow amplification of evolutionarily homologous DNA regions in a taxonomically broad set of seed plants. The ability to amplify homologous regions in a widely divergent set of species has a number of applications, e.g., phylogenetic reconstruction and comparison of protein evolution in a broad set of organisms. Ultimately, this approach to identifying conserved primer pairs could provide the community of systematists with a universal set of DNA sequences that can be used for assembling the tree of life.

Joint work with Weijia Xu, Wenguo Liu and C. Randal Linder.

| | | |
|---|---|---|
| poster session | **William J. Murphy**<br>National Cancer Institute-Frederick | Reconstructing the Genomic Architecture of Mammalian Ancestors Using Multispecies Comparative Maps |

*Abstract*: Rapidly developing comparative maps in selected mammal species are providing an opportunity to reconstruct the genomic architecture of mammalian ancestors and study rearrangements that transformed this ancestral genome into existing mammalian genomes. Here we apply the recently developed Multiple Genome Rearrangement algorithm (MGR) to human, mouse, cat and cattle comparative maps (with 311-470 shared markers) to impute the ancestral mammalian genome. Reconstructed ancestors consist of 70-100 conserved segments shared across the genomes that have been exchanged by rearrangement events along the ordinal lineages leading to modern species genomes. Genomic distances between species, dominated by inversions (reversals) and translocations, are presented in a first multispecies attempt using ordered mapping data to reconstruct the evolutionary exchanges that preceded modern placental mammal genomes.

Joint work with Guillaume Bourque (Centre de Recherches Mathématiques, Université de Montréal Montréal, Canada H3C 3J7), Glenn Tesler and Pavel Pevzner (Department of Computer Science and Engineering, University of California, San Diego La Jolla, California 92093-0114), and Stephen J. O'Brien (Laboratory of Genomic Diversity, National Cancer Institute, Frederick, Maryland 21702).

| | | |
|---|---|---|
| poster session | **Nikolas Nikolaidis**<br>Pennsylvania State University | Evolution of the Hsp70 Gene Superfamily in Two Sibling Species of Nematodes Caenorhabditis elegans and C. briggsae |

*Abstract*: The Hsp70 gene superfamily of C. briggsae was characterized in an attempt to investigate the evolutionary relationships with the respective one of its sibling species C. elegans. The phylogenetic analyses included also genes from Drosophila melanogaster and Saccharomyces cerevisiae to clarify the long-term evolution of hsp70s. The Hsp70s are classified into three monophyletic groups according to their sub-cellular localization, namely, cytoplasm (CYT), endoplasmic reticulum (ER) and mitochondria (MT). The Hsp110 genes can be classified into the polyphyletic CYT group and the monophyletic ER group. The two nematode species encode two Hsp70 and two Hsp110 proteins localized in the ER and their highly heat-inducible genes contain introns. The different Hsp70 and Hsp110 groups appear to evolve following the model of independent or divergent evolution. These models can also explain the evolution of the ER and MT genes. On the other hand, the CYT genes are divided into heat-inducible and constitutively expressed genes. The constitutively expressed genes probably have evolved by the birth-and-death process and the rates of gene birth-and-death are different among all organisms studied. The heat-inducible genes show an intra-species phylogenetic clustering, suggesting sequence homogenization, probably by gene conversion-like events. In addition, these genes show high levels of sequence conservation in both intra- and inter-species comparisons, and in most comparisons the amino acid sequence similarity was higher than the nucleotide. These results suggest that purifying selection also played a crucial role in sequence conservation of the Hsp70s. Therefore, we suggest that the CYT heat-inducible genes have apparently followed a mixed evolutionary pattern

with a combination of purifying selection, birth and death, and gene conversion-like events.

Joint work with Masatoshi Nei.

| | | |
|---|---|---|
| poster session | **Ron Y. Pinter**<br>Technion, Israel Institute of Technology | Evaluating a Class of Length-Sensitive Algorithms for Sorting by Reversal |

*Abstract*: Traditionally, comparative genomic studies that are based on sorting by reversal (SBR) use the unit cost model (merely counting the number of reversals) to measure the distance between genomes. In [1] we have introduced a length-sensitive model in which the cost f(x) of each reversal depends on the length, x, of the reversed sequence; the overall cost of the SBR process is the total of the individual reversals costs. Initially we looked at f(x)=x, but following feedback we received from plant biologists we studied the cost f(x)=x**a. We have a class of algorithms [2] that find an approximate cost (finding the minimal cost is NP-hard) for any positive value of the exponent a, but the question of which value of a is best is of great interest.

We decided to evaluate this by using the cost as a distance that is fed to a tool that builds phylogenetic trees and compare the results to evolutionary trees found using other methods. This gives rise to numerous issues, such as finding a good match score between trees that is appropriate to this study, how many common genes are necessary, if the number of common genes for the whole sample is too low - how do we put together partial results (i.e. combining trees that were built on subsets of sample) and how small can the subsets be, how to "learn" the best value of a, etc. The poster will describe the method and the results on two data sets and discuss their merits.

[1] Ron Y. Pinter and Steven Skiena. Genomic Sorting with Length-Weighted Reversals. Genome Informatics 13: 103-111 (2002).

[2] Michael A. Bender, Dongdong Ge, Simai He, Haodong Hu, Ron Y. Pinter, Steven Skiena, and Firas Swidan. Improved Bounds on Sorting with Length-Weighted Reversals. To appear in the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA'04), January 2004.

| | | |
|---|---|---|
| poster session | **Amal A. Shervington**<br>University of Central Lancashire | Induced CYP1A1 Gene Expression in Lung Cancer Cell Lines |

*Abstract*: The gene CYP1A1 (cytochrome P450, family A polypeptide 1), encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases that catalyze numerous reactions involved in drug metabolism and synthesis of cholesterol, steroids and lipids. The enzyme is reported to be present predominantly in extrahepatic tissues in humans and in experimental animals (1). CYP1A1 is of toxicological importance because it catalyses the bioactivation of polyaromatic hydrocarbon (PHA) constituents e.g. Benzo[a]pyrene and other combustion products abundant in tobacco smoke to mutagens and canrcinogens (2).

Several studies of the oncogenic significance of CYP1A1 have found correlation between inducibility of the enzyme and lung cancer susceptibility in smokers (3). The expression and activity of CYP1A1 were examined using either peripheral blood lymphocytes as surrogate for lung cancer tissue (3) or lung biopsy specimens from human subjects. CYP1A1 transcripts were detected in lung cancer tissue either by reverse transcription polymerase chain reaction (RT-PCR) or northern blot hybridization (4).

In our laboratory we used four different lung cell lines: A549 Adenocarcinoma; H460 large cell carcinoma; COR-L23/5010 drug resistance large cell carcinoma and CCD-32Lu normal lung cells as a control. We measured the level of CYP1A1 transcript using the LightCycler (quantitative PCR). mRNA extracted from 106 cells using mRNA capture kit (Roche) were used to generate cDNA by Reverse Transcription System (Roche) with CYP1A1 primers (designed using primer3 web site) and amplified by the LightCycler using CYP1A1

The size of the CYP1A1 amplicon expected were 166bp, which was expressed at a highly induced level in the A549 Adenocarcinoma and to a less extent in the H460 large cell carcinoma. Very faint bands can be seen in L23/5010 drug resistance large cell carcinoma. No CYP1A1 can be detected in the normal lung cells. An amplicon of 300bp was amplified only in the control and not in the cancerous cell lines. Further work is required to characterise the 300bp band and to identify its significance.

Our results have shown an induced level of CYP1A1 in the adenocarcinoma cell line which is absent from the control, indicating that CYP1A1 is expressed at elevated level in some cancer cell line but not in the control.

Numerous citation have emphasised on the induction level of CYP1A1 in peripheral blood lymphocytes and lung cancer tissue but there have been no or few reports on the level of CYP1A1 in established cancer such as cancerous cell lines. Our

study has shown elevated level of CYP1A1 in some of the cancerous cell lines, which may suggest an active role for the CYP1A1 in the maintenance of cancer.

Joint work with Kulthum Mohammed.

| | | |
|---|---|---|
| poster session | **Derek E. Wildman**<br>Wayne St. University School of Medicine | An Objective View of Humankind's Place in Primate Evolution |

*Abstract*: In order to accurately place humankind in a phylogenetic classification of Primates it is necessary to know the phylogenetic relationships among all members of the order. We present the phylogenetic relationships and times of divergence for extant members of the order as determined by DNA nucleotide sequence data, and we focus particularly on the relationships within the family Hominidae. Local molecular clock analyses using fossil calibrations calculate that the time of origin for the order Primates as a crown group is 63 million years ago. Anthropoid primates (New World monkeys, Old World monkeys, and apes including humans) originated approximately 40 million years ago.

Phylogenetic and local molecular clock analyses from a sample of 97 genes show that humans and chimpanzees form a clade that most recently shared a common ancestor between 5 and 6 million years ago. These coding DNA data separate the human-chimpanzee clade from the gorilla clade between 6 and 7 million years ago. This African ape clade separated from the orangutan clade between 13 and 15 million years ago. We calculated the percent nonsynonymous DNA identity between humans and chimpanzees to be 99.4%, synonymous identity to be 98.4%, and total DNA sequence identity to be 99.1%. Interestingly, phylogenetic analysis grouped humans and chimpanzees together when only nonsynonymous sites were analyzed. This result suggests that at the protein level humans and chimpanzees are functionally more similar to each other than either taxon is to any other ape. Additionally, of these 97 genes, 30 show evidence of positive selection during the descent of catarrhine primates. An equal number (n=14) of these genes show elevated nonsynonymous rates of substitution on the human and chimpanzee lineages.

Divergences between humans and chimpanzees are placed in perspective by comparing their date of divergence with those found across the class Mammalia. The age of genus level crown groups for mammals ranged from 2 to 21 million years old. The mean crown group time of origin is approximately 8 million years ago, and the 95% confidence interval falls between 6.61 and 9.71 million years ago. Thus, humans and chimpanzees more recently share a common ancestor than do many congeneric groups of mammals.

Joint work with Monica Uddin, Guozhen Liu, Lawrence I. Grossman, and Morris Goodman.

| | | |
|---|---|---|
| poster session | **Tiffani L. Williams**<br>University of New Mexico | Searching for Optimal Trees Under Maximum Parsimony |

*Abstract*: See http://www.ima.umn.edu/complex/abstracts/williams/poster-abstract.pdf.

**Tuesday, October 21**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

| | | |
|---|---|---|
| 9:00 | **Coffee** | Reception Room EE/CS 3-176 |
| 9:30 | **Bret Larget**<br>University of Wisconsin - Madison | A Statistical Approach to the Estimation of Phylogeny from Genome Arrangements |

*Abstract*: The determination of evolutionary relationships is a fundamental problem in evolutionary biology. Genome arrangement data offers a source of information for estimating phylogenetic trees that may be especially useful for distantly related species. A statistical approach to phylogenetic information is concerned with assessment of uncertainty in estimated phylogenetic trees. We describe a Bayesian framework for phylogenetic inference from genome arrangement data using Markov chain Monte Carlo and discuss our results on several data sets.

| | |
|---|---|
| 10:20 | **Discussion** |

| | | |
|---|---|---|
| 10:30 | **Coffee Break** | Reception Room EE/CS 3-176 |

| | | |
|---|---|---|
| 11:00 | **Li-San Wang**<br>University of Pennsylvania | Distance-Based Genome Rearrangement Phylogeny |

*Abstract*: Evolution operates on whole genomes through mutations that change the order and strandedness of genes within the genomes. These events are examples of "rare genomic changes," which have low frequency and high signal-to-noise ratio. Thus analyses of gene-order data present new opportunities for discoveries about deep evolutionary events, provided that sufficiently accurate methods can be developed to reconstruct evolutionary trees.

In this talk I will present our results in distance-based genome rearrangement phylogeny reconstruction. We approach the problem by developing new statistically-based true evolutionary distance estimators. These estimators are based on the distributions of genomic distances including breakpoint and inversion distances under Markov Models. In our simulation study, we obtain highly accurate trees by using these new distance estimators, even when the amount of evolution in the dataset is high.

This is joint work with Robert K. Jansen and Tandy Warnow at the University of Texas, and Bernard M.E. Moret at the University of New Mexico.

| | | |
|---|---|---|
| 11:50 | **Discussion** | |

| | | |
|---|---|---|
| 12:00 | **Lunch Break** | |

| | | |
|---|---|---|
| 1:30 | **Stephen J. O'Brien**<br>Laboratory of Genomic Diversity, National Cancer Institute-Frederick | The Landscape of Comparative Genomics in Mammals |

*Abstract*: Dense genetic maps of human, mouse and rat genomes that are based on coding genes, microsatellite and single nucleotide polymorphism (SNP) markers have been complemented by precise gene homologue alignment with moderate resolution maps of livestock, companion animals and additional mammal species. Comparative genetic assessment expands the utility of these maps in gene discovery, in functional genomics, and in tracking the evolutionary forces that sculptured the genome organization of modern mammalian species.

| | | |
|---|---|---|
| 2:20 | **Discussion** | |

| | | |
|---|---|---|
| 2:30 | **Coffee Break** | Reception Room EE/CS 3-176 |

| | | |
|---|---|---|
| 3:00 | **Pavel Pevzner**<br>University of California San Diego | Transforming Men into Mice: Lessons from Human and Mouse Genomic Sequences |

*Abstract*: Despite some differences in appearance and habits, men and mice are genetically very similar. In a pioneering paper, Nadeau and Taylor, 1984 estimated that surprisingly few genomic rearrangements (about 200) have happened since the divergence of human and mouse 75 million years ago. The genomic sequences of human and mouse provide evidence for a larger number of rearrangements than previously thought and shed some light on previously unknown features of mammalian evolution. In particular, they provide evidence for extensive re-use of breakpoints from the same relatively short regions and reveals a great variability in the rate of micro-rearrangements along the genome. Our analysis also implies the existence of a large number of very short "hidden" synteny blocks that were invisible in comparative mapping data and were ignored in previous studies of chromosome evolution. These results suggest a new model of chromosome evolution that postulates that breakpoints are chosen from relatively short fragile regions that have much higher propensity for rearrangements than the rest of the genome.

This is a joint work with Glenn Tesler.

| | | |
|---|---|---|
| 3:50 | **Discussion** | |
| 4:00 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 4:30 | **Second Chances** | Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions. |

---

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

---

**Wednesday, October 22**

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

| | | |
|---|---|---|
| 9:00 | **Coffee** | Reception Room EE/CS 3-176 |
| 9:30 | **Elisabeth R.M. Tillier**<br>Ontario Cancer Institute/University of Toronto | Models and Methods for Phylogenomics |

*Abstract*: I will present a number of new approaches to some fundamental problems in comparative genomics and sequence analysis:
1. Using gene order information to confirm orthologous identifications.
2. Using phylogenetic profiles for the phylogenetic analysis of whole genomes.
3. Development of substitution models for the analysis of protein and RNA sequences.

| | | |
|---|---|---|
| 10:20 | **Discussion** | |
| 10:30 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 11:00 | **Aoife McLysaght**<br>Trinity College Dublin, Ireland | Poxviruses and Adaptive Genome Evolution |

*Abstract*: We used complete sequence from twenty poxviruses to investigate the evolution of these virus genomes. We examined the pattern of genome content and genome arrangement evolution in the context of the virus phylogeny. We also examined the patterns of positive selection acting on genes in these genomes. We show that the rate of genome evolution is not constant over time, and that it may be possible to relate patterns of genome evolution and adaptive evolution acting on genes.

| | |
|---|---|
| 11:50 | **Discussion** |

| | | |
|---|---|---|
| 12:00 | **Lunch Break** | |

| | | |
|---|---|---|
| 1:30 | **Dannie Durand**<br>Carnegie Mellon University | Gene Clusters in Comparative Genomics: Accident or Design? |

*Abstract*: Large scale gene duplication, the duplication of whole genomes and subchromosomal regions, is a major force driving the evolution of genetic functional innovation. Whole genome duplications are widely believed to have played an important role in the evolution of the maize, yeast and vertebrate genomes. Two or more linked clusters of similar genes found in distinct regions on the same genome are often presented as evidence of large scale duplication. However, as the gene order and the gene complement of duplicated regions diverge progressively due to insertions, deletions and rearrangements, it becomes increasingly difficult to distinguish remnants of common ancestral gene order from coincidental similarities in genomic organization. In this talk, I present computational approaches to validating gene clusters in comparative genomics.

| | | |
|---|---|---|
| 2:20 | **Discussion** | |
| 2:30 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 3:00 | **Ross Overbeek**<br>Fellowship for Interpretation of Genomes (FIG) | Exploiting Gene Clusters to Curate Annotations |

*Abstract*: Previously, we argued that gene clustering on prokaryotic genomes was the key to locating "missing genes," and we demonstrated that the technique worked remarkably well. The use of clusters is also the key to straightening out many of the assignments that could not be made precisely based only on similarities and motifs. We will consider the case of gene clusters related to leucine degradation as an example; they occur in phylogentically diverse organisms, and many of the genes involved currently have inaccurate or imprecise annotations. Comparative analysis of clusters, as well as occurrence profiles, can be used to methodically construct chains of assignments that follow from a few basic observations. This sets the stage where a single carefully chosen wet lab confirmation can confirm or reject a large number of assignments, often removing ambiguities from tens if not hundreds of genes.

| | | |
|---|---|---|
| 3:50 | **Discussion** | |
| 4:00 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 4:30 | **Second Chances** | Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions. |

## Thursday, October 23

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

| | | |
|---|---|---|
| 9:00 | **Coffee** | Reception Room EE/CS 3-176 |
| 9:15 | **Ron Y. Pinter**<br>Technion, Israel Institute of Technology | Evaluating a Class of Length-Sensitive Algorithms for Sorting by Reversal |

*Abstract*: Traditionally, comparative genomic studies that are based on sorting by reversal (SBR) use the unit cost model (merely counting the number of reversals) to measure the distance between genomes. In [1] we have introduced a length-sensitive model in which the cost f(x) of each reversal depends on the length, x, of the reversed sequence; the overall cost of the SBR process is the total of the individual reversals costs. Initially we looked at f(x)=x, but following feedback we received from plant biologists we studied the cost f(x)=x**a. We have a class of algorithms [2] that find an approximate cost (finding the minimal cost is NP-hard) for any positive value of the exponent a, but the question of which value of a is best is of great interest.

We decided to evaluate this by using the cost as a distance that is fed to a tool that builds phylogenetic trees and compare the results to evolutionary trees found using other methods. This gives rise to numerous issues, such as finding a good match score between trees that is appropriate to this study, how many common genes are necessary, if the number of common genes for the whole sample is too low - how do we put together partial results (i.e. combining trees that were built on subsets of sample) and how small can the subsets be, how to "learn" the best value of a, etc. The poster will describe the method and the results on two data sets and discuss their merits.

[1] Ron Y. Pinter and Steven Skiena. Genomic Sorting with Length-Weighted Reversals. Genome Informatics 13: 103-111 (2002).

[2] Michael A. Bender, Dongdong Ge, Simai He, Haodong Hu, Ron Y. Pinter, Steven Skiena, and Firas Swidan. Improved Bounds on Sorting with Length-Weighted Reversals. To appear in the Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA'04), January 2004.

| | | |
|---|---|---|
| 9:35 | **Discussion** | |
| 9:45 | **Jijun Tang**<br>University of New Mexico | Large-scale Phylogeny Reconstruction from Arbitrary Gene-order Data |

*Abstract*: See http://www.ima.umn.edu/complex/abstracts/tang/TangAndMoret.pdf.

| | | |
|---|---|---|
| 10:05 | **Discussion** | |
| 10:15 | **Allan Force**<br>Benaroya Research Inst., Virginia Mason | Origin of Subfunctions and Modular Genes |

*Abstract*: Evolutionary explanations for the origin of modular genetic and developmental pathways almost always invoke some sort of long-term selective advantage, e.g., as a functional prerequisite to the evolution of phenotypic complexity or as an enhancer of evolvability. However, simple theoretical results demonstrate that even in the absence of any direct selective advantage, genetic modularity can spontaneously emerge through the acquisition of new gene subfunctions. Provided that population size is sufficiently small, random genetic drift and mutation can conspire to produce changes in the underlying genetic architecture of a species without necessarily altering the phenotype. Extensive genetic modularity may then accrue in a near-neutral fashion in permissive population- genetic environments, potentially opening novel pathways to morphological evolution. These results provide additional support for the proposition that many aspects of gene and genome complexity in multicellular eukaryotes may have arisen passively as population size reductions accompanied an increase in organism size, with the adaptive exploitation of such complexity occurring secondarily.

| | | |
|---|---|---|
| 10:35 | **Discussion** | |
| 10:45 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 11:00 | **Elizabeth Ann Housworth**<br>Measures of Conserved Synteny | Indiana University |

*Abstract*: Measures of conserved synteny are important for estimating the relative rates of chromosomal evolution in various lineages. We present a natural way to view the synteny conservation between two species from an Oxford grid–an r x c table summarizing the number of orthologous genes on each of the chromosomes 1 through r of the first species that are on each of the chromosomes 1 through c of the second species. This viewpoint suggests a natural statistic, which we call syntenic correlation, designed to measure the amount of synteny conservation between two species. This measure allows syntenic conservation to be compared across many pairs of species. We also discuss incorporating the dependency of the numbers of orthologues observed in the chromosome pairings between the two species into the estimates of the true number of conserved syntenies given the observed number of conserved syntenies.

| | | |
|---|---|---|
| 11:50 | **Discussion** | |
| 12:00 | **Lunch Break** | |
| 1:30 | **Fiona Brinkman**<br>Simon Fraser University, Burnaby, BC, Canada | Analysis of Horizontal Gene Transfers of Potential Relevance to Microbial Virulence |

*Abstract*: We have been using genome-wide bioinformatic approaches to identify horizontal gene transfers that are of interest for their potential role in bacterial virulence and the evolution of pathogenic microbes. Analyses of both bacteria-eukaryotic and bacteria-bacteria gene transfers are summarized, revealing possible patterns in the types of genes most often transferred between species. The implications are discussed, both in the context of the evolution of virulence and what is likely the most effective approach to control infectious disease agents.

| | | |
|---|---|---|
| 2:20 | **Discussion** | |
| 2:30 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 3:00 | **Emmanuelle Lerat**<br>University of Arizona | Lateral Gene Transfers and Organismal Phylogeny in Bacteria: Implications for Ancestral Genome Reconstruction |

*Abstract*: Genome reconstruction is of particular interest from a biological perspective. This knowledge can illuminate the history of events that led to the present contents and organization of genomes. The principle of reconstruction methods is the inference of rearrangements that occurred during the history of the genome. This makes the strong assumption that genes are faithfully transmitted with their genome through generations. However in bacteria, lateral (or horizontal) gene transfers (LGT) are known to be very numerous. LGT might be an obstacle in the attempt to establish genome history, because some homologous genes may be transmitted between different species. It has even been argued that LGT may prevent the establishment of organismal relationships based on individual gene phylogenies. Thus to reconstruct ancestral genomes in bacteria seems to be particularly hazardous unless LGT is taken into account. It is therefore very important to test the hypothesis of vertical transmission of the genes used in genome reconstruction. In order to determine the impact of LGT on the potential organismal phylogeny, an approach to multigene phylogeny using complete genomes is necessary to identify the genes that have been, without ambiguity, vertically transmitted and that are thus good candidates to be used in genome reconstruction. This will allow a real biological interpretation of the genome reconstruction but also facilitate the reconstruction itself.

| | | |
|---|---|---|
| 3:50 | **Discussion** | |
| 4:00 | **Coffee Break** | Reception Room EE/CS 3-176 |

| | | |
|---|---|---|
| 4:30 | **Second Chances** | Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions. |

---

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

---

| | | |
|---|---|---|
| 6:00 pm | **Workshop Dinner** | Gardens of Salonica<br>19 Fifth Street NE, Minneapolis, Tel (612) 378-0611. |

## Friday, October 24

**All talks are in Lecture Hall EE/CS 3-180 unless otherwise noted.**

| | | |
|---|---|---|
| 9:00 | **Coffee** | Reception Room EE/CS 3-176 |
| 9:30 | **Kenneth H. Wolfe**<br>University of Dublin, Trinity College | Genome Evolution and Sorting Out Ancient Polyploidy in Yeasts |

*Abstract*: Yeasts are a good model system for investigating gene order and chromosomal evolution because their genomes are compact and relatively eas ing in a metabolic pathway was put together during the evolution of species that can grow vigorously without oxygen.

| | | |
|---|---|---|
| 10:20 | **Discussion** | |
| 10:30 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 11:00 | **Jens Lagergren**<br>SBC (Stockholm Bioinformatics Center), & KTH (Kunliga Tekniska Högskolan) | Bayesian Gene/Species Tree Reconciliation and Orthology Analysis Using MCMC |

*Abstract*: Comparative genomics in general and orthology analysis in particular are becoming increasingly important parts of gene function prediction. Previously, orhtology analysis and reconciliation has been performed only with respect to the parsimony model. This discards many plausible solutions and sometimes precludes finding the correct one. In many other areas in bioinformatics probabilistic models have proven to be both more realistic and powerful than parsimony models.

We introduce a probabilistic gene evolution model based on a birth-death process in which a gene tree evolves "inside" a species tree. Based on this model, we develop a tool with the capacity to perform practical orthology analysis, based on Fitch's original definition, and more generally for reconciling pairs of gene and species trees. Our gene evolution model is biologically sound and intuitively attractive. We develop a Bayesian analysis based on MCMC which facilitates approximation of an a posteriori distribution for reconciliations. That is, we can find the most probable reconciliations and estimate the probability of any reconciliation, given the observed gene tree. This also gives a way to estimate the probability that a pair of genes are orthologs. To the best of our knowledge, this is the first successful introduction of this type of probabilistic methods, which flourish in phylogeny analysis, into reconciliation and orthology analysis.

The MCMC algorithm has been implemented and performs very well on synthetic as well as biological data. Using standard correspondences, our results carry over to allele trees as well as biogeography.

| | | |
|---|---|---|
| 11:50 | **Discussion** | |
| 12:00 | **Lunch Break** | |
| 1:30 | **Evan Eichler**<br>Case Western Reserve University | Recent Segmental Duplications and the Fragile Breakage Model of Human Genome Evolution |

*Abstract*: It has been estimated that 5% of the human genome consists of interspersed duplicated material that has arisen over the last 30-40 million years of evolution. A large proportion of these duplications exhibits an extraordinarily high degree of sequence identity at the nucleotide level (¿95%) and are interspersed over large genomic distances (¿1 Mb). The distribution of these duplications is non-random in the human genome. Through processes of non-allelic homologous recombination, these same regions are targets for rapid evolutionary turnover creating hotspots of mammalian chromosomal evolution and sites of genomic instability associated with disease within the human population. Preliminary analyses have suggested that the amount of segmental duplication may be a relatively unique property of our genome. We have developed systematic experimental and computational tools to examine duplication content from human and other sequenced vertebrate species. An analysis of the breakpoints of these duplications shows a significant enrichment of Alu-repeat elements, providing new insight into their mechanism of origin and preeminence within the primate genome. In additions based on our analysis of syntenic breakpoints between the mouse and human genome, we find that 25% (122/461) of mouse-human synteny breakpoints contain 10 kb of duplicated sequence. This association is highly significant (P¡0.0001) when compared to a simulated random breakage model. These data support a non-random model of chromosomal evolution that implicates a predominance of both small-scale duplication and large-scale evolutionary rearrangements within specific regions of the human genome. Such properties should be considered when trying to reconstruct the evolutionary history of mammalian genomes.

| | | |
|---|---|---|
| 2:20 | **Discussion** | |
| 2:30 | **Coffee Break** | Reception Room EE/CS 3-176 |
| 3:00 | **Second Chances** | Speakers of the day respond to further questions, suggestions, re-frame their main points, look toward future directions. |
| 3:30 | **Concluding Remarks by Organizers** | |
| 3:40 | **End of Workshop** | |

**Monday, October 27**

The 10:30 IMA break will be in Lind Hall 400.

**Tuesday, October 28**

The 10:30 IMA break will be in Lind Hall 400.

**IMA POSTDOC SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 11:15-12:15 | **Mohammad Kazim Khan**<br>Kent State University | On SPRT and CUSUM Procedures and Some Open Problems |

*Abstract*: Let $Y_1, Y_1, \cdots, Y_{\nu-1}$ be iid random variables with a distribution function (df) $F_0(y)$, and let $Y_\nu, Y_{\nu+1}, \cdots$ be iid random variables with a df $F_1(y)$, where $\nu$ is an unknown time index of a change in distribution with the change in parametric value. For a suitable function $\psi$ let $X_j = \psi(Y_j)$ denote some convenient data reduction or it may be defined by certain optimality consideration such as the Sequential Probability Ratio Test. For detecting a change-point in the distribution, Page (1954) defined his famous cusum (cumulative sum) procedure. From the inherent renewal property of the cusum, Page noted that $EN = EM/P(S_M \geq h)$, where $N$ is the cusum stopping rule and $M$ is the SPRT and $S_n$ represents the partial sum of the information $X_1, X_2, ...$. The constant $h$ is the trigering constant for the cusum. This link between the SPRT and the cusum is quite useful in approximating and/or evaluating $EN$. However, a deeper connection between $N$ and $M$ is known that I will try to present. The purpose of this exposition is to further exploit such a relationship between $N$ and $M$ to study the properties of some one sided and two sided cusums with several applicable examples. We will see how exact results can be computed in discrete time settings. There are some open problems which I will outline.

The IMA Postdoc Seminar is organized by Balaji Gopalakrishnan and Antar Bandyopadhyay.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

**Wednesday, October 29**

The 10:30 IMA break will be in Lind Hall 400.

**BROWN BAG SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 12:00 | **Huiqiang Jiang**<br>University of Minnesota | Singular Elliptic Equation in Thin Film Rupture |

*Abstract*: We consider nonnegative solutions of a singular elliptic equation, which arise in thin film rupture and minimal surface theory. We get a general estimate of the size of singular (zero) set. And other results and open problems will also be discussed.

**COMPLEX SYSTEMS SEMINAR, Lind Hall 409:**

| | | |
|---|---|---|
| 3:30 pm | **Shmuel Friedland**<br>University of Illinois at Chicago | Missing Value Estimation Methods for DNA Microarrays |

*Abstract*: In this talk we first survey the known methods for estimation for missing values in gene microaarays:

1. Row average.
2. Clustering methods.
3. Singular value decomposition (principal-component analysis).

The draw back of these methods is that the recovery of the missing data is done independently, i.e. the completion of each missing value does not influence the completion of other values.

Next we discuss a new method for a completion of missing values in which the completion is done simultaneously. It is inspired by the methods for solutions of the inverse eigenvalue problems.

**Thursday, October 30**

The 10:30 IMA break will be in Lind Hall 400.

| | | |
|---|---|---|
| 12:45-2:15 pm | **Random Matrices**<br>G. Anderson and O. Zeitouni | This course (Math 8660) meets in Physics 133 |

**Friday, October 31**

The 10:30 IMA break is canceled; 3:00 Hallowe'en party instead.

**IMA/MCIM INDUSTRIAL PROBLEM SEMINAR, 570 Vincent Hall:**

| | | |
|---|---|---|
| 1:25pm | **Nikos Paragios**<br>Siemens Corporate Research Princeton, NJ | Variational Methods, Implicit Representation & Visual Grouping |

*Abstract*: Segmentation and extraction of structures of interest refers to a core component of Imaging and Vision with a variety of applications. One can think that this task is equivalent with the separation of a bounded domain (image/volume) into regions with consistent properties. Such properties can be defined on an arbitrary space, like intensity properties, texture properties, motion properties, stereo properties, etc.

An elegant tool to perform such grouping is by the propagation of curves aiming at separating regions with consistent characteristics. Such propagation is either derived from the minimization of an objective function, or defined according to the application objectives (geometric flows). Implicit representations and level set methods is an emerging technique to perform this task. In this talk, we will describe variants of this technique to perform visual grouping. First, a connection between curve/surface propagation and level set methods will be established. Then, we will propose a general formulation to address grouping using implicit representations. Such formulation can account for various information cues. Boundary, Regional (intensity, motion, stereo, texture), prior knowledge (shape of the structure of interest) and user interaction will be the major components of our approach.

The 3:00 IMA Hallowe'en Party will be in Lind Hall 400.

*******************************************************************************

## PART III: CURRENT IMA PARTICIPANTS

FIRST YEAR POSTDOCTORAL MEMBERS

| NAME | PREVIOUS INSTITUTION |
|---|---|
| Gerard Awanou | University of Georgia |
| Karen Ball | Indiana University |
| Antar Bandyopadhyay | UC Berkeley |
| Tim Garoni | University of Melbourne |
| Chuan-Hsiang Han | North Carolina State University |
| Lea Popovic | UC Berkeley |

SECOND YEAR POSTDOCTORAL MEMBERS

| NAME | PREVIOUS INSTITUTION |
|---|---|
| Olga Brezhneva | Russian Academy of Sci. |
| Lisa Evans | Georgia Tech |
| Balaji Gopalakrishnan | Georgia Tech |
| Herve Kerivin | University Blaise Pascal-France |
| Tamon Stephen | University of Michigan |
| Jing Wang | University of Minnesota |

POSTDOCTORAL MEMBERS IN INDUSTRIAL MATHEMATICS

| NAME | PREVIOUS INSTITUTION | INDUSTRIAL AFFILIATION |
|---|---|---|
| Yusuf Altundas | University of Pittsburgh | Schlumberger |
| Lili Ju | Iowa State University | VA Hospital |
| Haewon Nam | Texas A & M University | GE |
| Jun Zhao | Texas A & M University | Schlumberger |

LONG TERM VISITORS

| NAME | HOME INSTITUTION |
|---|---|
| Soohan Ahn | Seoul National University (SRCCS) |
| Montaz Ali | Witwatersrand University |
| Greg Anderson | University of Minnesota |
| Maury Bramson | University of Minnesota |
| Laura Chihara | Carleton College |
| Shmuel Friedland | University of Illinois - Chicago |
| Naresh Jain | University of Minnesota |
| Christina Kendziorski | University of Wisconsin - Madison |
| Mohammad Kazim Khan | Kent State University |
| Dohyun Kim | Seoul National University (SRCCS) |
| Thomas G. Kurtz | University of Wisconsin - Madison |
| Taerim Lee | Korea National Open University |
| Richard P. McGehee | University of Minnesota |
| Michael Newton | University of Wisconsin - Madison |
| Amir Niknejad | University of Illinois - Chicago |
| Greg Rempala | University of Louisville |
| Arnd Scheel | University of Minnesota |
| Simon Tavare | University of Southern California |
| Stephen J. Willson | Iowa State University |
| Yuhong Yang | Iowa State University |
| Ofer Zeitouni | University of Minnesota |

VISITORS IN RESIDENCE (as of 29 September 2003)

| | | |
|---|---|---|
| Lars Arvestad | Universitet of Stockholm | 10/19/03 – 10/24/03 |
| Pierre Baldi | U of California - Irvine | 10/20/03 – 10/23/03 |
| Anne Bergeron | Universite du Quebec a Montreal | 10/19/03 – 10/25/03 |
| Ann-Charlotte Berglund | Stockholm U | 10/20/03 – 10/24/03 |
| Dimitris Bertsimas | MIT | 10/10/03 – 10/12/03 |
| Guillaume Bourque | Universite de Montreal | 10/19/03 – 10/24/03 |
| Fiona S. Brinkman | Simon Fraser U | 10/22/03 – 10/24/03 |
| Steven Cannon | U of Minnesota | 10/20/03 – 10/24/03 |
| Alberto Caprara | DEIS, U of Bologna | 10/18/03 – 10/22/03 |
| Dimitra Chalkia | Pennsylvania State U | 10/18/03 – 10/24/03 |
| Avril Coghlan | Trinity College Dublin | 10/19/03 – 10/24/03 |
| Jack R. Collins | National Cancer Institute | 10/19/03 – 10/24/03 |
| Indraneel Das | United Technologies Research | 10/09/03 – 10/10/03 |
| Ramana V. Davuluri | Ohio State U | 10/19/03 – 10/24/03 |
| Dannie Durand | Carnegie Mellon | 10/19/03 – 10/24/03 |
| Evan Eichler | Case Western Reserve U | 10/22/03 – 10/24/03 |
| Nadia El-Mabrouk | Universite de Montreal | 10/19/03 – 10/24/03 |
| Scott C. Fahrenkrug | U of Minnesota | 10/19/03 – 10/24/03 |
| Allan Force | Virginia Mason Research Center | 10/19/03 – 10/24/03 |
| Anant Godbole | East Tennessee State U | 10/18/03 – 10/24/03 |
| Josefa Gonzalez | Universitat Autonoma de Barcelona | 10/18/03 – 10/25/03 |
| Lawrence F Gray | U of Minnesota | 10/11/03 – 10/12/03 |
| Roderic Guigo | Inst. Municipal de Investigacion Medica | 10/19/03 – 10/24/03 |
| Michael Hallett | McGill U | 10/19/03 – 10/24/03 |
| Tzvika Hartman | Weizmann Institute of Science | 10/19/03 – 10/24/03 |
| Dirk Holste | MIT | 10/19/03 – 10/24/03 |
| Elizabeth Housworth | Indiana U | 10/19/03 – 10/24/03 |
| Richard D. James | Heriot-Watt U | 10/10/03 – 10/12/03 |
| Sabera Kazi | Honeywell | 10/20/03 – 10/24/03 |
| Jon Kettenring | Bell Communications Research | 10/10/03 – 10/12/03 |
| Martin Kreitman | U of Chicago | 10/19/03 – 10/24/03 |
| Jens Lagergren | Royal Institute of Technology, Stockholm | 10/19/03 – 10/24/03 |
| Bret Larget | U of Wisconsin - Madison | 10/19/03 – 10/24/03 |
| Emmanuelle Lerat | Universite Lyon 1 | 10/19/03 – 10/25/03 |
| C. Randal Linder | U of Texas - Austin | 10/19/03 – 10/25/03 |
| Bill Long | Cray, Inc. | 10/20/03 – 10/24/03 |
| Michael Lynch | U of Indiana | 10/19/03 – 10/24/03 |
| Wojciech Makalowski | Pennsylvania State U | 10/18/03 – 10/24/03 |
| Bob Mau | U of Wisconsin - Madison | 10/19/03 – 10/25/03 |
| Georgiana May | U of Minnesota | 10/20/03 – 10/24/03 |
| Aoife Mclysaght | Trinity College Dublin | 10/18/03 – 10/24/03 |
| Joao Meidanis | U of Campinas | 10/19/03 – 10/25/03 |
| Julia Mixtacki | Universität Bielefeld | 10/19/03 – 10/24/03 |
| Bernard Moret | U of New Mexico | 10/19/03 – 10/23/03 |
| James C. Mullikin | National Institutes of Health | 10/20/03 – 10/24/03 |
| Bill Murphy | National Cancer Institute | 10/19/03 – 10/24/03 |
| Gene Myers | Celera Genomics Corporation | 10/19/03 – 10/24/03 |
| Luay Nakhleh | U of Texas - Austin | 10/19/03 – 10/24/03 |
| Nikolas Nikolaidis | Pennsylvania State U | 10/18/03 – 10/24/03 |
| Stephen J. O'Brien | National Cancer Institute | 10/19/03 – 10/24/03 |
| Ross Overbeek | Argonne National Laboratory | 10/19/03 – 10/24/03 |
| Wei Pan | U of Minnesota | 10/19/03 – 10/24/03 |
| Nikos Paragios | Siemens Corporate Research | 10/30/03 – 10/31/03 |
| Pavel Pevzner | U of California - San Diego | 10/19/03 – 10/24/03 |

| | | |
|---|---|---|
| Ron Y. Pinter | Technion, Israel Institute of Technology | 10/19/03 – 10/24/03 |
| Juan F. Poyatos | Spanish National Cancer Centre (CNIO) | 10/19/03 – 10/25/03 |
| Raj Rajagopal | 3M | 10/20/03 – 10/24/03 |
| David Sankoff | U of Ottawa | 10/19/03 – 10/24/03 |
| Jeanette Schmidt | Incyte Genomics, Inc. | 10/18/03 – 10/24/03 |
| Jennifer Seffernick | U of Minnesota | 10/20/03 – 10/24/03 |
| Igor V. Sharakhov | U of Notre Dame | 10/19/03 – 10/24/03 |
| Amal Shervington | U of Central Lancashire | 10/19/03 – 10/26/03 |
| Christopher Stark | National Science Foundation | 10/10/03 – 10/12/03 |
| Jens Stoye | Universität Bielefeld | 10/18/03 – 10/24/03 |
| Dharmashankar Subramanian | Honeywell | 10/17/03 – 10/17/03 |
| De Witt L. Sumners | Flordia State U | 10/10/03 – 10/12/03 |
| William W. Symes | Rice U | 10/10/03 – 10/12/03 |
| Eitan Tadmor | U of Maryland | 10/11/03 – 10/12/03 |
| Jijun Tang | U of New Mexico | 10/18/03 – 10/24/03 |
| Eric Tannier | INRIA | 10/18/03 – 10/24/03 |
| Glenn Tesler | U of California - San Diego | 10/19/03 – 10/24/03 |
| Peter Tiffin | U of Minnesota | 10/20/03 – 10/24/03 |
| Elisabeth Tillier | Princess Margaret Hospital | 10/18/03 – 10/22/03 |
| Ali Tofigh | Kungliga Tekniska Hogskolan | 10/19/03 – 10/25/03 |
| Philippe Tondeur | U of Illinois - Urbana-Champaign | 10/10/03 – 10/12/03 |
| Vamsi Veeramachaneni | Pennsylvania State U | 10/18/03 – 10/24/03 |
| Li-San Wang | U of Texas - Austin | 10/18/03 – 10/24/03 |
| Tandy Warnow | U of Texas - Austin | 10/19/03 – 10/24/03 |
| Derek E. Wildman | Wayne State U | 10/18/03 – 10/24/03 |
| Ruth Williams | U of California - San Diego | 10/10/03 – 10/12/03 |
| Tiffani L. Williams | U of New Mexico | 10/18/03 – 10/24/03 |
| Ken Wolfe | Trinity College Dublin | 10/20/03 – 10/24/03 |
| Stacia K. Wyman | U of Texas - Austin | 10/19/03 – 10/24/03 |
| Weijia Xu | U of Texas - Austin | 10/19/03 – 10/24/03 |
| Mihalis Yannakakis | Stanford U | 10/10/03 – 10/12/03 |
| Weixiong Zhang | Washington U - St. Louis | 10/22/03 – 10/24/03 |
| Yongqing Zhang | U of Minnesota | 10/19/03 – 10/24/03 |

See also URL: `http://www.ima.umn.edu/people/`