

8th Bioinformatics Day on "Computational Problems in Biology" October 5th 2006

1PM - 5 PM in the Oxford Centre for Gene Function, Oxford.

October 5th a Bioinformatics Day with focus on Computer Science will take place. It is our intention to have such a half-day before every term. The day will be dedicated to interesting topics and will also be used to announce Genomics/Bioinformatics in the following term: seminars, study group and other meetings. If you intend to attend this event, please send an email to Madeline Mitchell (mitchell@stats.ox.ac.uk) by with your name and affiliation. This is necessary, so we can arrange coffee and biscuits appropriately and since there is a limit of a 100 in the seminar room in OCGF. Information on bioinformatics days can be found at <http://www.stats.ox.ac.uk/mathgen/bioinformatics/> under activities.

1.00pm Jotun Hein: Short Welcome

1.10pm Lior Pachter: **Computational Challenges in Comparative Genomics – the case of Drosophila**

The Drosophila comparative genomics project consists of the comparison and analysis of 12 whole genomes for the purpose of identifying functional elements and studying genome evolution. These genomes have recently been made available online, and a self-assembled analysis group has begun preliminary analyses, with results posted at <http://rana.lbl.gov/drosophila/wiki/>. This exciting project offers the opportunity for computer scientists and statisticians to develop and explore new methods for comparative genomics in a challenging setting dominated by large quantities of data. We will highlight a few interesting problems.

1.30pm Rahael Hauser: **Large Deviation Techniques for Computing the Asymptotic Similarity Rate of Random Sequences**

Optimal sequence alignments are usually chosen as maximisers of certain scoring functions among a set of admissible alignments. To make such alignments meaningful predictors of a true similarity between the aligned sequences, their maximising score should be significantly larger than the typical scores achieved by optimal alignments of independent random – and hence completely unrelated – sequences. However, the mathematical understanding of optimal alignment scores of random sequences is very incomplete. In this talk we discuss some recent progress in this field and show how large deviations techniques play a key role in the design of algorithms for the computation of some key constants that appear in this context.

1.50pm Dan Levy (with R. Mihaescu and L. Pachter): **Why Neighborjoining Works**

We show that the neighbor-joining algorithm is a robust quartet method for constructing trees from distances. This leads to a new performance guarantee that contains Atteson's optimal radius bound as a special case and explains many cases where neighbor-joining is successful even when Atteson's criterion is not satisfied. We also provide a proof for Atteson's conjecture on the optimal edge radius of the neighbor joining algorithm. The strong performance guarantees we provide also hold for the quadratic time fast neighbor-joining algorithm, thus providing a theoretical basis for inferring very large phylogenies with neighbor-joining.

2.10pm Rune Lyngsø: **Folding RNA Alignments**

Comparative information revealing compatible and compensating mutations significantly improves the accuracy of RNA secondary structure prediction. Numerous methods exist for predicting structure based on an alignment of a set of sequences. These all put the user at the mercy of the alignment used. Other methods try to simultaneously align the set of sequences and predict a common structure. These all suffer from prohibitive complexities. I will present efficient algorithms for structure prediction based on profile HMM probability distributions over multiple alignments of a set of sequences. Furthermore I will introduce a mathematical formulation of the alignment based structure prediction problem that can be seen as a first step towards investigations of the computational complexity of the problem.

2.30pm Vasile Palade: **Multi-Classifer Systems for Bioinformatics Problems**

Multi-Classifer Systems have fast been gaining popularity among researchers for their ability to fuse together multiple classification outputs for better accuracy and classification. I will first present some current issues on the design of multi-classifier systems before moving on to present our multi-classifier developments for some Bioinformatics problems, like gene identification in DNA sequences, microarray data analysis and gene regulatory networks.

2.50-3.30pm **BREAK**

3.30 pm Rob Freeman: **Analyzing Avian Navigation**

How animals navigate is of interest to numerous fields (animal behaviour, cognitive psychology, conservation, epidemiology, robotics). This talk will cover Machine Learning and Pattern Recognition techniques which we have applied to the problem of Avian Navigation. The talk will discuss techniques for the analysis of positional information, the application of these techniques to data acquired from GPS tracking of homing pigeons and sea birds and summarize methods for the analysis of the interactions between the recorded tracks and landmarks using aerial Photos and bathymetry.

3.50pm Chiu Fan Lee: **Molecular dynamics simulation of large biomolecules with discrete potentials**

Molecular dynamics simulation is an important tool in understanding biomolecules but its use is still very much limited by computational constraint. In this talk, I will present the method of discrete molecular dynamics which allows for clever exploitation of scheduling technique and as such provides considerable speed-up over conventional simulation method.

4.10 pm Peter Kohl: **Computational Modelling of the Heart: An Experimentalist's View**

The heart works as an externally homogeneous pump. This functional homogeneity arises in an internally highly heterogeneous system, in which no two cells have completely matching patterns of activity. In order to understand the workings of the heart it is necessary, therefore, to investigate cardiac structure and function at various levels of integration, from molecule to man. This poses considerable challenges for biological research, which may only be addressed in direct iteration of 'wet and dry modeling studies'. The central role of model selection and interdisciplinarity will be highlighted.

4.30pm Nicolas Wu: **A Model of Bacterial Chemotaxis using Process Algebra**

This talk will present basic notions of process algebra and will discuss its suitability for modelling problems in Systems Biology. As an example, the development of a simple model of chemotaxis in E. coli will be demonstrated using Communicating Sequential Processes, a process algebra, illustrating the principles of process refinement.

5.00 pm + **Beers at Lamb and Flag**