



## Bioinformatics Seminar

---

Speaker: David Sankoff, University of Ottawa

Title: Genome rearrangement algorithms in statistical and biological perspectives.

Date: Monday, 11 September 2006

Time & Location:

Refreshments: 11 am in the Theory of Computation Lab at MIT's Building 32, Stata Center Room G-575

Talk: 11:30 am the Theory of Computation Lab at MIT's Building 32, Stata Center, Room G-575

URL: <http://www-math.mit.edu/compbiosem/>

---

### Abstract:

To see whether the syntenic comparison of two genomes can reveal some signal of the specific rearrangement processes responsible for their evolutionary divergence we should carry out the rearrangement analysis of these genomes in parallel with the same analysis on pairs of randomized genomes. The total number of syntenic blocks, i.e., maximal regions homologous in the two genomes, is an indicator of the evolutionary divergence of two genomes, but contains no information about the evolutionary processes that caused this divergence.

In classical genetics, the main rearrangement events include inversions (reversals) of chromosomal segments, and reciprocal translocations of the prefixes or suffixes of two chromosomes. For a fixed number of syntenic blocks in two genomes, the number and nature of the events that intervene to differentiate them can be inferred from the number of cycles in the bicoloured graph - one colour edge for the block adjacencies in each genome - they induce.

It is these cycles induced by the two data genomes that should be assessed in comparison with those in random genomes to see if there remains an evolutionary signal. For random genomes, we have shown that the limiting expectation of the number of cycles  $c$  is

$$c = \max[G, H] + 2GH / (2G + 2H - 1) + 1/2 \log[n + \min(G, H) / (G + H)],$$

where  $G$  and  $H$  are the number of linear chromosomes in the two genomes. We discuss how this result was built up, using simulations, random graph theory and combinatorial recurrences. The evolutionary model allows only generalized inversions (including translocations, fusions and fissions) and general block interchange (including transposition).

When no signal about the specific events is apparent, i.e., when  $c$  is not significantly large, we can at least estimate the relative proportion of inversions versus translocations for fixed  $n$ . Our estimators are based on the numbers of chromosomes in one genome sharing blocks with each chromosome in the other genome.

---

The seminar is co-hosted by Professor Peter Clote of Boston College's Biology and Computer Science Departments and MIT Professor of Applied Math Bonnie Berger. Professor Berger is also affiliated with CSAIL & HST.

Massachusetts Institute  
of Technology  
77 Massachusetts Avenue  
Cambridge, MA 02139

*For General Questions, please contact [kvdickey@mit.edu](mailto:kvdickey@mit.edu)*