MIT
Department of Mathematics
& The Theory of
Computation Group
At CSAIL

Bioinformatics Seminar

Speaker: Michael Brudno, University of Toronto
Title: Reconstructing Ancestral Genome Order via Maximum Matching
Date: Monday, 28 November 2005
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:
The ancestral order of two genomes can be used, for example, for progressive
alignment. The difficulty of applying progressive alignment to whole genomes or other
sequences with rearrangements is due to the lack of a general ordering on the
alignments - it is possible to order based on either of the two genomes, but not both. We
present a novel alignment algorithm that reconstructs a likely ordering of the genome of
the ancestor of two organisms. While the genome of this ancestor is not known, one can
use the outgroup genomic sequences to establish this order (if reconstructing the
ancestor of mouse and rat, then human, dog, and chicken are all outlgroups). By making
several assumptions we formulate the problem of ordering a set of alignments based on
a likely ancestral order as a maximum matching problem.

By using the algorithm above we order these according to their most recent common
ancestor, and then align the resulting chain of alignments. This approach has several
advantages over previous algorithms: 1) it does not assume a base genome, to which all
other genomes are aligned, but creates a symmetric alignment equally valid for all
genomes, 2) it penalizes various rearrangement events progressively based on an
evolutionary tree, creating a set of alignments that mirrors the evolutionary history of the
sequences, and 3) it is able to align short, low similarity syntenic areas based on their
adjacency to higher similarity areas even when there has been a rearrangement event
between the two areas. I will present the results of some first analyses of whole genome
alignment using this algorithm.