A Flexible Approach to Genome Map Assembly

Eric Harley
eharley@db.toronto.edu

Anthony J. Bonner
bonner@db.toronto.edu

University of Toronto
Department of Computer Science
Toronto, Ont, Canada M5S 1A4

Abstract

A major goal of the Human Genome Project is to construct detailed physical maps of the human genome. A physical map is an assignment of DNA fragments to their locations on the genome. Complete maps of large genomes require the integration of many kinds of experimental data, each with its own forms of noise, experimental error and anomalies. To facilitate the assembly of physical maps, we are developing a flexible approach to map assembly based on logic programming and data visualization. Logic programming provides a convenient and mathematically rigorous way of reasoning about data, while data visualization provides layout algorithms for assembling and displaying genome mapping data. The approach has several benefits: (*i*) it frees the user (or programmer) from data management problems, (*ii*) it allows him to see anomalies in the data visually, (*iii*) it helps him to rapidly develop and test rules for filtering out anomalous data and for combining data from different sources, and (*iv*) it displays a genome map as a graph with linear structure. To demonstrate the approach, this paper describes numerous rules for map assembly implemented in a data-visualization system called Hy+. Using these rules, we have successfully assembled contigs (partial maps) from real and simulated mapping data—data that is noisy, imprecise and contradictory.

Appears in *Proceedings of the International Symposium on Intelligent Systems for Molecular Biology (ISMB)*, pages 161–169. AAAI Press, Menlo Park, California. Held at Stanford University, August 14–17 1994.

This paper is available at the following URL:
ftp://db.toronto.edu/pub/bonner/papers/genome.mapping/ismb94.ps.gz