PatternHunter: Faster And More Sensitive Homology Search

Bin Ma
Department of Computer Science
University of Western Ontario, Canada
bma@cs.uwo.ca

Abstract

Genomics and proteomics studies routinely depend on homology searches based on the strategy of finding short seed matches which are then extended. The exploding genomic data growth presents a dilemma for DNA homology search techniques: increasing seed size decreases sensitivity whereas decreasing seed size slows down computation.

We present a new homology search algorithm "PatternHunter" that uses a novel seed model for increased sensitivity and new hit-processing techniques for significantly increased speed. At Blast levels of sensitivity, PatternHunter is able to find homologies between sequences as large as human chromosomes, in mere hours on a desktop.

PatternHunter is available at http://www.bioinformaticssolutions.com. It runs on all platforms that support Java.