

Class 07: Gene Alignments (2)

Held: Thursday, 17 September 2009

Summary: Today we explore the BLAST algorithm in two ways: Through an analysis of the first paper on BLAST and through experiments using an implementation of that algorithm.

Related Pages:

- EBoard.
- Reading: Altschul et al. 1990: Basic local alignment search tool..
- Due: Response to Altschul et al..

Notes:

- For next Tuesday: Work on the “On Your Own” project from Chapter 2.
- You will *not* need to turn in today’s Web Exploration.

Overview:

- The BLAST Paper.
- Simulating the BLAST algorithm by hand.
- Web Exporation.

Having a BLAST

Our goal is to “tease apart” the work represented by the BLAST paper.

Some Standard Questions

- What is the problem domain that the paper addresses?
- What are the authors’ primary claims about their work?
- What is the structure of the paper? (Logical, rhetorical, ...)
- Why are we reading this paper? What is important about this paper?

Approximate Matching and Searching

- The problem seems to be finding all “reasonable” matches of one sequence in a larger sequence or collection of sequences.
- What measurements might we use for “reasonable”?
- What is an MSP (maximal segment pair)?
- What is a locally maximal segment pair?
- Why do we care about MSPs?
- What is an obvious way to find an MSP for two sequences?

- What alternatives are there?

The BLAST Algorithm

- What are the parameters to the algorithm? (Letters, meanings, ...)
- What is the overall structure of the BLAST algorithm? (The authors claim that the algorithm has three stages.)
 - Stage 1:
 - Stage 2:
 - Stage 3:
- How do they accomplish each stage?
- What is the running time of the algorithm?

Exploring the Algorithm

- Helpful tool: PAM Matrix Calculator
- Input sequence: AMANAPLANPANAMA
- Database: ZZZZZZZZMANNNAFLANNANANAXXXXXXX
- How do we generate the word list from the input sequence?
- Suppose our word list only contains PLAN and FLAN (everything else is too high frequency).
 - How do we search the database?
- Suppose we've matched the FLAN (a variant of the middle PLAN with high score)
 - How do we expand the word match to an approximate MSP?

Analyzing the BLAST Algorithm

- What kinds of analyses are they doing?
- Why are they doing these analyses?
- What do the analyses suggest?
- What are the potential drawbacks of using BLAST?

Web Exploration

- Let's all have fun with the Web exploration (using this algorithm we just learned about).

Copyright © 2009 Vida Praitis and Samuel A. Rebelsky. This work is licensed under a Creative Commons Attribution-NonCommercial 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.