

A typical thesis

- A typical thesis in algorithms in bioinformatics:
 - Review of topic, what is the biological motivation, what is the computational problem. This involves finding, reading and surveying papers (15-20 pages).
 - Implementation. Describe your solution of the key computational problems in order to clarify the contribution(s) in the thesis (15-20 pages)
 - Experiments. Investigate computational properties (time, space etc.), investigate biological relevance (15-20 pages).
 - Typically a thesis is accompanied by a www-page which gives access to source code or an www-interface ...

Future Topics

- Alignment of whole chromosomes. Classic methods are too slow. Use data structures, e.g. suffix trees, to make it possible to detect large scale similarities efficiently. Take large scale events such as reversals and transpositions into account. A lot of methods and literature ...
- Detecting of large scale events in genomic data, e.g. recombinations, transpositions and transpositions. This is related to aligning whole chromosomes.
- Algorithms for database search, e.g. heuristics for local alignment such as BLAST and FASTA. Implementation and investigation of techniques used in e.g. PatternHunter. A lot of methods and literature ...

Future Topics

- Algorithms for multiple alignment, e.g. how to find the best combination of tree and multiple alignment (which is also the parsimony tree reconstruction problem). Look at kalign.
- Comparative gene finding utilizing more than two sequences. Utilizing heuristic search methods to make it possible to predict gene structure in long sequences, e.g. using BLAST to decompose the problem in larger building blocks. Whole genome annotations.
- Structural RNA alignment. Similar to comparative gene finding but the aim is now to identify sequences which share similar structure. Explore SCFG's (similar to HMMs but for context-free languages). Can be used in RNA gene finding.

Future Topics

- Structural alignment of proteins.
- Threading methods for comparing protein structures and protein sequences, e.g. how well does a sequence align to a particular protein.
- Implementation and investigation of “forward” dynamic programming e.g. using cache-oblivious algorithms for breadth-first search. Does this yield better and simpler implementations of the “standard algorithms” in bioinformatics?
- Implementation of an HMM (or SCFG) framework which makes it easy to prototype and develop HMM (or SCFG) based algorithms.

Future Topics

- A framework for easy visualization of fundamental algorithms in bioinformatics ...
- Tools for visualizing complex data sets, e.g. www based access to concrete biological data sets
- Simulators for sequence evolution
- You can also look for project ideas at:

<http://www.stats.ox.ac.uk/mathgen/bioinformatics/projects/>

Discussion

Most exciting topic(s)? Least exciting topic(s)?

Which topic(s) did we spent too much/little time on?

The lectures? The reading material?

3 hours or 2+1 hours per week? Exercises or lectures?

Topics of the mandatory projects?

Written reports? Presentations in class?

Type of exam?

Finally

- Hand-in mandatory project 2 to me (at my office in the BiRC building) on Monday, October 22.
- Exam is on Monday October 29 in **Shannon-157**. Schedule is on www-page.
- AiBTaS starts **Nov 7, 09:15-12:00**, in Shannon -159, see
 - http://www.daimi.au.dk/~cstorm/courses/AiBTaS_e07/