

Computational Issues in Molecular Biology

CSE 397 / 497

Spring Semester 2004
Professor Daniel Lopresti
Department of Computer Science and Engineering
Lehigh University, Bethlehem, PA 18015
Email: dal9@lehigh.edu ~ Homepage: <http://www.cse.lehigh.edu/~lopresti>

Reading List

Tu 1/20 Introduction molecular biology I
Th 1/22 No class: Professor Lopresti in California

Tu 1/27 Introduction to molecular biology II

- (1) Chapter 1 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “The Blueprint for Life?” by D. G. Feitelson and M. Treinin, *IEEE Computer*, July 2002, pp. 34-40.
<http://ieeexplore.ieee.org/iel5/2/21883/01016899.pdf?isNumber=21883&prod=JNL&arnumber=1016899&arSt=34&ared=40&arAuthor=Feitelson%2C+G.%3B+Treinin%2C+M>
- (3) If you don't already know the material: Chapter 2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

=====

Th 1/29 Basics of sequence comparison and alignment

- (1) Section 3.1-3.2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “A General Method Applicable to the Search for Similarities in the Amino Acid Sequences of Two Proteins” by S. B. Needleman and C. D. Wunsch.
not available on WWW – see Blackboard archive
- (3) “The String-to-String Correction Problem” by R. A. Wagner and M. J. Fischer.
<http://doi.acm.org/10.1145/321796.321811>

Tu 2/3 Advanced sequence comparison and alignment

- (1) Section 3.3 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “The emerging landscape of bioinformatics software systems” by L.S. Heath and N. Ramakrishnan, *IEEE Computer*, July 2002, pp. 41-45.
<http://ieeexplore.ieee.org/iel5/2/21883/01016900.pdf?isNumber=21883&prod=JNL&arnumber=1016900&arSt=41&ared=45&arAuthor=Heath%2C+L.S.%3B+Ramakrishnan%2C+N>

=====

Th 2/5 Block editing models and algorithms

- (1) “The String-to-String Correction Problem with Block Moves,” W. F. Tichy, *ACM Transactions on Computer Systems*, vol. 2, no. 4, November 1984, pp. 309-321.
<http://doi.acm.org/10.1145/357401.357404>
- (2) “Block Edit Models for Approximate String Matching,” D. Lopresti and A. Tomkins, *Theoretical Computer Science*, vol. 181, no. 1, 1997, pp. 159-179.
<http://www.cse.lehigh.edu/~lopresti/Publications/1997/tcs97.pdf>

Tu 2/10 Database search: BLAST, FAST, and GenBank

- (1) Section 3.5 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “Rapid and Sensitive Protein Similarity Searches,” D. J. Lipman and W. R. Pearson, *Science*, vol. 227, no. 4693, 1985, pp. 1435-1441.
<http://links.jstor.org/sici?sici=0036-8075%2819850322%293%3A227%3A4693%3C1435%3ARASPSS%3E2.0.CO%3B2-S>
- (3) “GenBank,” D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell and D. L. Wheeler, *Nucleic Acids Research*, 2003, Vol. 31, No. 1, pp. 23-27.
<http://nar.oupjournals.org/cgi/reprint/31/1/23>

=====

Th 2/12 Cross-domain string matching

- (1) "Cross-Domain Approximate String Matching," D. Lopresti and G. Wilfong, *Proceedings of the Sixth International Symposium on String Processing and Information Retrieval*, September 1999, Cancún, Mexico, pp. 120-127.
<http://www.cse.lehigh.edu/~lopresti/Publications/1999/spire99.pdf>

Tu 2/17 Sequence comparison & alignment: Arthur Loder

- (1) "Alignment of Whole Genomes," A. L. Delcher, S. Kasif, R. D. Fleischmann, J. Peterson, O. White, and S. L. Salzberg, *Nucleic Acids Research*, vol. 27, no. 11, 1999, pp. 2369-2376 .
<http://www.tigr.org/software/mummer/MUMmer.pdf>
- (2) "A Space-Economical Suffix Tree Construction Algorithm," E. M. McCreight, *Journal of the ACM*, vol. 23, no. 2, April 1976, pp. 262-272.
<http://doi.acm.org/10.1145/321941.321946>

=====

Th 2/19 Sequence comparison & alignment: Jesse Wolfgang

- (1) "The Multiple Sequence Alignment Problem in Biology," H. Carrillo and D. Lipman, *SIAM Journal on Applied Mathematics*, vol. 48, no. 5, October 1988, pp. 1073-1082.
<http://links.jstor.org/sici?sici=0036-1399%28198810%2948%3A5%3C1073%3ATMSAPI%3E2.0.CO%3B2-O>
- (2) "Trees, Stars, and Multiple Biological Sequence Alignment," S. F. Altschul and D. J. Lipman, *SIAM Journal on Applied Mathematics*, vol. 49, no. 1, February 1989, pp. 197-209.
<http://links.jstor.org/sici?sici=0036-1399%28198902%2949%3A1%3C197%3ATSAMBS%3E2.0.CO%3B2-O>

**Tu 2/24 Sequencing and sequence assembly: Lan Nie
Scribe: Shi Chen**

- (1) Section 4.1-4.2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

- (2) "Genome sequence assembly: algorithms and issues," M. Pop, S.L. Salzberg, and M. Shumway, *IEEE Computer*, July 2002, pp. 47-54.
<http://ieeexplore.ieee.org/iel5/2/21883/01016901.pdf?isNumber=21883&prod=JNL&arnumber=1016901&arSt=47&ared=54&arAuthor=Pop%2C+M.%3B+Salzberg%2C+S.L.%3B+Shumway%2C+M.>

=====

Th 2/26 Sequencing and sequence assembly: Upmanyu Misra
Scribe: Deepthi Bollu

- (1) Sections 4.3-4.4 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) "The greedy path-merging algorithm for sequence assembly," D. H. Huson, K. Reinert, E. W. Myers, *Proceedings of RECOMB 2001*, pp. 157-163.
<http://doi.acm.org/10.1145/369133.369190>

Tu 3/2 Physical mapping of DNA: Shanna Terry
Scribe: Jesse Wolfgang

- (1) Sections 5.1-5.2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) "The enhanced double digest problem for DNA physical mapping," M-Y Kao, J. Samet, and W-K Sung, *Proceedings of the 7th Scandinavian Workshop on Algorithm Theory (WAT00)*, 2000, pages 383-392.
http://www.daimi.au.dk/~cstorm/bioinf_f02/project3/kao:sam:sun:2000.pdf

=====

Th 3/4 Physical mapping of DNA: Shi Chen
Scribe: Tom Morrisette

- (1) Sections 5.3-5.4 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

Tu 3/9 Spring break: no class
Th 3/11 Spring break: no class

Tu 3/16 Phylogenetic trees: Benjamin Loyle
Scribe: Patrick Schmid

- (1) Section 6.1 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “Designing Fast Converging Phylogenetic Methods,” Luay Nakhleh, Usman Roshan, Katherine St. John, Jerry Sun, Tandy Warnow, *Proc. 9th Int'l Conf. on Intelligent Systems for Mol. Biol. (ISMB 01)*, also appear in special edition of *Bioinformatics*, Vol. 17 Suppl. 1 2001, pp. S190-S198.
<http://citeseer.nj.nec.com/nakhleh01designing.html>

=====

Th 3/18 Phylogenetic trees : Bulent Moller
Scribe: Lan Nie

- (1) Sections 6.2-6.4 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “A Fast Algorithm for the Computation and Enumeration of Perfect Phylogenies when the Number of Character States is Fixed,” Sampath Kannan and Tandy Warnow, *SIAM Journal on Computing*, vol. 26, no. 6, 1997, pp. 1749-1763.
<http://epubs.siam.org/sam-bin/dbq/article/27906>

Tu 3/23 Genome rearrangements: Kevin Gaittens
Scribe: Arthur Loder

- (1) Sections 7.1-7.2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

=====

Th 3/25 Genome rearrangements: Scott Taubman

- (1) Section 7.3 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

Tu 3/30 RNA and protein structure prediction: Tom Morrisette
Scribe: Kevin Gaittens

- (1) Section 8.1 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.
- (2) “The Use of Dynamic Programming Algorithms in RNA Secondary Structure Prediction,” Chapter 7 in *Mathematical Models for DNA Sequences*, M. S. Waterman, Boca Raton, FL: CRC Press.
not available on WWW – see Blackboard archive

=====

Th 4/1 RNA and protein structure prediction: Dan Lopresti

- (1) Sections 8.2 and 8.3 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

Tu 4/6 DNA microarrays: Patrick Schmid
Scribe: Upmanyu Misra

- (1) “Introduction to Microarrays,” Section 4.1 in *Discovering Genomics, Proteomics, & Bioinformatics*, A. M. Campbell and L. J. Heyer, San Francisco, CA: Benjamin Cummings, 2003.
not available on WWW – see Blackboard archive
- (2) DNA Microarray Methodology - Flash Animation
<http://www.bio.davidson.edu/courses/genomics/chip/chip.html>

=====

Th 4/8 DNA microarrays: Stephen Hookway
Scribe: Charles Omsby

- (1) “Basic microarray analysis: grouping and feature reduction,” Soumya Raychaudhuri, Patrick D. Sutphin, Jeffrey T. Chang and Russ B. Altman, *TRENDS in Biotechnology*, vol.19, no.5 May 2001, pp. 189-193.
<http://smi-web.stanford.edu/people/sxr/pdfs/tibTech.pdf>
- (2) “Self Organizing Maps,” Tom Germano, March 23, 1999.
<http://davis.wpi.edu/~matt/courses/soms/>

Tu 4/13 DNA computing: Deepthi Bollu

- (1) "Molecular Computation of Solutions to Combinatorial Problems," Leonard M. Adleman, *Science*, vol. 266, no. 5187, November 11, 1994, pp. 1021-1024.
- (2) Section 9.1 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

=====

Th 4/15 No class: Professor Lopresti in Pittsburgh

Tu 4/20 DNA computing: Charles Omsby

- (1) "DNA Solution of Hard Computational Problems," Richard J. Lipton, *Science*, vol. 268, no. 5210, April 28, 1995, pp. 542-545.
- (2) Section 9.2 in *Introduction to Computational Molecular Biology*, J. Setubal and J. Meidanis, Pacific Grove, CA: Brooks/Cole Publishing, 1997.

=====

Th 4/22 History

- (1) "The Invention of the Genetic Code," Brian Hayes, *American Scientist*, vol. 86, no. 1, January-February 1998, pp. 8-14.
<http://www.amsci.org/amsci/issues/Comsci98/compsci9801.html>

Tu 4/27 Ethics

- (1) "Does Genetic Research Threaten Our Civil Liberties?," Philip Bereano, ActionBioscience.org, August 2000.
<http://www.actionbioscience.org/genomic/bereano.html>
- (2) "Toward a New Vocabulary of Human Genetic Variation," by Pamela Sankar and Mildred K. Cho, *Science*, vol. 298, November 2002, pp. 1337-1338.
<http://www.sciencemag.org/cgi/reprint/298/5597/1337.pdf>
- (3) "Procreative Beneficence: Why We Should Select the Best Children," by Julian Savulescu, *Bioethics*, vol. 15, no. 5/6, 2001.

http://www.blackwellpublishing.com/content/BPL_Images/Journal_Samples/BIOT0269-9702~15~5&6~251%5C251.pdf

(4) “How DNA Evidence Works,” by Ann Meeker-O'Connell, HowStuffWorks.com.

<http://www.howstuffworks.com/dna-evidence.htm>

(5) “Fight Errupts Over DNA Fingerprinting,” by Leslie Roberts, *Science*, vol. 254, no. 5039, December 1991, pp. 1721-1732.

<http://links.jstor.org/sici?sici=0036-8075%2819911220%293%3A254%3A5039%3C1721%3AFEODF%3E2.0.CO%3B2-W>

=====

Th 4/29 Course wrap-up and (optional) student presentations of final projects