CS 6824:
Intermediate Paper List

Lenwood S. Heath

October 18, 2012

1 Paper 1

Limitations of Next-Generation Genome Sequence Assembly [1]

2 Paper 2

How to Apply de Bruijn Graphs to Genome Assembly [6]

3 Paper 3

Efficient Storage of High Throughput DNA Sequencing Data Using Reference-Based Compression [9]

4 Paper 4

A Genome Compression Algorithm Supporting Manipulation [8]

5 Paper 5

Assembly Algorithms for Next-Generation Sequencing Data [20]

6 Paper 6

Parallelized SHort Read Assembly of Large Genomes Using de Bruijn Graphs [17]

7 Paper 7

GAGE: A Critical Evaluation of Genome Assemblies and Assembly Algorithms [22]

8 Paper 8

Next-Generation Transcriptome Assembly [19]

9 Paper 9

Basic Local Alignment Search Tool [2]

10 Paper 10

LAGAN and Multi-LAGAN: Efficient Tools for Large-Scale Multiple Alignment of Genomic Data [4]

11 Paper 11

Profile Hidden Markov Models [7]

12 Paper 12

Genomics-Based Approaches to Gene Discovery in Innate Immunity [23]

13 Paper 13

Pairwise Sequence Alignment for Very Long Sequences on GPUs [14]

14 Paper 14

CUDA Compatible GPU Cards as Efficient Hardware Accelerators for Smith-Waterman Sequence Alignment [18]

15 Paper 15

Biological Network Comparison Using Graphlet Degree Distribution [21]

16 Paper 16

Integrative Network Alignment Reveals Large Regions of Global Network Similarity in Yeast and Human [12]

17 Paper 17

Global Alignment of Multiple Proteins Interaction Networks With Application to Functional Orthology Detection [24]

18 Paper 18

IsoRankN: Spectral Methods for Global Alignment of Multiple Protein Networks [15]

19 Paper 19

MEME: Discovering and Analyzing DNA and Protein Sequence Motifs [3]

20 Paper 20

GADEM: A Genetic Algorithm Guided Formation of Spaced Dyads Couple With an EM ALgorithm for Motif Discovery [13]

21 Paper 21

Intronic Motif Pairs Cooperate Across Exons to Promote Pre-mRNA Splicing [10]

22 Paper 22

Quantitative Evaluation of All Hexamers as Exonic Splicing Elements [11]

23 Paper 23

Computational Approaches to Unveiling Ancient Genome Duplications [25]

24 Paper 24

Nonrandom Divergence of Gene Expression Following Gene and Genome Duplications in the Flowering Plant *Arabidopsis thaliana* [5]

25 Paper 25

Evolution of Gene Families Based on Gene Duplication, Loss, Accumulated Change, and Innovation [26]

26 Paper 26

A Bayesian Model for Gene Family Evolution [16]

References

[1] Can Alkan, Saba Sajjadian, and Evan E. Eichler. Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1):61–65, 2011.

[2] Stephen F. Altschul, Warren Gish, Webb Miller, Eugene W. Myers, and David J. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, October 1990.

[3] Timothy L. Bailey, Nadya Williams, Chris Misleh, and Wilfred W. Li. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research*, 34(suppl 2):W369–W373, July 2006.

[4] Michael Brudno, Chuong B. Do, Gregory M. Cooper, Michael F. Kim, Eugene Davydov, Nisc C. Program, Eric D. Green, Arend Sidow, and Serafim Batzoglou. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Research*, 13(4):721–731, April 2003.

[5] Tineke Casneuf, Stefanie De Bodt, Jeroen Raes, Steven Maere, and Yves Van de Peer. Nonrandom divergence of gene expression following gene and genome duplications in the flowering plant *Arabidopsis thaliana*. *Genome Biology*, 7(2):R13+, 2006.

[6] Phillip E. C. Compeau, Pavel A. Pevzner, and Glenn Tesler. How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology*, 29(11):987–991, November 2011.

[7] S. R. Eddy. Profile hidden Markov models. *Bioinformatics*, 14(9):755–763, January 1998.

[8] Lenwood S. Heath, Ao-ping Hou, Huadong Zia, and Liqing Zhang. A genome compression algorithm supporting manipulation. In *Proc LSS Comput Syst Bioinform Conf.*, pages 38–49, August 2010.

[9] Markus Hsi-Yang Fritz, Rasko Leinonen, Guy Cochrane, and Ewan Birney. Efficient storage of high throughput DNA sequencing data using reference-based compression. *Genome Research*, 21(5):734–740, May 2011.

[10] Shengdong Ke and Lawrence Chasin. Intronic motif pairs cooperate across exons to promote pre-mRNA splicing. *Genome Biology*, 11(8):R84+, 2010.

[11] Shengdong Ke, Shulian Shang, Sergey M. Kalachikov, Irina Morozova, Lin Yu, James J. Russo, Jingyue Ju, and Lawrence A. Chasin. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Research*, 21(8):1360–1374, August 2011.

[12] Oleksii Kuchaiev and Nataša Pržulj. Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 27(10):1390–1396, March 2011.

[13] Leping Li. GADEM: A genetic algorithm guided formation of spaced dyads coupled with an EM algorithm for motif discovery. *Journal of Computational Biology*, 16(2):317–329, February 2009.

[14] Zhenyu Li, Yanxiang Chen, Desheng Mu, Jianying Yuan, Yujian Shi, Hao Zhang, Jun Gan, Nan Li, Xuesong Hu, Binghang Liu, Bicheng Yang, and Wei Fan. Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics*, 11(1):25–37, January 2012.

[15] Chung-Shou Liao, Kanghao Lu, Michael Baym, Rohit Singh, and Bonnie Berger. IsoRankN: spectral methods for global alignment of multiple protein networks. *Bioinformatics*, 25(12):i253–i258, June 2009.

[16] Liang Liu, Lili Yu, Venugopal Kalavacharla, and Zhanji Liu. A Bayesian model for gene family evolution. *BMC Bioinformatics*, 12(1):426+, 2011.

[17] Yongchao Liu, Bertil Schmidt, and Douglas L. Maskell. Parallelized short read assembly of large genomes using de Bruijn graphs. *BMC bioinformatics*, 12(1):354+, August 2011.

[18] Svetlin Manavski and Giorgio Valle. CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment. *BMC Bioinformatics*, 9(Suppl 2):S10–9, March 2008.

[19] Jeffrey A. Martin and Zhong Wang. Next-generation transcriptome assembly. *Nature reviews. Genetics*, 12(10):671–682, October 2011.

[20] Jason R. Miller, Sergey Koren, and Granger Sutton. Assembly algorithms for next-generation sequencing data. *Genomics*, 95(6):315–327, June 2010.

[21] Nataša Pržulj. Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183, January 2007.

[22] Steven L. Salzberg, Adam M. Phillippy, Aleksey Zimin, Daniela Puiu, Tanja Magoc, Sergey Koren, Todd J. Treangen, Michael C. Schatz, Arthur L. Delcher, Michael Roberts, Guillaume Marçais, Mihai Pop, and James A. Yorke. GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3):557–567, December 2011.

[23] Todd Scheetz, Jennifer A. Bartlett, Jesse D. Walters, Brian C. Schutte, Thomas L. Casavant, and McCray. Genomics-based approaches to gene discovery in innate immunity. *Immunological Reviews*, 190(1):137–145, 2002.

[24] Rohit Singh, Jinbo Xu, and Bonnie Berger. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proceedings of the National Academy of Sciences*, 105(35):12763–12768, September 2008.

[25] Yves Van de Peer. Computational approaches to unveiling ancient genome duplications. *Nat Rev Genet*, 5(10):752–763, October 2004.

[26] Damian Wójtowicz and Jerzy Tiuryn. Evolution of gene families based on gene duplication, loss, accumulated change, and innovation. *Journal of Computational Biology*, 14(4):479–495, May 2007.