

## Introduction to Bioinformatics (Graduate Section)

For the time being, I will not require written reviews. However, unlike previous years, I will not assign papers in advance to specific reviewers. Rather I will call on one participant at random to lead the discussion of a particular paper.

When you are called, you should assume that everyone has actually read the paper so you do not have present a detailed summary of the paper. Instead, focus on the following:

- very brief synopsis of paper - not a detailed description of every point; focus on the logical argument and the following questions: What data? What analysis? What is their argument and does it support their conclusions? Why is this paper important?
- Any confusing or difficult to understand sections
- strong points of the paper
- weak points of the paper
- validity of conclusions

59500 Paper schedule		
		Tentative (Don't read yet)
Week	Lecture Topic	Papers
<b>6 September</b>		<ul style="list-style-type: none"> <li>• Needleman SB, Wunsch CD., A general method applicable to the search for similarities in the amino acid sequence of two proteins. <i>J Mol Biol</i> 1970 48:443-453. <a href="#">PDF</a></li> <li>• Fitch WM, Smith TF. Optimal sequence alignments. <i>Proc Natl Acad Sci USA</i> 1983 80:1382-6. <a href="#">PDF</a></li> </ul>
<b>13 September</b>	Sequence Alignment and Scoring	<ul style="list-style-type: none"> <li>• Henikoff S, Henikoff JG., Amino acid substitution matrices from protein blocks. <i>Proc Natl Acad Sci USA</i> 1992, 89, 10915-10919. <a href="#">PDF</a></li> <li>• Chang MSS , Benner SA, Empirical Analysis of Protein Insertions and Deletions Determining Parameters for the Correct Placement of Gaps in Protein Sequence Alignments, <i>J Mol Biol</i> 2004 341:617–631. <a href="#">PDF</a></li> </ul>
<b>20 September</b>	Database searching	<ul style="list-style-type: none"> <li>• Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, 1997, <i>Nucleic Acids Res</i> 25:3389-3402. <a href="#">PDF</a></li> <li>• Kent WJ. BLAT--the BLAST-like alignment tool. <i>Genome Res</i> 2002 12:656-64. <a href="#">PDF</a></li> </ul>
<b>27 September</b>	Motifs	<ul style="list-style-type: none"> <li>• Nevill-Manning CG, Wu TD, Brutlag DL. Highly specific protein sequence motifs for genome analysis. <i>Proc Natl Acad Sci USA</i> 95, 5865-5871, 1998. <a href="#">PDF</a></li> <li>• Trachana K, Larsson TA, Powell S, Chen WH, Doerks T, Muller J, Bork P. Orthology prediction methods: a quality assessment using curated protein families. <i>Bioessays</i> 33, 769-80, 2011. <a href="#">PDF</a></li> </ul>

## 59500 Paper schedule

Tentative (Don't read yet)

Week	Lecture Topic	Papers
4 October	Trees	<ul style="list-style-type: none"> <li>• L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A. C. Wilson, <i>Science</i> 253, 1503 (1991). <ul style="list-style-type: none"> <li>◦ Templeton AR, Human origins and analysis of mitochondrial DNA sequences, <i>Science</i> 255, 737 (1992)</li> <li>◦ Hedges SB, Kumar, S, Tamura K, <i>Science</i> 255, 737-739 (1992)</li> </ul> </li> <li>• Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading. <i>Systematic Zool</i> 27, 401-410, 1978.</li> </ul>
11 October	Multiple Alignment	<ul style="list-style-type: none"> <li>• Thompson JD, Higgins DG, Gibson TJ, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, <i>Nucleic Acids Research</i> 22, 4673-4680, 1994.</li> <li>• RC Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. <i>Nucleic Acids Res.</i> 32, 1792-1797, 2004.</li> </ul>
18 October	Profiles & HMMs	<ul style="list-style-type: none"> <li>• LS Johnson, SR Eddy, E Portugaly. Hidden Markov model speed heuristic and iterative HMM search procedure. <i>BMC Bioinformatics</i> 11, 431, 2010.</li> <li>• Kawaji H, Schönbach C, Matsuo Y, Kawai J, Okazaki Y, Hayashizaki Y, Matsuda H. Exploration of novel motifs derived from mouse cDNA sequences. <i>Genome Res.</i> 12, 367-378, 2002</li> </ul>
25 October	Genome Sequencing and annotation	<ul style="list-style-type: none"> <li>• AM Phillippy, MC Schatz and M Pop, Genome assembly forensics: finding the elusive mis-assembly, <i>Genome Biology</i> 9,R55, 2008.</li> <li>• Lomsadze A, Ter-Hovhannisyann V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. <i>Nucleic Acids Res.</i> 33, 6494-6506, 2005.</li> </ul>
1 November	Gene Finding and Annotation	<ul style="list-style-type: none"> <li>• Kamoun C, Payen T, Hua-Van A, Filée J. Improving prokaryotic transposable elements identification using a combination of de novo and profile HMM methods. <i>BMC Genomics</i> 14:700, 2013.</li> <li>• R Overbeek, M Fonstein, M D'souza, G D. Pusch, and N Maltsev, The use of gene clusters to infer functional coupling, <i>Proc. Natl. Acad. Sci. USA</i>, 96, 2896–2901, 1999</li> </ul>
8 November	Protein Structure Prediction	<ul style="list-style-type: none"> <li>• Xue B, Williams RW, Oldfield CJ, Dunker AK, Uversky VN. Archaic chaos: intrinsically disordered proteins in Archaea. <i>BMC Syst Biol.</i> 2010 May 28;4 Suppl 1:S1. <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2880407/pdf/1752-0509-4-S1-S1.pdf">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2880407/pdf/1752-0509-4-S1-S1.pdf</a></li> <li>• Zhang X, Mesirov JP, Waltz DL. Hybrid system for protein secondary structure prediction. <i>J Mol Biol.</i> 1992 Jun 20;225(4):1049-63. <a href="http://www.sciencedirect.com/science/article/pii/002228369290104R#">http://www.sciencedirect.com/science/article/pii/002228369290104R#</a></li> </ul>
15 November	Protein Modeling	<ul style="list-style-type: none"> <li>• Bryant SH, Lawrence CE. An empirical energy function for threading protein sequence through the folding motif. <i>Proteins</i> 16, 92-112, 1993</li> <li>• Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. <i>J.Mol.Biol</i> 268, 209-225, 1997.</li> </ul>

## 59500 Paper schedule

Tentative (Don't read yet)

Week	Lecture Topic	Papers
<b>22 Nov</b>	Thanksgiving holiday	Thanksgiving holiday - TBA - there will be a written review for this week
<b>29 November</b>	Protein Homology modeling and docking	<ul style="list-style-type: none"> <li>• Zhou JM, Lee E, Kanapathy-Sinnaiaha F, Park Y, Kornblatt JA, Lim Y, Ibrahim RK. Structure-function relationships of wheat flavone O-methyltransferase: Homology modeling and site-directed mutagenesis. <i>BMC Plant Biology</i>, 10:15, 2010. <a href="http://www.biomedcentral.com/content/pdf/1471-2229-10-156.pdf">http://www.biomedcentral.com/content/pdf/1471-2229-10-156.pdf</a></li> <li>• Geisler M, Wilczynska M, Karpinski S, Kleczkowski LA. Toward a blueprint for UDP-glucose pyrophosphorylase structure/function properties: homology-modeling analyses. <i>Plant Mol Biol</i>. 56:783-794, 2004. <a href="http://www.springerlink.com/content/g64688w486218k56/fulltext.pdf">http://www.springerlink.com/content/g64688w486218k56/fulltext.pdf</a></li> </ul>
<b>6 December</b>	Systems Biology	<ul style="list-style-type: none"> <li>• Barabasi A-L, Albert R. Emergence of scaling in random networks. <i>Science</i> 286, 509-512, 1999.</li> <li>• Michael P. H. Stumpf, Thomas Thorne, Eric de Silva, Ronald Stewart, Hyeong Jun An, Michael Lappe and Carsten Wiuf Estimating the size of the human interactome. <i>PNAS</i> 105, 6959-6964, 2008.</li> </ul>