

EN 600.439/639: Computational Genomics

JP = "Jones and Pevzner" = "An Introduction to Bioinformatics Algorithms"

G = "Gusfield" = "Algorithms on Strings, Trees and Sequences"

Week	Mon: 3 - 4:15pm	Wed: 3 - 4:15pm	Preparation
1/28	<p>Computational genomics</p> <p>What is genomics? Why is it important? What is computational genomics? What do computer scientists do in this field? What role does DNA sequencing play?</p>	<p>Biological background</p> <p>Evolution Genotype and phenotype DNA and the genome The Central Dogma: DNA, RNA, Protein Diversity of species Differentiation</p> <p>Homework 0 assigned</p>	<p>Watch:</p> <ul style="list-style-type: none"> Videos 1 – 3 (Link on Piazza page in "Videos" section of "Resources" tab.) (Optional): Videos 4 – 6 <p>Read:</p> <ul style="list-style-type: none"> JP: Ch. 3 (Optional): "Molecular Biology for Computer Scientists." Link on Piazza page in "Readings" section of "Resources" tab.
2/4	<p>Exact matching</p> <p>DNA sequencing and the point-of-origin problem Naïve exact matching Boyer-Moore</p>	<p>Inverted indexing</p> <p>Online versus offline Indexing Inverted index Contiguous seeds Index-assisted naïve algorithm</p> <p>Homework 0 due</p>	<p>Read:</p> <ul style="list-style-type: none"> G: 1.1, 2.2 "Inverted genome indexes." Link on Piazza page in "Readings" section of "Resources" tab.
2/11	<p>Suffix indexing</p> <p>Trees and tries Suffix tree Applications of suffix tree</p> <p>Homework 1 assigned</p>	<p>Suffix indexing</p> <p>Suffix array Burrows-Wheeler Transform</p>	<p>Read:</p> <ul style="list-style-type: none"> G: Ch. 5 (all) G: Ch. ??? Reading for BWT
2/18	<p>Suffix indexing</p> <p>FM Index Methods used in practice</p> <p>Alignment</p> <p>Approximate matching and alignment</p>	<p>Dynamic programming alignment</p> <p>Edit distance Needleman-Wunsch (global alignment) Smith-Waterman (local alignment)</p>	<p>Read:</p> <ul style="list-style-type: none"> G: 11 "Full-text indexes for alignment" part 1. Link on Piazza page in "Readings" section of "Resources" tab.

2/25	Dynamic programming alignment Variations Scoring Time and space improvements Real-world performance	Index-assisted alignment Trees and tries again 1-mismatch example Full algorithm Performance improvements Homework 1 due	Read: <ul style="list-style-type: none"> G: 12.2, 12.7 "Full-text indexes for alignment" part 2. Link on Piazza page in "Readings" section of "Resources" tab.
3/4	Index-assisted alignment Two views of alignment Real software packages Variants on the alignment problem	Assembly Human genome project Shortest common superstring Greedy assembly Repeats	Read: <ul style="list-style-type: none"> JP: 8.1-8.4
3/11	Midterm Exam in class On everything up to but not including assembly	Assembly Repeats Assembly graphs Hamiltonian path	Read: <ul style="list-style-type: none"> JP: 8.8-8.9 "Assembly of large genomes using second-generation sequencing." Link on Piazza page in "Readings" section of "Resources" tab.
3/18	Spring break		
3/25	Assembly De Bruijn graphs Eulerian path path Homework 2 assigned Project ideas due	Assembly Guest lecturer?	Read: <ul style="list-style-type: none"> "How to apply de Bruijn graphs to genome assembly." Link on Piazza page in "Readings" section of "Resources" tab.
4/1	Classifying DNA elements Genome signal and background Markov assumption and Hidden Markov Model Finding CpG islands	Classifying DNA elements Simple genes HMM classifiers for gene finding Exons and introns More complex HMM classifiers	Read: <ul style="list-style-type: none"> JP: 11.1-11.3

		Homework 2 due	
4/8	Classifying DNA elements What does all the DNA do? Regulatory DNA HMMs classifiers for discovering regulatory DNA Homework 3 assigned	The Seqs DNA sequencers: not just for sequencing genomes Capture sequencing RNA-seq: Sequencing mRNA ChIP-seq: Sequencing bound DNA Bisulfite sequencing: Revealing the epigenome	
4/15	RNA-seq Eukaryotic genes and alternative splicing Spliced alignment	RNA-seq The assembly-quantitation co-problem Assembling parsimonious transcripts	
4/22	Indexing and compressing reads Reference-based methods Sketch methods	Success stories & desiderata Better models Better methods for analyzing big sequence data ...for storing big sequence data ...for making big sequence data useful Homework 3 due	
4/29	Final project presentations (1)	Final project presentations (2)	
5/3 is last day of class – Final project due 11:59pm on 5/3			