

Review

BMI/CS 576

www.biostat.wisc.edu/bmi576/

Colin Dewey

cdewey@biostat.wisc.edu

Fall 2008

Sequence Assembly

- How does fragment assembly work?
- Overlap graphs
- Hamiltonian paths/cycles
- Eulerian paths/cycles
- Greedy algorithms
- What are the main difficulties in assembly?

Sequence Alignment

- What is an alignment?
- Scoring functions for alignments
- Dynamic programming
- Global alignment
- Local alignment
- Affine gap alignment

Probabilistic Sequence Models

- Markov chains
 - Use in discrimination
 - Parameter estimation
 - Higher order ($>1^{\text{st}}$) markov chains
- Hidden Markov models
 - Forward algorithm
 - Viterbi algorithm
 - Baum-Welch (EM) algorithm

Phylogeny

- Distance-based methods
 - UPGMA
 - Neighbor-joining
- Parsimony-based methods
 - Fitch's algorithm
 - Weighted parsimony
 - Branch and bound

Multiple sequence alignment

- Scoring functions for multiple alignment
 - Sum of pairs
 - Minimum entropy
- Methods
 - Dynamic programming
 - Star alignment
 - Progressive (using a tree) alignment
 - Profile HMMs

Microarrays/Gene expression

- What is a microarray? Gene expression?
- Methods
 - Hierarchical clustering
 - K-Means
 - EM clustering
- False discovery rate (FDR)
- Classification
 - Decision tree

Biological networks

- What types of biological networks are there?
- Bayesian networks
 - Joint probability factorization (CPDs)
 - Markov blankets
 - Learning
 - Bayesian analysis – conjugate priors
 - KL-divergence, mutual information
 - Sparse candidate algorithm
 - Inference
 - Elimination algorithm

Protein structure prediction

- Primary, secondary, tertiary, quaternary structure
- What factors determine 3D structure?
- Techniques
 - Homology modeling
 - Threading
 - Branch and bound algorithms
 - Fragment assembly (Rosetta)
 - Molecular dynamics