

# CMSC 858P Project2 Proposal

Phylogenetic tree construction using  
neighbor joining

Fatih Kaya

# Phylogenetic trees

- Evolutionary relationships between genes and organisms
- The actual pattern of evolutionary history is the evolutionary tree

# Construction

- Reconstruction is a statistical problem
- Three classes of methods
  - Maximum likelihood
  - Maximum parsimony
  - Distance based methods

# Construction

- Maximum likelihood
  - Maximize the probability of the genetic data given the tree
  - Result in most accurate trees
  - Need long computation time to construct a tree
- Maximum Parsimony
  - Looks at the evolutionary changes required
  - The best tree is the one which needs the fewest changes
  - Long computation time is needed to construct a tree

# Construction

- Distance based methods (Neighbor Joining)
  - Based on the minimum-evolution criteria
  - Find a tree such that total branch lengths of paths between sequences fit a matrix of pairwise distances between sequences.
  - Bottom-up clustering method

# Neighbor Joining

- Build the distance matrix from the sequence data by acquiring a set of pairwise distance values. Then NJ tries to fit a tree to this matrix.
- Recursive algorithm
  - Use the distance matrix to find the closest pair
  - Add new node
  - Update distance matrix

# Neighbor Joining

- Advantages
  - simple
  - fast
  - fit for sequences having high similarity scores
- Disadvantages
  - all sites are equally treated
  - not applicable to distantly related sequences
  - some of the information is lost